Nadav Ahituv *Editor*

Gene Regulatory Sequences and Human Disease

Springer

Gene Regulatory Sequences and Human Disease

Nadav Ahituv
Editor

Gene Regulatory Sequences and Human Disease

Springer

Editor
Nadav Ahituv
Department of Bioengineering
and Therapeutic Sciences
Institute for Human Genetics
University of California San Francisco
San Francisco, CA, USA

ISBN 978-1-4614-1682-1 e-ISBN 978-1-4614-1683-8
DOI 10.1007/978-1-4614-1683-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012937214

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Advanced sequencing technologies can now allow us to obtain individual genomic sequences. These genomes contain an overwhelming amount of nucleotide variation per individual, with the majority of the variants being noncoding. However, our current ability to provide a functional interpretation to these noncoding variants is extremely lagging compared to protein coding sequences. This book provides seminal examples of how these noncoding variants and epigenetic changes can be associated with human disease by altering gene regulation. While the current number of examples is very limited, the methodologies and techniques described in this book can serve as a model for researchers to associate additional noncoding variants with human disease. In addition, future development of technologies that will enable to functionally characterize noncoding gene regulatory variants in a high-throughput manner will move this field forward and expand our knowledge of gene regulation. Combined, this will lead to a better understanding of the "gene regulatory code." Other than allowing us to obtain a better diagnosis and understanding of the genetic causes of human disease, it will be of extreme importance to numerous other biological disciplines. Biologists have long been in need of defined sequences that drive precise patterns of expression. Such sequences can be used to express recombinant proteins or to overexpress various proteins in precise locations. These sequences could also be used to target specific molecules to certain tissues for gene therapy purposes. Gene regulatory elements are also important developmental regulators, and the understanding of the factors that regulate these developmental genes can increase our knowledge of development. In evolution, regulatory sequences are thought to be a major contributor to the evolution of form. An increased understanding of the "gene regulatory code" will vastly assist these and other disciplines.

I want to thank all the authors for their outstanding contribution to this book. I would like to acknowledge UCSF and my funding sources (NICHD, NHGRI, NIGMS, NIDDK, NINDS) who allow me to undertake this scientific journey. I would also like to thank my current lab members (Ramon Y. Birnbaum, Mee J. Kim, Karl Murphy, Nir Oksenberg, Robin P. Smith, Julia E. VanderMeer) for their helpful comments and suggestions. Finally, I would like to thank my family (Sarit, Amit, and Yonatan) for their support.

San Francisco, CA, USA Nadav Ahituv

Contents

Contributors

Nadav Ahituv Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA

Jeanne Amiel Department of Genetics, University Paris Descartes and INSERM U-781, Paris, France

Sabina Benko Department of Genetics, University Paris Descartes and INSERM U-781, Paris, France

Jonathan E. Bird Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Rockville, MD, USA

Walid D. Fakhouri Department of Microbiology and Molecular Genetics and Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA

Thomas B. Friedman Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Rockville, MD, USA

Christopher T. Gordon Department of Genetics, University Paris Descartes and INSERM U-781, Paris, France

Huilin Jin Kennedy Institute of Rheumatology, University of Oxford, 65 Aspentea Road, London

Ian D. Krantz Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Stavros Lomvardas Department of Anatomy, University of California San Francisco, San Francisco, CA, USA

Gabriela G. Loots Biology and Biotechnology Division, Lawrence Livermore National Laboratory Livermore, CA, USA

School of Natural Sciences, University of California Merced, Merced, CA, USA

Stanislas Lyonnet Department of Genetics, University Paris Descartes and INSERM U-781, Paris, France

Angeliki Magklara Department of Anatomy, University of California San Francisco, San Francisco, CA, USA

Levi H. Makala Georgia's Health Sciences University, Augusta, GA, USA

Andrew S. McCallion Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Marcelo A. Nobrega Department of Human Genetics, University of Chicago, Chicago, IL, USA

Betty S. Pace Georgia's Health Sciences University, Augusta, GA, USA

Stuart H. Ralston Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK

Brian C. Schutte Department of Microbiology and Molecular Genetics and Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA

Zachary E. Stine McKusick – Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Julia E. VanderMeer Department of Bioengineering and Therapeutic Sciences and Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA

Nora F. Wasserman Department of Human Genetics, University of Chicago, Chicago, IL, USA

Dongbin Xu Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Daniel Zemke Department of Microbiology and Molecular Genetics and Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA

Chapter 1
Gene Regulatory Elements

Nadav Ahituv

Abstract While the annotation and functional characterization of the 2% of our genome that encodes for protein has been extremely successful, the remaining 98% still remains primarily uncharted territory. Within this territory reside gene regulatory sequences that instruct genes when, where, and at what levels to turn on or off. There is abundant evidence, as described in this book, that nucleotide and epigenetic changes in these gene regulatory sequences can lead to human disease. In this chapter, we will define the different types of gene regulatory elements (promoters, enhancers, silencers, and insulators) and how to identify and functionally characterize them.

Keywords Promoter • Enhancer • Silencer • Insulator • Locus control region • Transcription factors • Nucleosome positioning • DNase I hypersensitive sites • ChIP • 3C

1.1 Introduction

The human genome consists of ~3.2 billion base pairs and encompasses around 20,500 genes (Clamp et al. 2007) which make up only ~1.6% of its content. Repetitive sequences make up an additional ~50%, and the remaining 48% is composed of DNA sequences with primarily unknown function. One vital function that is clearly

N. Ahituv (✉)
Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, 513 Parnassus Ave., Health Sciences East building, Room HSE901H, Box 0794, San Francisco, CA 94143, USA
e-mail: nadav.ahituv@ucsf.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_1,

embedded in these sequences is gene regulation, instructing genes when and where to turn on or off at specific levels. There is a variety of clinical and molecular data, as described in this book, demonstrating that regulatory elements can harbor mutational events that lead to human disease. However, the identification of distinctive nucleotide or epigenetic changes within these elements that lead to human disease has been extremely limited due in part to the vast genomic noncoding space in which to search for them, their scattered distribution, the unavailability of an established regulatory code, and the difficulties in linking these elements to specific genes.

The identification and functional characterization of these gene regulatory elements is not only important for their association with human disease but also for several other biological disciplines. Biologists have long been in need of defined sequences that drive precise patterns of expression. Such sequences can be used to express recombinant proteins (e.g., *Cre* to generate tissue-specific knockouts) or to overexpress various proteins in precise locations. These sequences could also be used to target specific DNA sequences to certain tissues for gene therapy purposes. Gene regulatory elements are also important developmental regulators, and the understanding of the factors that regulate these developmental genes can increase our knowledge of development. In evolution, regulatory sequences are thought to be a major contributor to the evolution of form (Carroll 2005). More and more examples are now being reported in various organisms, including humans (Prabhakar et al. 2008; McLean et al. 2011), that highlight the effect these sequences can have on morphological differences between species. An increased understanding of these gene regulatory elements will vastly assist these disciplines.

1.2 The Different Kinds of Gene Regulatory Elements

There are several different types of gene regulatory elements. In this section, we will define the most commonly characterized elements: promoters, enhancers, silencers, and insulators. The general dogma is that these regulatory elements get activated by the binding of *transcription factors*, proteins that bind to specific DNA sequences, and control mRNA transcription. There could be several transcription factors that need to bind to one regulatory element in order to activate it. In addition, several other proteins, called *transcription cofactors*, bind to the transcription factors themselves to control transcription.

The binding of these sequences changes the *nucleosome positioning* in that region. The *nucleosome* consists of 147 bp of DNA wrapped around a histone core. The binding of transcription-related proteins repositions the nucleosome and changes it into a more open state. As later described for DNase hypersensitive sites (Sect. 1.3.1), this change in nucleosome state can be used for the discovery of active gene regulatory elements. The nucleosome core consists of histone proteins which can have various posttranslational modifications. These modifications affect the state of this genomic region and can also be used to detect various gene regulatory elements as described in detail in the chromatin immunoprecipitation section (Sect. 1.3.3).

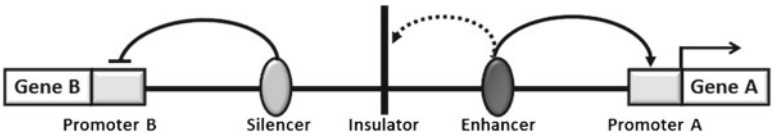

Fig. 1.1 *Schematic representation of the various gene regulatory elements described in this chapter.* Two different promoters are located next to the two different genes. By binding to *promoter A*, an enhancer actively regulates *Gene A* and leads to its transcription, as indicated by the *black arrow* above *Gene A*. An insulator prevents this enhancer from activating *Gene B*, which is not transcribed due to regulation by a silencer element

1.2.1 Promoters

The best characterized gene regulatory element is the promoter, in part due to its established location relative to the gene that it regulates. The *promoter* resides at the beginning of the gene and serves as the site where the transcription machinery assembles and transcription of the gene begins (Fig. 1.1). The *core promoter* is defined as the minimal stretch of DNA sequence that is sufficient to allow the RNA polymerase II machinery to initiate transcription (Butler and Kadonaga 2002; Smale and Kadonaga 2003). It is typically 35–40 base pairs (bp) long and encompasses several sequence motifs, such as the following: TATA box, TFIIB recognition element (BRE), initiator element (Inr), and the downstream core promoter element (DPE). It is important to note that not all of these elements need to be present in order to define a core promoter.

The *TATA box* was the first core promoter element to be discovered. Its consensus sequence is TATAAA, but this can be modified in many cases as long as the sequence remains A/T rich. In humans, it is thought to be bound predominantly by the TATA-binding protein (TBP) which is part of the transcription factor IID (TFIID) multiprotein complex that contributes to transcription initiation (Smale and Kadonaga 2003). 32% of all human promoters (Suzuki et al. 2001) are thought to contain a TATA box. The *TFIIB recognition element (BRE)* is a binding site for the transcription factor IIB (TFIIB) and is thought to facilitate the incorporation of this transcription factor to the transcription initiation complex (Lagrange et al. 1998). It has a consensus sequence of G/C-G/C-G/A-C-G-C-C and is usually located immediately upstream to the TATA box. The *initiator element (Inr)* is usually −3 to +5 bp from the transcription start site (TSS) with a typical consensus sequence of C/T-C/T-A-N-T/A-C/T-C/T. It is thought to be recognized by and interact with TFIID and RNA Polymerase II at different steps during the transcription process. The *downstream core promoter element (DPE)* is a binding site for TFIID that is usually found in promoters lacking a TATA box. It is located 28–32 bp upstream of the Inr, and the distance between both of these elements was shown to be important for proper TFIID binding and transcription (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). Its consensus sequence is A/G-G-A/T-C/T-G/A/C.

CpG islands are defined as sequences that are at least 200 long with a G/C percentage that is greater than 50% (Gardiner-Garden and Frommer 1987). The human genome is estimated to have ~29,000 of these islands (Smale and Kadonaga 2003), half of which are near gene promoters (Suzuki et al. 2001). Promoters with CpG islands usually lack a TATA box and a DPE and have multiple binding sites for the transcription factor SP1 that is thought to direct the transcription machinery. CpG islands are usually unmethylated if the gene they regulate is expressed.

The *proximal promoter* is the region that is in the immediate vicinity (−250 to +250 bp) of the TSS of the gene. It can contain several transcription factor binding sites (TFBS) and is thought to serve as a tethering element for distant enhancers, enabling them to interact with the core promoter (Calhoun et al. 2002). Several additional promoter elements have been discovered. Since this book primarily deals with more distant regulatory elements, see Smale and Kadonaga (2003) and Riethoven (2010) for a more detailed review of various promoter elements.

1.2.2 Enhancers

Enhancers turn on the promoters at specific locations, times, and levels and can be simply defined as the "promoters of the promoter" (Fig. 1.1). They can be tissue specific or regulate gene expression in multiple tissues. They often have modular expression patterns, and a gene that is active in many tissues is likely influenced by multiple enhancers (Pennacchio et al. 2006; Visel et al. 2009a). They can also regulate gene activity at various time points. They can regulate in *cis*, meaning that they regulate a gene on the same chromosomal region as they are located, or in *trans*, regulating a gene that is located on a different chromosome as was shown for the H olfactory receptor enhancer (Lomvardas et al. 2006). *Cis* enhancers can be 5′ or 3′ to the regulated gene, in introns or even within the coding exon of the gene they regulate (Neznanov et al. 1997; Tumpel et al. 2008). These enhancers can be near the promoter or very far away, with some like the Sonic hedgehog (*SHH*) limb enhancer being as far as ~1,000,000 bp away from the gene it regulates (Lettice et al. 2003). Enhancer function is generally considered to be independent of location or orientation relative to the gene they regulate.

A given enhancer can have an additional enhancer or enhancers with overlapping activity, called a *shadow enhancer/s* (Hong et al. 2008). The enhancer closest to the gene is usually considered to be the primary enhancer while other, more distant, enhancer or enhancers with similar activity are the shadow enhancer/s. Shadow enhancers are thought to protect the essential activities of the primary enhancer in adverse genetic conditions or environmental pressure (Hobert 2010). In addition, they can also have their own unique regulatory activities. The existence of shadow enhancers might also explain why in certain cases the removal of potentially important enhancers in the mouse genome can lead to no apparent phenotype (Ahituv et al. 2007; Cretekos et al. 2008).

Enhancers are thought to function through the recruitment of transcription factors and subsequent physical interactions with the gene promoter. This physical interaction is thought to be carried out through DNA looping. The DNA looping is mediated by transcription cofactors which activate *cohesin*, a protein that links DNA sequences to one another (Wood et al. 2010; Dorsett 2011). Cohesin, along with Nipped-B homolog (NIPBL), its DNA loading factor, allows the binding of the enhancer to the promoter. In humans, mutations in *NIPBL* as described in Chap. 11 of this book, entitled "Cohesin and Human Diseases," can lead to a range of gene regulatory defects that cause Cornelia de Lange syndrome (MIM #122470). Following enhancer-promoter binding, it is then thought that a conformational change takes place in Mediator, a multiprotein complex consisting of over 30 proteins that subsequently activates transcription (Kagey et al. 2010; Malik and Roeder 2010). In addition to looping, other mechanisms such as enhancer transcription and chromatin modifications over large regions that encompass both enhancer and promoter have also been suggested for enhancer function (Bulger and Groudine 2011; Ong and Corces 2011).

1.2.3 Silencers

Opposite to enhancers, *silencers* are thought to turn off gene expression at specific time points and locations (Fig. 1.1). Not much is known about silencers in humans, primarily due to the lack of a good functional assay to characterize them. Similar to enhancers, they are also thought to be orientation independent and can be located almost anywhere with regard to the genes that they regulate. Transcription factors bind to the silencer sequences and along with transcription cofactors they act to repress the expression of the gene. These "negative" transcription factors are thus called *repressors*, and their cofactors are termed *corepressors* (Privalsky 2004).

Several different mechanisms have been proposed for silencer function. Recent work suggests that, similar to enhancers, they can interact with the promoter through DNA looping (Lanzuolo et al. 2007; Tiwari et al. 2008). Repressor and corepressor proteins can silence a gene by competing for the binding to a specific promoter (Li et al. 2004; Harris et al. 2005). They can also influence the local chromatin region by establishing repressive chromatin marks (Srinivasan and Atchison 2004).

1.2.4 Insulators

Insulators, also called boundary elements, are DNA sequences that create *cis*-regulatory boundaries that prevent the regulatory elements of one gene from

affecting neighboring genes (Fig. 1.1). Based on their function, they can be generally divided into two types: (1) *enhancer-blocking insulators* that obstruct enhancers by inhibiting their interaction with promoters and (2) *barrier insulators* that prevent the spread of heterochromatin.

Enhancer-blocking insulators have three different proposed models of function: (a) *promoter decoy*, where the insulator recruits the transcription machinery away from the promoter by having it bind to the insulator instead (Geyer 1997); (b) *physical barrier*, where the insulator sequence acts as a physical barrier that blocks the enhancer signal from reaching the promoter (Kong et al. 1997; Ling et al. 2004; Zhao and Dean 2004); and (c) *loop domain*, where the insulator sequences form loops with each other or other DNA sequences that interfere with enhancer function (Blanton et al. 2003; Byrd and Corces 2003; Ameres et al. 2005). In addition to binding to each other or other DNA sequences, it is thought that insulators could also form insulation loops by tethering the chromatin to various nuclear structures, such as the nucleolus (Yusufzai et al. 2004) and the nuclear lamina (Guelen et al. 2008). For a more detailed review of enhancer-blocking insulators, see Bushey et al. (2008).

Barrier insulators insulate genomic regions against silencing by *heterochromatin*, a chromatin state caused by tightly packing the DNA into a repressed/silenced form. Heterochromatin is marked by high levels of methylation in H3K9 and H3K27 histone residues (see Sect. 1.3.3 for more details) and CpG methylation along with a general lack of acetylation (a mark for active regions). Barrier insulators are thought to disrupt the spread of heterochromatin in the region they reside. They generally do this by modifying the substrates used to generate heterochromatin. This is done by recruiting histone acetyltransferases (HATs) or ATP-dependent nucleosome-remodeling complexes (Oki et al. 2004) or by removing nucleosomes (Bi et al. 2004). For a more detailed review of barrier insulators, see Gaszner and Felsenfeld (2006).

Probably the most widely studied insulator-associated protein is the *CCCTC-binding factor* (CTCF), which got its name due to its ability to recognize three regularly spaced repeats of CCCTC. CTCF is a ubiquitously expressed protein that has 11 zinc fingers and uses different combinations of them to identify and bind different DNA sequences. Binding of CTCF to DNA can protect that region from methylation, hence its ability to insulate (Filippova 2008). CTCF-binding sites, which serve as potential insulator regions, have been mapped throughout the human genome in several cell lines (Barski et al. 2007; Kim et al. 2007; Heintzman et al. 2009; Ernst et al. 2011) using chromatin immunoprecipitation (ChIP; technique described in detail in Sect. 1.3.3). Analysis of the location of these sites in various cell lines found that they remain largely unchanged, suggesting that insulator activity remains more or less constant in these different cell lines (Heintzman et al. 2009). CTCF-binding sites were also shown to overlap with cohesin-binding sites, suggesting that these two proteins interact together to achieve insulator function. In addition, CTCF depletion was shown to affect the genomic distribution of cohesin, but not the opposite, suggesting that cohesin localization is governed by CTCF (Parelho et al. 2008; Wendt et al. 2008).

1.2.5 Locus Control Region

A *locus control region* (LCR) is a region where various gene regulatory elements are clustered together to regulate a certain gene or several genes. One example is the well-characterized beta-globin LCR, described in Chap. 2 of this book entitled "The Hemoglobin Regulatory Regions." This region is composed of various regulatory elements that together regulate the expression of different globin genes during human development, a process known as hemoglobin switching (Sankaran et al. 2010).

1.3 Techniques to Identify Gene Regulatory Elements

Several techniques have been developed in order to identify gene regulatory elements. Building on advancements in molecular biology, primarily DNA sequencing technologies, these techniques are continuously being refined and made cost-efficient. These advancements are currently allowing the field to move from a "one-by-one" or "region-by-region" approach to a whole-genomic one.

1.3.1 DNase I Hypersensitive Sites

DNase I is an endonuclease that cleaves both single- and double-stranded DNA. Active or "open" chromatin regions are associated with nucleosome-free regions and hence known to be more attainable to DNase I. These regions were thus termed *DNase I hypersensitive sites* (DHSs). In the 1970s, it was already recognized that chromatin regions that contain active genes are twice as sensitive to DNase I digestion as regions where genes are inactive (Weintraub and Groudine 1976). In addition to active genes, DHSs can identify all types of active regulatory elements such as promoters, enhancers, silencers, and insulators. DHSs do not reveal the regulatory function of the sequence or the identity of the transcription factors that could be binding to it, but they can show whether a certain region in a specific cell type or tissue is active. With advancements in molecular biology, DHSs can now be discovered on a genomic scale using techniques such as DNase-chip (Crawford et al. 2006) which uses microarrays to hybridize captured DNase I hypersensitive sequences and DNase-Seq (Song and Crawford 2010) that uses massively parallel sequencing.

1.3.2 Comparative Genomics

The increasing availability of genomic data from multiple vertebrate genomes facilitated the ability to carry out comparative genomic studies on the human genome. These genomic comparisons allowed for the identification of noncoding regions in the human genome that have been conserved throughout evolution, suggesting that

this conservation is due to an important function (Boffelli et al. 2004; Dermitzakis et al. 2005). One such function could be gene regulation. The use of comparative genomic tools has been extremely successful in identifying regulatory elements in the human genome (Thomas et al. 2003; Woolfe et al. 2005; Pennacchio et al. 2006; Birney et al. 2007; Visel et al. 2008). These comparisons generally look at sequence conservation between two evolutionary distant species or between multiple closely related species. In general, genomic comparisons between species that are more distantly related through evolution yield fewer conserved sequences, but the sequences that are shared between them are more likely to be functional.

1.3.3 Chromatin Immunoprecipitation (ChIP)

ChIP is rapidly becoming the most effective and widely used tool to map putative regulatory sequences on a genomic scale. In ChIP, DNA-binding proteins are cross-linked to the DNA, and an antibody that is specific to a protein of interest is used to pull down the protein along with the DNA sequences that it is bound to. Following reversal of the cross-links, these DNA sequences can be identified either through quantitative PCR (ChIP-qPCR), by binding to a DNA microarray (ChIP-chip), or using massively parallel DNA sequencing technologies (ChIP-Seq) (Johnson et al. 2007; Visel et al. 2009b). Among these techniques, ChIP-Seq is rapidly becoming the most commonly used tool to map putative regulatory sequences on a genomic scale. As described later for the various regulatory elements, carrying out ChIP-Seq for various chromatin marks is rapidly becoming the gold standard in the identification of potential gene regulatory elements.

The most commonly used chromatin marks are histone modifications (listed in detail in Table 1.1). Each 147 bp of DNA is wrapped around eight histone proteins that are composed of the following protein pairs: H2A, H2B, H3, and H4, and sealed off by histone H1, making what is termed a nucleosome. These histones have protruding tails that have various modifications such as methylation, acetylation, phosphorylation, ubiquination, and sumoylation. These modifications can determine the status of the chromatin. For example, open chromatin which is indicative of enhancer activity can be identified based on one methyl group on the fourth lysine of histone H3. This is abbreviated as H3K4me1. Closed chromatin which is indicative of silencing can be identified by three methyl groups on the lysine in position 27 of H3 and is abbreviated as H3K27me3. Antibodies developed against these modifications can allow researchers to carry out ChIP and determine the chromatin state and regulatory potential (active, silenced, etc.) of specific sequences.

1.3.4 Chromatin Conformation Capture (3C)

In order to properly understand gene regulation, we must view it in three-dimensional space. As mentioned above, chromatin loops have been shown to be one of the

Table 1.1 Gene regulatory marks used for ChIP

Element	State	Mark	Selected references
Promoter	Active	H2A.Z histone variant	Barski et al. (2007)
		H2BK5me1	Barski et al. (2007)
		H3K4me2	Bernstein et al. (2005), Barski et al. (2007), Birney et al. (2007), Heintzman et al. (2007), Ernst et al. (2011)
		H3K4me3	Bernstein et al. (2005), Barski et al. (2007), Guenther et al. (2007), Mikkelsen et al. (2007), Heintzman et al. (2009), Ernst et al. (2011)
		H3K9me1	Barski et al. (2007)
		H3K9ac	Bernstein et al. (2005), Guenther et al. (2007), Ernst et al. (2011)
		H3K14ac	Guenther et al. (2007)
		H3K27me1	Barski et al. (2007)
		H4K20me1	Barski et al. (2007)
	Repressed	H3K27me3	Barski et al. (2007)
		H3K79me3	Barski et al. (2007)
Enhancer		p300 (EP300)	Heintzman et al. (2009)
		CBP	Kim et al. (2010)
		H2A.Z	Barski et al. (2007), Ernst et al. (2011)
		H3K4me1	Birney et al. (2007), Heintzman et al. (2007), Heintzman et al. (2009)
		H3K4me2	Bernstein et al. (2005), Barski et al. (2007), Birney et al. (2007), Heintzman et al. (2007), Ernst et al. (2011)
		H3K27ac	Heintzman et al. (2009), Creyghton et al. (2010), Ernst et al. (2011), Rada-Iglesias et al. (2011)
Silenced regions		DNA methylation	Tiwari et al. (2008)
		H3K9me2	Barski et al. (2007)
		H3K9me3	Barski et al. (2007), Ernst et al. (2011)
		H3K27me2	Barski et al. (2007)
		H3K27me3	Barski et al. (2007), Mikkelsen et al. (2007), Ernst et al. (2011)
Insulator		CTCF	Barski et al. (2007), Kim et al. (2007), Heintzman et al. (2009), Ernst et al. (2011)

major mechanisms by which the various regulatory elements (promoters, enhancers, silencers, and insulators) regulate. To unravel the physical interactions of the various regulatory elements in the nucleus, *chromatin conformation capture* (3C) and several derivatives of this technique have been developed. They are primarily based on cross-linking DNA-binding proteins with DNA (similar to ChIP) so that both the

regulatory element and the other region of DNA with which it interacts are bound together, bridged by the proteins that facilitate this interaction. The DNA is then cut randomly with restriction enzymes and ligated in conditions where the segments of DNA bridged by the protein cross-linked bundle will preferentially ligate to one another rather than to random free DNA. These newly ligated DNA segments are then analyzed in order to identify which regions of DNA have been joined, implying that they physically interact. The subsequent analysis of these sequences is what determines whether this technique is known as 3C, 4C, or 5C. 3C uses real-time PCR with primers matching two specific candidate regions in order to determine whether they interact with one another (Vassetzky et al. 2009). 4C generates circular DNA molecules following ligation, and the PCR primers are used in order to determine the identity of the DNA sequences that interacts with a specific sequence by having them faced outward on opposite ends of the restriction enzyme fragment (Vassetzky et al. 2009). For example, this technique can be used to determine the identity of the DNA sequences/regulatory elements that bind to a specific promoter by designing primers specific to that promoter. 5C can detect numerous chromatin interactions at the same time by using several primers (van Berkum and Dekker 2009). With the advent of massively parallel sequencing technologies, whole-genome adaptations of this technique have been introduced such as Hi-C (Lieberman-Aiden et al. 2009) and ChIA-PET (Fullwood et al. 2009).

1.4 Techniques to Functionally Characterize Gene Regulatory Elements

1.4.1 Promoters

The standard promoter assay is straightforward. The candidate promoter sequence is placed in front of a reporter gene (Fig. 1.2a), which is used as an indicator of where the promoter is active by inserting this construct into a cell culture or animal model. Various reporter genes are used for this assay, usually depending on the context of the assay. In cell culture, the most commonly used reporter gene is luciferase, in zebrafish fluorescent proteins such as GFP or mCherry, and in mice the frequently used reporter gene is LacZ. The DNA sequence that is typically assayed for promoter activity covers at least 250–500 bp upstream of the transcription start site, so as to include the proximal promoter and other potential promoter-associated sequences.

1.4.2 Enhancers

Enhancers can be functionally characterized by various methods: deletion series, enhancer traps, cell culture enhancer assays, and in vivo electroporation or transgenic

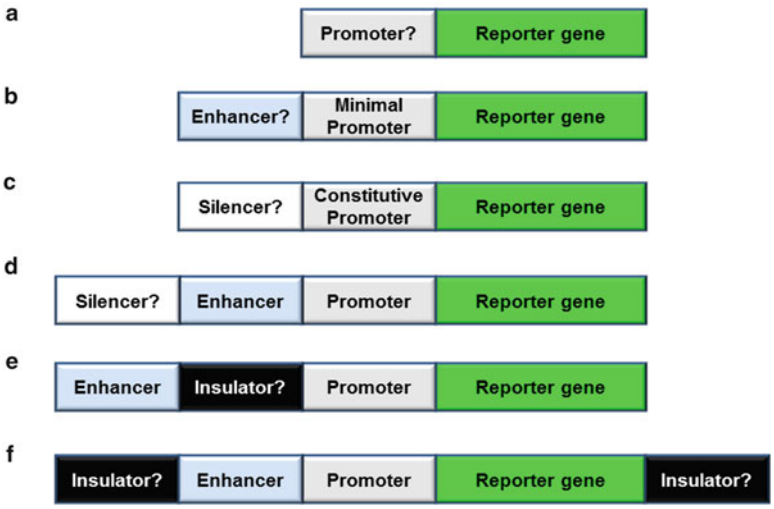

Fig. 1.2 *DNA construct designs that are commonly used to functionally characterize gene regulatory elements.* (**a**) A DNA sequence assayed for promoter activity is placed in front of a reporter gene. (**b**) A DNA sequence assayed for enhancer activity is placed in front of a minimal promoter (a promoter that is not sufficient to drive expression without the presence of a functional enhancer) and a reporter gene. (**c**) A DNA sequence assayed for silencer activity is placed in front of a constitutive promoter (a promoter that is always active) and a reporter gene or in front of a characterized enhancer and promoter followed by a reporter gene (**d**). (**e**) DNA sequences can be assayed for insulator activity by blocking the ability of a characterized enhancer to turn on a reporter gene via a characterized promoter. They can also be assayed for barrier insulator activity by placing them on both sides of a known enhancer and promoter that drive reporter gene expression and measuring the activity of this reporter gene for consistency and length of expression following stable integration (**f**)

enhancer assays. *Deletion series* uses a long DNA construct, such as a yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC), that encompasses the genomic region harboring a potential enhancer or enhancers and some kind of reporter gene that is inserted into this DNA fragment. These are then injected into an organism in order to observe the reporter gene expression pattern. Once an expression pattern of interest is observed, the YAC or BAC can be modified such that certain sequences are deleted and these new constructs are reinjected to compare for changes in reporter gene expression due to this manipulation. An analogous approach is the use of overlapping YACs or BACs that also encompass a reporter gene. Each construct is injected separately and observed for its respective reporter gene expression, allowing the ability to locate tissue-specific enhancers based on their shared genomic region (Mortlock et al. 2003). These overlapping constructs could also be deleted, as mentioned above, to further pinpoint the enhancer location. Combined, these assays are able to broadly identify the location of the enhancer/s, but usually fall short of determining its exact sequence.

The majority of enhancer assays are based on a common design: the assayed sequence is placed in front of a minimal promoter (a promoter that is not sufficient to

drive expression without the presence of a functional enhancer), followed by a reporter gene (Fig. 1.2b), similar to reporters used for promoter assays. If the assayed sequence is an enhancer, it will turn on the minimal promoter which in turn will drive the reporter gene expression. *Enhancer traps* use the random genomic integration of a DNA sequence containing a minimal promoter and the reporter gene to uncover enhancers around the region of integration (Parinov et al. 2004; Korzh 2007). Similar to the deletion series, the disadvantage of this method is that it does not pinpoint the exact location of the enhancer.

To assay specific sequences for enhancer activity, candidate sequences are typically cloned into a vector containing a minimal promoter and a reporter gene as outlined above. This vector is then introduced into a model system via different methods. *Cell culture enhancer assays* usually use transfection, electroporation, or virus integration to insert constructs into cells and assay for enhancer activity using luciferase as the reporter gene. The advantage of cell culture-based enhancer assays is that they are relatively cheap, can be done in a high-throughput manner, and can be quantitative. However, they are carried out in cell lines which can lose a lot of their tissue characteristics; they do not provide "whole-organism" properties, and the results can be highly variable due to factors such as different DNA preparation methods, number of cell passages, cell culture conditions, and others.

In vivo enhancer assays provide the advantage of being able to test the assayed sequence in the context of the whole organism. Due to the length of time and technical aspects of observing adult reporter gene expression in vertebrates, the majority of in vivo-based enhancer assays are carried out at developmental time points. Zebrafish transgenic enhancer assays often use the Tol2 transposon for genomic integration along with a fluorescent reporter gene for visualization (Fisher et al. 2006). In frogs, transgenic enhancer assays can be carried out by using standard sperm nuclear transplantation and can utilize transposons (Khokha and Loots 2005). Chicken enhancer assays are usually based on electroporation of the assayed construct into the embryonic tissue of interest, followed by visualization using fluorescence or LacZ (Uchikawa 2008). Mouse transgenic enhancer assays tend to use standard mouse transgenic techniques (Nagy et al. 2002) and LacZ as the reporter gene. These mouse assays are usually transient (embryos are removed at a certain time point for reporter gene detection), avoiding the costs associated with maintaining mouse lines (Pennacchio et al. 2006). The main caveat for the majority of these assays is that they are not able to quantitatively measure enhancer activity because there can be multiple integrations of the enhancer construct per animal, leading to variation in the reporter intensity.

1.4.3 Silencers

Silencers can be detected by placing the sequence to be assayed in front of a constitutive promoter (a promoter that is always active) and a reporter gene and comparing the activity of that reporter gene with and without the assayed sequence

(Fig. 1.2c; Petrykowska et al. 2008). If the sequence is a silencer, lower reporter gene expression levels due to the assayed sequence silencing the constitutive promoter would be observed. As previously described, silencers can also interfere with the binding of a nearby activating site. In order to detect these kinds of silencers, an enhancer blocker assay was developed where the assayed sequence is placed in front of a characterized enhancer, promoter, and a reporter gene (Fig. 1.2d; Petrykowska et al. 2008). If the sequence silences by enhancer blocking, there should be a reduction in reporter gene activity in this assay. Both these assays can work well in cell culture, using a luciferase reporter, because reporter expression can be quantified. However, these techniques are not straightforward for in vivo silencer assays. In both zebrafish and mouse transgenic assays mentioned above, there is a high degree of variability between embryos in the number of inserted transgenes. This variability does not allow quantitative measurements of reporter gene expression differences and has hampered the development of in vivo silencer assays.

1.4.4 Insulators

As mentioned previously, insulators can be divided into two types: enhancer blockers that obstruct enhancers by inhibiting their interaction with promoters and barrier insulators that prevent the spread of heterochromatin. To assay for enhancer blockers, the sequence being analyzed is placed in between a characterized enhancer and promoter, which is followed by a reporter gene (Fig. 1.2e). If the sequence is an enhancer blocker insulator, it should block the ability of the enhancer to activate the promoter and thus lead to reduced reporter gene expression versus a vector that does not contain this sequence. Barrier insulator assays consist of placing an enhancer, promoter, and a reporter gene in between two copies of the presumed barrier insulator (Fig. 1.2f) and having it stably integrate into the genome. The reporter gene expression is then monitored for consistency and length of expression. A sequence will be considered a barrier insulator if it provides for consistent reporter gene expression over a certain period of time (Recillas-Targa et al. 2002; Gaszner and Felsenfeld 2006).

1.5 Summary

We are in the midst of revolutionary times, with novel DNA sequencing technologies increasing our ability to sequence DNA at enormous rates. Due to these technological advances, individual whole-genome sequences are readily available at an affordable price. This availability will have the most immediate impact on two major fields: predicting human disease risk and pharmacogenomics. The pioneering work described in this book that led to the detection of various human disease-causing regulatory mutations and the techniques described in this chapter will

provide us with a starting foundation to analyze gene regulatory mutations in whole-genome data sets. As these whole-genome data sets expand and high-throughput functional gene regulatory assays develop, we will gain an increased understanding of how gene regulatory mutations lead to human phenotypes.

Abbreviations

BRE	TFIIB recognition element
Inr	Initiator element
DPE	Downstream core promoter element
TBP	TATA-binding protein
TFIID	Transcription factor IID
TFIIB	Transcription factor IIB
TSS	Transcription start site
TFBS	Transcription factor binding sites
SHH	Sonic hedgehog
NIPBL	Nipped-B homolog
CTCF	CCCTC-binding factor
LCR	Locus control region
DHSs	DNase I hypersensitive sites
ChIP	Chromatin immunoprecipitation
3C	Chromatin conformation capture
YAC	Yeast artificial chromosome
BAC	Bacterial artificial chromosome

References

Ahituv N, Zhu Y et al (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5(9):e234

Ameres SL, Drueppel L et al (2005) Inducible DNA-loop formation blocks transcriptional activation by an SV40 enhancer. EMBO J 24(2):358–367

Barski A, Cuddapah S et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823–837

Bernstein BE, Kamal M et al (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120(2):169–181

Bi X, Yu Q et al (2004) Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. Mol Cell Biol 24(5):2118–2131

Birney E, Stamatoyannopoulos JA et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799–816

Blanton J, Gaszner M et al (2003) Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev 17(5):664–675

Boffelli D, Nobrega MA et al (2004) Comparative genomics at the vertebrate extremes. Nat Rev Genet 5(6):456–465

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144(3):327–339

Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev 11(22): 3020–3031
Bushey AM, Dorman ER et al (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell 32(1):1–9
Butler JE, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 16(20):2583–2592
Byrd K, Corces VG (2003) Visualization of chromatin domains created by the gypsy insulator of Drosophila. J Cell Biol 162(4):565–574
Calhoun VC, Stathopoulos A et al (2002) Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proc Natl Acad Sci USA 99(14):9243–9247
Carroll SB (2005) Evolution at two levels: on genes and form. PLoS Biol 3(7):e245
Clamp M, Fry B et al (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proc Natl Acad Sci USA 104(49):19428–19433
Crawford GE, Davis S et al (2006) DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods 3(7):503–509
Cretekos CJ, Wang Y et al (2008) Regulatory divergence modifies limb length between mammals. Genes Dev 22(2):141–151
Creyghton MP, Cheng AW et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107(50):21931–21936
Dermitzakis ET, Reymond A et al (2005) Conserved non-genic sequences – an unexpected feature of mammalian genomes. Nat Rev Genet 6(2):151–157
Dorsett D (2011) Cohesin: genomic insights into controlling gene transcription and development. Curr Opin Genet Dev 21(2):199–206
Ernst J, Kheradpour P et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43–49
Filippova GN (2008) Genetics and epigenetics of the multifunctional protein CTCF. Curr Top Dev Biol 80:337–360
Fisher S, Grice EA et al (2006) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc 1(3):1297–1305
Fullwood MJ, Liu MH et al (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462(7269):58–64
Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196(2):261–282
Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7(9):703–713
Geyer PK (1997) The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 7(2):242–248
Guelen L, Pagie L et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453(7197):948–951
Guenther MG, Levine SS et al (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130(1):77–88
Harris MB, Mostecki J et al (2005) Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J Biol Chem 280(13):13114–13121
Heintzman ND, Stuart RK et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39(3):311–318
Heintzman ND, Hon GC et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243):108–112
Hobert O (2010) Gene regulation: enhancers stepping out of the shadow. Curr Biol 20(17):R697–R699
Hong JW, Hendrix DA et al (2008) Shadow enhancers as a source of evolutionary novelty. Science 321(5894):1314
Johnson DS, Mortazavi A et al (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830):1497–1502

Kagey MH, Newman JJ et al (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314):430–435
Khokha MK, Loots GG (2005) Strategies for characterising cis-regulatory elements in Xenopus. Brief Funct Genomic Proteom 4(1):58–68
Kim TH, Abdullaev ZK et al (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128(6):1231–1245
Kim TK, Hemberg M et al (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182–187
Kong S, Bohl D et al (1997) Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. Mol Cell Biol 17(7):3955–3965
Korzh V (2007) Transposons as tools for enhancer trap screens in vertebrates. Genome Biol 8(Suppl 1):S8
Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol 20(13):4754–4764
Lagrange T, Kapanidis AN et al (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12(1):34–44
Lanzuolo C, Roure V et al (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9(10):1167–1174
Lettice LA, Heaney SJ et al (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12(14):1725–1735
Li L, He S et al (2004) Gene regulation by Sp1 and Sp3. Biochem Cell Biol 82(4):460–471
Lieberman-Aiden E, van Berkum NL et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293
Ling J, Ainol L et al (2004) HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. J Biol Chem 279(49):51704–51713
Lomvardas S, Barnea G et al (2006) Interchromosomal interactions and olfactory receptor choice. Cell 126(2):403–413
Malik S, Roeder RG (2010) The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11(11):761–772
McLean CY, Reno PL et al (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471(7337):216–219
Mikkelsen TS, Ku M et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448(7153):553–560
Mortlock DP, Guenther C et al (2003) A general approach for identifying distant regulatory elements applied to the Gdf6 gene. Genome Res 13(9):2069–2081
Nagy A, Gertsenstein M et al (2002) Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor, New York
Neznanov N, Umezawa A et al (1997) A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice. J Biol Chem 272(44):27549–27557
Oki M, Valenzuela L et al (2004) Barrier proteins remodel and modify chromatin to restrict silenced domains. Mol Cell Biol 24(5):1956–1967
Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12(4):283–293
Parelho V, Hadjur S et al (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132(3):422–433
Parinov S, Kondrichin I et al (2004) Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. Dev Dyn 231(2):449–459
Pennacchio LA, Ahituv N et al (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118):499–502
Petrykowska HM, Vockley CM et al (2008) Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome Res 18(8):1238–1246

Prabhakar S, Visel A et al (2008) Human-specific gain of function in a developmental enhancer. Science 321(5894):1346–1350
Privalsky ML (2004) The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annu Rev Physiol 66:315–360
Rada-Iglesias A, Bajpai R et al (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470(7333):279–283
Recillas-Targa F, Pikaart MJ et al (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci USA 99(10):6883–6888
Riethoven JJ (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. Methods 674:33–42
Sankaran VG, Xu J et al (2010) Advances in the understanding of haemoglobin switching. Br J Haematol 149(2):181–194
Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72:449–479
Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold 2010(2):pdb.prot5384
Srinivasan L, Atchison ML (2004) YY1 DNA binding and PcG recruitment requires CtBP. Genes Dev 18(21):2596–2601
Suzuki Y, Tsunoda T et al (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res 11(5):677–684
Thomas JW, Touchman JW et al (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424(6950):788–793
Tiwari VK, McGarvey KM et al (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol 6(12):2911–2927
Tumpel S, Cambronero F et al (2008) A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. Proc Natl Acad Sci USA 105(51):20077–20082, Epub 2008 Dec 22
Uchikawa M (2008) Enhancer analysis by chicken embryo electroporation with aid of genome comparison. Dev Growth Differ 50(6):467–474
van Berkum NL, Dekker J (2009) Determining spatial chromatin organization of large genomic regions using 5C technology. Methods Mol Biol 567:189–213
Vassetzky Y, Gavrilov A et al (2009) Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. Methods Mol Biol 567:171–188
Visel A, Prabhakar S et al (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 6:6
Visel A, Akiyama JA et al (2009a) Functional autonomy of distant-acting human enhancers. Genomics 93(6):509–513
Visel A, Blow MJ et al (2009b) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–858
Weintraub H, Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. Science 193(4256):848–856
Wendt KS, Yoshida K et al (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451(7180):796–801
Wood AJ, Severson AF et al (2010) Condensin and cohesin complexity: the expanding repertoire of functions. Nat Rev Genet 11(6):391–404
Woolfe A, Goodson M et al (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3(1):e7
Yusufzai TM, Tagami H et al (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell 13(2):291–298
Zhao H, Dean A (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res 32(16): 4903–4919, Print 2004

Chapter 2
The Hemoglobin Regulatory Regions

Betty S. Pace and Levi H. Makala

Abstract All animals that use hemoglobin for oxygen transport synthesize different hemoglobin types during the various stages of development. In humans, two gene clusters direct the production of hemoglobin including the α-locus which contains the embryonic ζ gene and two adult α genes on chromosome 16. A second cluster, the β-globin locus located on chromosome 11, contains the ε, $^{G}\gamma$, $^{A}\gamma$, δ, and β genes. The globin genes are arranged from 5′ to 3′ according to the order of their expression and are developmentally regulated to produce different hemoglobin species during ontogeny. Two switches in the type of hemoglobin synthesized during development occur, a process known as hemoglobin switching. Through research efforts over the last two decades, several insights have been gained into the molecular mechanisms of hemoglobin switching. However, the entire process has not been fully elucidated. Studies of naturally occurring globin gene promoter mutations and transgenic mouse investigations have contributed to our understanding of the effect of DNA mutations on globin gene expression. Furthermore, the developmental regulation of globin gene expression has shaped research efforts to establish therapeutic modalities for individuals affected with sickle cell disease and β-thalassemia. Here, we will review the progress made toward understanding molecular mechanisms that control globin gene expression and the consequences of mutations on hemoglobin switching.

Keywords Hemoglobin • ε-Globin • γ-Globin • β-Globin • α-Globin • Hereditary persistence of fetal hemoglobin • Hemoglobin switching • Thalassemia • Sickle cell disease

B.S. Pace (✉) • L.H. Makala
Georgia's Health Sciences University, 1120 15th Street, CN-4112, Augusta, GA 30912, USA
e-mail: bpace@georgiahealth.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_2,

2.1 Introduction

2.1.1 Hemoglobin Proteins

Hemoglobin is a 64.4-kDa tetramer iron-containing oxygen-transport protein, composed of two pairs of α-like and β-like globin polypeptide chains. A heme group, ferroprotoporphyrin IX, is linked covalently at a specific site to each chain. In humans, the α-locus on chromosome 16 encodes the embryonic ζ-globin and adult α1- and α2-globin genes (Fig. 2.1); the β-locus on chromosome 11 encodes the functional ε-, ^{G}γ-, ^{A}γ-, δ-, and β-globin genes, expressed sequentially from 5′ to 3′ in a tissue- and developmental-specific manner during ontogeny (Stamatoyannopoulos and Grosveld 2001). Expression of the α-globin genes is controlled by the HS-40 enhancer element located 40 Kb upstream of ζ-globin and the β-locus by the locus control region (LCR) positioned 6–20 kb upstream of ε-globin. Two major switches in the type of hemoglobin synthesized during development occur as a result of changes in globin gene expression in the β-locus (Fig. 2.2; Wood et al. 1985): (1) from ε- to γ-globin expression around 6 weeks of gestation and (2) from γ- to β-globin expression before birth producing embryonic, fetal, and adult hemoglobin, respectively. The developmental changes in globin gene expression are collectively called hemoglobin switching.

The globin genes are relatively small genes comprising three coding exons and two introns. The exons code for 141 and 146 amino acids in the α- and β-like globin chains, respectively. To achieve switching, the α and β cluster are expressed in a coordinated fashion to produce different hemoglobin types during development. In primitive erythroblasts in the yolk sac, embryonic hemoglobin is produced during the first 8–10 weeks after conception (Stamatoyannopoulos and Grosveld 2001);

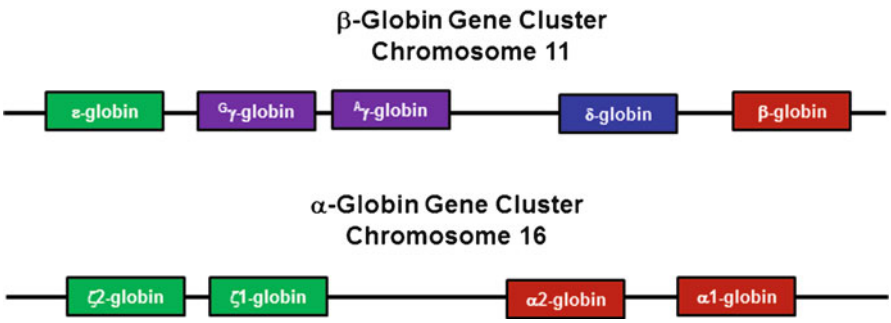

Fig. 2.1 *β-like and α-like globin loci on chromosome 11 and 16.* Shown are the functional human hemoglobin genes located on chromosome 11 (β-like globin gene locus) and chromosome 16 (α-like globin gene locus). Both gene loci are expressed in a coordinated manner to produce the various types of hemoglobin synthesized during the different stages of development

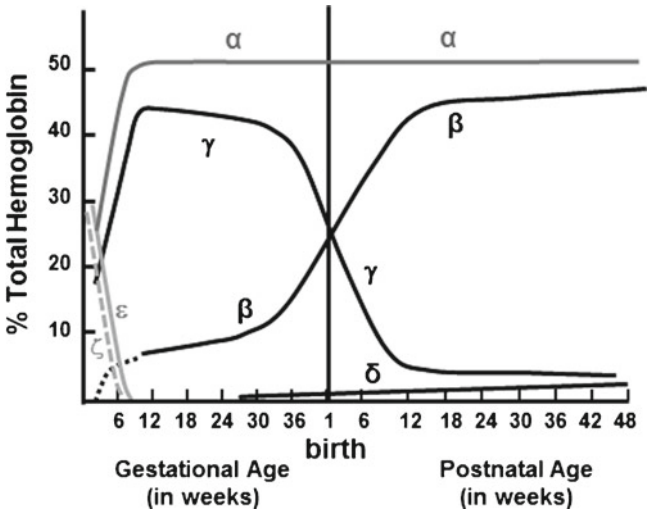

Fig. 2.2 *The developmental switch in globin gene expression.* Embryonic globins are produced in the first few weeks of in utero development. The first switch in the type of hemoglobin produced occurs at 6 weeks, where ε-globin is silenced and fetal hemoglobin synthesis increases (γ-globin), and predominates for the remainder of gestation. At birth, a second switch occurs from fetal to adult hemoglobin A (β-globin) and low-level hemoglobin A_2 (δ-globin) production is constant for the remainder of life (Reproduced with modifications by permission from Stamatoyannopoulos and Grosveld 2001)

it is composed of embryonic ζ-globin and ∈-globin chains ($\zeta_2\in_2$; Fig. 2.2). The first hemoglobin switching event occurs as ζ- and ∈-globin expression is silenced and α- and γ-globin synthesis begins, leading to the production of Hb F ($\alpha_2\gamma_2$). Simultaneous with the switch, the site of erythropoiesis transitions from the yolk sac to the fetal liver and spleen. The second switch in hemoglobin production occurs when γ-globin gene expression is silenced and increased synthesis of the adult forms of hemoglobin ($\alpha_2\delta_2$ and $\alpha_2\beta_2$) is observed in the bone marrow.

At birth, Hb F comprises 80–90% of the total hemoglobin synthesized, and it gradually decreases to ~1% by 10 months in normal infants (Maier-Redelsperger et al. 1994). Hb F is a heterogeneous mixture of γ-globin polypeptide chains containing either glycine ($^{G}\gamma$) or alanine ($^{A}\gamma$) at residue 136. The oxygen affinity of Hb F is greater than Hb A, but the former functions well to maintain normal tissue oxygenation. At birth, $^{G}\gamma$-chains predominate; however, a switch to Hb F consisting predominantly of $^{A}\gamma$-chains arises during the first 10 months as well. Furthermore, as Hb F levels decline, it becomes restricted to a subset of erythrocytes termed F-cells and is distributed in a heterocellular pattern. Family studies show that F-cell numbers are genetically controlled; however, the genes involved in this process are poorly understood (Wood 1993).

2.1.2 Developmental Regulation of Globin Gene Expression

The β-locus is developmentally regulated by a region 5–25 kb upstream of ε-globin known as the locus control region (LCR). The LCR consists of five developmentally stable DNase I hypersensitive sites (HSs) of which HS1 to HS4 are erythroid specific (Kollias et al. 1986; Grosveld et al. 1987; Forrester et al. 1987). Two other sites, HS6 and HS7, located 6 and 12 kb, respectively, upstream of HS5 have also been described (Bulger et al. 1999), but it is unclear whether they constitute components of the β-globin LCR. In addition, an erythroid-specific HS localized 40 kb upstream of ζ-globin required for α-like globin gene expression was identified (Higgs et al. 1990).

Hemoglobin switching is thought to be mediated by the LCR through the competition of various developmental stage-specific transcription factors that mediate interaction with the individual globin gene promoters (Enver et al. 1990; Townes and Behringer 1990; Strouboulis et al. 1992). The tissue- and developmental-specific expression pattern of the individual globin genes is achieved through the action of transcription factors on regulatory sequences in the immediate flanking region of individual genes and more distal sequences that regulate the entire locus such as the LCR. The most widely accepted model of LCR function is based on looping of intervening sequence between globin gene promoters and the LCR to form active transcription complexes (Fraser and Grosveld 1998).

Each LCR HS contains a 200–500-bp core region of DNase I hypersensitivity; however, only HS2 acts as a classical enhancer element (Tuan et al. 1989; Ney et al. 1990). In addition to other ubiquitous DNA-binding proteins, each HS has one or more binding motif for two hematopoietic-restricted proteins: GATA-binding protein 1 (GATA1) and nuclear factor erythroid-2 (NFE2) (Martin and Orkin 1990; Ney et al. 1990; Goodwin et al. 2001). GATA1 consensus binding sites are present both in globin regulatory elements that activate or silence gene expression. Friend of GATA (FOG; ZFPM1) is co-expressed and interacts with GATA1 to promote erythroid and megakaryocytic differentiation (Tsang et al. 1997). The combination of GATA1 sites and a GGTGG motif occurs a number of times in the β-locus and appears to be associated with erythroid specificity. Of the HSs in the LCR, HS3 is believed to be involved in γ-globin activation during fetal-stage development (Ellis et al. 1996). HS3 is bound by erythroid Kruppel-like factor (EKLF; KLF1), an erythroid-specific member of the Sp1 family of Kruppel-like zinc finger proteins that binds the β-globin promoter CACCC box to facilitate adult Hb A synthesis (Miller and Bieker 1993). EKLF is an active factor at the CACCC element in vivo and is thought to induce changes in chromatin structure required to accomplish β-globin activation (Miller and Bieker 1993). It is postulated that EKLF bound to HS3 may provide a competitive advantage for LCR-β-globin promoter interactions over γ-globin to facilitate hemoglobin switching after birth (Jackson et al. 2003).

2.1.3 ε-Globin Gene Regulation

The ε-globin gene is normally expressed in the embryonic yolk sac. Two ε chains combined with two ζ chains constitute what is called the embryonic hemoglobin Gower I. Two ε chains combined with two α chains form the embryonic hemoglobin Gower II. Both of these embryonic hemoglobins are replaced by Hb F and Hb A hemoglobin later in development. Very little is known about the activation of ε-globin in the embryonic stage of development. More is known about the mechanism whereby the ε-globin gene is silenced when definitive stage erythropoiesis starts in the fetal liver. A silencer transcriptional control element has been reported between −182 and −467 bp 5′ of the canonical ε-globin gene cap site (GenBank accession #NG_000007.3; GI:28380636). Deletion of this region produced continued ε-gene expression in adult life (Raich et al. 1992). It was later shown that the GATA1 site at −208, YY1 site at −269, and the CACCC site at −379 (GenBank accession #NG_000007.3; GI:28380636) are presumably bound by Sp1 and play a major role in normal ε-globin silencing during development (Gong et al. 1991; Yu et al. 1991; Peters et al. 1993; Raich et al. 1995).

2.1.4 γ-Globin Gene Regulation

Extensive research has been conducted to understand γ-globin gene regulation because it provides a rational basis for molecular strategies to induce Hb F after birth which is of therapeutic value in the treatment of sickle cell disease and β-thalassemia. Each γ-gene contains a canonical TATA box, a duplicated CAAT box, and a single CACCC box (Fig. 2.3a; Stamatoyannopoulos and Grosveld 2001). A large number of proteins have been identified and are capable of binding the 200-bp region relative to the cap site of the γ-globin promoters (Fig. 2.3a). In theory, therapeutic γ-globin reactivation could be accomplished by inhibition of repressor proteins to prevent silencing or enforced expression of *trans*-activators. Pace and associates investigated a proximal signal transducer and activator of transcription 3 (STAT3) binding site in the 5′-untranslated region of γ-globin (Ferry et al. 1997) as a potential silencer. They demonstrated γ-gene silencing by interleukin-6, a known activator of STAT3 signaling. In subsequent investigations, a repressor role for the dominant-negative STAT3β isoform was established through its binding at the putative binding site 5′TTCTGGAA-3′ located between nucleotides +9 to +16 in the γ-globin 5′-untranslated region (Foley et al. 2002).

Another important regulatory element is the stage selector element (SSE; Fig. 2.3a) located between −53 and −34 of the γ-globin promoter (GenBank accession #NG_000007.3; GI:28380636). The SSE is a sequence that was found to have an in vitro role as a fetal stage-specific site involved in the γ-to β-globin gene switch at birth (Jane et al. 1993). When bound by the stage selector protein (SSP),

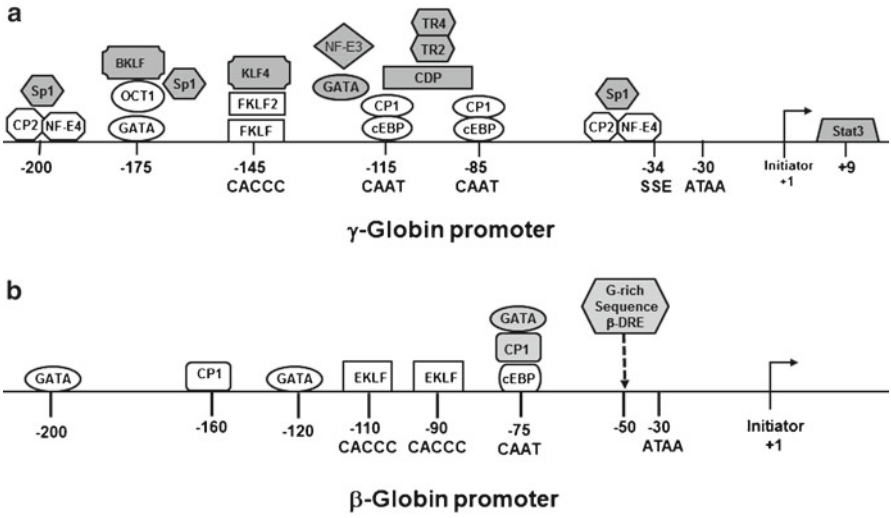

Fig. 2.3 *Transcription factor binding in the γ-globin and β-globin promoter regions.* (**a**) Shown are the different DNA-binding proteins that have been demonstrated to bind to the minimal γ-globin promoter either as monomers, homodimers, or heterodimers. Also shown are ubiquitous- and hematopoietic-specific transcription factors. NF-E4, nuclear factor erythroid 4; BKLF, basic Kruppel-like factor; OCT1, octamer 1; FKLF, fetal Kruppel-like factor; CDP, CCAAT displacement protein; cEBP, CCAAT enhancer-binding protein; Stat3, signal transducer and activators of transcription 3. (**b**) The β-globin gene promoter, showing the position of conserved boxes and the binding motifs for functionally important transcriptional activators. The G-rich sequence is the βDRE. The CAAT box binds CP1, GATA-1, and NF-E6 or DSF (cEBP). The promoters are not drawn to scale

a competitive advantage for γ-globin expression occurred in vitro. The SSP is a heterodimer composed of ubiquitous transcription factor CP2 (TFCP2) (Jane et al. 1995) and the erythroid-specific *trans*-activator NFE4, known to alter histone acetylation (Zhao et al. 2004). The binding of SSP and SP1 to the SSE is mutually exclusive; however, the level of methylation at the CpG dinucleotides at 55 and 50 influences their binding affinity (Jane et al. 1993). When the γ-promoter is hypomethylated in vitro, SSP binds to the SSE at the expense of Spl; by contrast, methylation of the CpG residues within the SSE enhances Spl binding to the exclusion of SSP ensuring the γ-gene is not reactivated. These studies mimic the normal changes in methylation status that occur in the CpG sites in the γ-promoter during development to facilitate hemoglobin switching (Enver et al. 1988). The importance of CpG methylation in gene silencing (Mavilio et al. 1983) has been shown in studies with hypomethylating agents such as 5-azacytidine and decitabine to reactivate γ-gene expression. However, mutations of the SSE in the presence of a competing β-globin gene had no effect on β-gene expression in transgenic mice (Ristaldi et al. 2001), thus failing to establish a functional role for the SSE in vivo. Further upstream to the

γ-globin promoter are the duplicated CCAAT boxes at −85 and −115 (Fig. 2.3a) (Mantovani 1998, 1999). These CCAAT boxes have been studied extensively since point mutations in this region lead to continued γ-globin expression in adults and a group of disorders known as hereditary persistence of fetal hemoglobin (HPFH). These boxes were shown to be recognized by several transcription factors including NF-Y/CBF (Mantovani 1998), C/EBP (Osada et al. 1996), and CCAAT displacement protein (CDP; Barberis et al. 1987; Superti-Furga et al. 1988; Neufeld et al. 1992; Aufiero et al. 1994). CP1 (CCAAT-binding factor) is a ubiquitous protein heterodimer that binds both CCAAT boxes as a positive regulator (Fucharoen et al. 1990), although in vivo data does not exists to substantiate this role. By contrast, CP1 binding is inhibited by CDP which binds to the region surrounding the CCAAT box. Studies have shown derepression of gene expression when the CDP-/CP1-binding site is deleted which suggests, in addition to blocking the interaction of CP1, that CDP may also repress transcription through a distinct *cis* element (Skalnik et al. 1991). A novel C/T mutation in the distal CCAAT motif of ^{G}γ-globin abolishes the binding of CP1, producing HPFH (Fucharoen et al. 1990). C/EBPs are ubiquitously expressed *trans*-activators that bind to the CCAAT box in a competitive manner with CDP as well.

CDP is a divergent homeodomain protein that is highly conserved through evolution and has properties of a potent transcriptional repressor. CDP is associated with histone deacetylase activity and a corepressor complex through interactions with histone deacetylases (Li et al. 1999), such as CREB-binding protein (CREBBP) and p300-/CREB-binding protein-associated factor (PCAF) (Li et al. 2000). CDP has also been reported to interact with a histone lysine methyltransferase, euchromatic histone-lysine N-methyltransferase 2 (*EHMT2*; *G9a*), in vivo and in vitro. The transcriptional repressor function of CDP is mediated through the activity of G9a. This is caused by increased methylation of histone H3 Lys-9 in the CDP regulatory region of the p21$^{waf1/cdi1}$ promoter (Nishio and Walsh 2004). These results indicate that G9a functions as a transcriptional corepressor within a CDP complex.

Other proteins such as Kruppel-like factor 4 (KLF4) have also been demonstrated to play a role in γ-globin gene regulation by several groups. In zebrafish, *KLF4* was shown to be essential for primitive erythropoiesis (Gardiner et al. 2005, 2007). Further evidence of the functional significance of this factor in erythropoiesis came from a very recent study where KLF4 positively regulated human β-globin gene expression (Marini et al. 2010). KLFs have also been implicated in γ-globin regulation (Asano et al. 1999, 2000; Zhang et al. 2005), although the precise mechanism of a KLF-CACCC-mediated regulation remains to be elucidated. Recently, Kalra et al. (2011) published data that support direct binding of KLF4 to the γ-gobin CACCC box and a model of antagonistic interaction between KLF4- and CREB-binding protein in γ-globin gene regulation.

Other groups have carried out studies to define repressors of γ-globin expression. Engels and associates identified the direct repeat erythroid-definitive (DRED) repressor complex that binds the DR1 motif located near the distal CAAT box (Fig. 2.3a; Tanabe et al. 2002). The complex is composed of TR2 and TR4, two

nuclear orphan receptors believed to bind DNA and recruit other repressors to achieve γ-gene silencing; an in vivo role for DRED has yet to be established. Other studies showed that a deletion of the upstream CACCC box inhibited γ-gene expression in definitive erythroid cells (Stamatoyannopoulos et al. 1993); Sp1 and basic Kruppel-like factor (KLF3) bind the CACCC box (Crossley et al. 1996); however, the latter does not affect γ-gene expression. Fetal Kruppel-like factor (FKLF) and FKLF-2 were also shown to bind the CACCC region (Fig. 2.3a), but their role in γ-gene expression in vivo has yet to be determined (Asano et al. 1999, 2000). Recently, B-cell CLL/lymphoma 11A (BCL11A) was reported as a major factor involved in switching through its ability to repress γ-globin expression (Sankaran et al. 2008, 2009). Chen and associates provided evidence that BCL11A binds the GGCCGG motif between nucleotide −56 and −51 on the γ-globin promoter (Chen et al. 2009). Subsequently, EKLF was shown to activate BCL11A gene expression resulting in upregulation of β-globin and repression of the γ-globin genes, thereby contributing to the process of hemoglobin switching (Borg et al. 2010; Zhou et al. 2010). Additional research will be required to ferret out the role of various *cis*-regulatory elements in developmentally regulated γ-gene expression. This research can lead to improved strategies for Hb F induction to treat the β-hemoglobinopathies.

2.1.5 β-Globin Gene Regulation

Several evolutionary conserved transcriptional regulatory elements within the proximal region of the β-globin promoter (Fig. 2.3b; GenBank accession #NG_000007.3; GI:28380636) have been identified including an initiator sequence TATA box at −30, a G-rich sequence at −50, CAAT box at −75, two CACC boxes at −90 to −110, and a directly repeated motif, the βDRE (Antoniou et al. 1995; Lewis and Orkin 1995; Stuve and Myers 1990). Of these regulatory sequences, only the βDRE sequence is found exclusively in the promoters of the adult genes. The CAAT box is important for promoter function in erythroid cells involving the binding of various transcription factors including CP1 (TFCP1), GATA1, and NFE4 (Antoniou and Grosveld 1990; deBoer et al. 1988). The CACCC box has been shown to bind several in vitro factors (Rodriguez et al. 2005; Vakoc et al. 2005); however, in vivo, EKLF was shown to be the transcription factor which binds to regulate β-globin expression (Miller and Bieker 1993).

Further upstream, the promoter contains additional binding sites for GATA1 (−120 and −200) and CP1 (−160). These sites are important for inducible β-globin promoter activity in erythroleukemia cells (Antoniou and Grosveld 1990; deBoer et al. 1988). In contrast, the gene remains inducible when the β-locus LCR is coupled directly to a minimal promoter (a promoter that is not sufficient to drive reporter expression without the presence of a functional enhancer) (Antoniou and Grosveld 1990; Levings and Bungert 2002); therefore, the role of these sequences in the context of the entire locus is not well established.

The β-globin gene also contains two enhancer elements. The first is located at the second intron and third exon border and the second several hundred bases downstream from the poly(A) site (Antoniou et al. 1988; deBoer et al. 1988; Behringer et al. 1987). Despite many studies, the role of the intron/exon enhancer element remains obscure. The second enhancer functions as a GATA1-binding site and stimulates the function of a linked promoter as shown by transfection experiments (Antoniou et al. 1988). Moreover, it functions as an adult stage-specific activator in transgenic mice experiments when the transgene is not linked to the LCR (Behringer et al. 1987; Kollias et al. 1986, 1987; Trudel and Constantini 1987; Magram et al. 1989). In addition, deletion of this 3′ β-globin enhancer in a yeast artificial chromosome (YAC) containing the entire β-globin locus produced a significant loss of β-globin expression in transgenic mice fetal liver and adult spleen (Liu et al. 1997), demonstrating that this element is required for β-gene expression during adult development.

2.2 Hemoglobin Variants and Human Diseases

The hemoglobinopathies are the most common single-gene disorders in the world (Flint et al. 1998). Over 1,500 structural hemoglobin variants exist (Hardison et al. 2002), the majority of which are clinically benign (Table 2.1). The globin gene server is a comprehensive public database where these mutations are cataloged (http://globin.cse.psu.edu/). Human hemoglobin variants are considered first in terms of the underlying mutations in globin gene structure and then in terms of their phenotypic expression. The clinical effect of mutations in the globin genes are discussed below.

Table 2.1 Hemoglobin structural variants (Adapted from the Globin Gene Server http://globin.cse.psu.edu/)

	Number
I. Structural variant by globin gene	
A. β-Like globin cluster	
$^{G}\gamma$-Globin	65
$^{A}\gamma$-Globin	55
δ-Globin	95
β-Globin	791
B. α-Like globin cluster	
α_1-Globin	313
α_2-Globin	373
II Structural variant by types	
Single nucleotide polymorphisms	1,227
Insertions	62
Deletions	180
Fusions	9

2.2.1 γ-Globin Gene Mutations

Experimental strategies to reactivate γ-gene expression can be developed from what is known about naturally occurring mutations that produce HPFH. Single nucleotide polymorphisms (SNPs) that occur in the γ-globin gene promoters (nondeletion HPFH) or large deletions in the β-locus (deletional HPFH) lead to a group of clinically heterogeneous disorders (Stamatoyannopoulos and Grosveld 2001). Many SNPs in the $^{G}\gamma$ or $^{A}\gamma$-globin gene promoter that produce HPFH have been identified. Experimental data support mechanisms for continued Hb F synthesis based on mutations in DNA–protein binding sites where either a new motif is created, which allows *trans*-activator binding, or a repressor protein-binding motif is destroyed. For example, a G/A mutation at −117 in the distal CAAT box, that is bound by GATA1 and the nuclear receptor subfamily 2, group F, member 2 (NR2F2; NF-E3) to silence γ-gene expression, leads to HPFH production (Gumucio et al. 1988; Mantovani et al. 1988). When CCAAT displacement protein binds the CAAT boxes to competitively displace CP1, it is believed to act as a transcriptional repressor (Aufiero et al. 1994). The −158 $^{G}\gamma$-globin C/T SNP interferes with this process leading to an HPFH phenotype as well (Sampietro et al. 1992), although the specific transcription factor(s) that binds this region has not been identified. Another SNP in the γ-globin promoter that produce HPFH includes the −175 (T/C) substitution associated with altered GATA1 and Octamer 1 binding (Stoming et al. 1989; Liu et al. 2005), suggesting a repressor role for GATA1; however, a direct correlation between GATA1-binding and promoter activity has been difficult to establish.

Further upstream, several mutations have been demonstrated in the −200 region that produce HPFH. This region is capable of forming a triplex structure, leaving the −206 to −217 γ-promoter sequence single-stranded (Ulrich et al. 1992). Triplex DNA is a secondary DNA structure which forms in an oligopyrimidine–oligpurine tract at acidic pH and negative DNA supercoiling (Ulrich et al. 1992). Sequences of this type are often found at regulatory regions in eukaryotic genomes and have been proposed to participate in the regulation of physiological processes such as transcription. Five different point mutations in the −200 region have been associated with HPFH; four of these mutations dramatically reduce the stability of the secondary DNA structure, suggesting that these mutations alter formation of the triplex by destabilizing critical Hoogsteen (triple-stranded) base pairs (Ulrich et al. 1992; Bacolla et al. 1995). This region is also thought to serve as the binding site for a repressor complex; mutations might result in displacement of repressor proteins allowing *trans*-activators to bind. Enhanced Sp1 and SSP binding to the −198 (T/C) and −202 (C/G) SNPs, respectively, might provide insights into the functional consequences of mutations in this region (Ronchi et al. 1989; Jane et al. 1993).

More recently, it was demonstrated that a repressor complex composed of GATA1, FOG1, and Mi2 is recruited to the $^{A}\gamma$-globin (−566) or $^{G}\gamma$-globin (−567) GATA1 sites when γ-globin expression was low but not when γ-globin is expressed at high levels. This suggests that γ-globin gene expression is silenced, in part, by

this complex (Harju-Baker et al. 2008), making it another potential region of interest in γ-globin regulation. Combined, these studies provide insights into strategies to develop therapeutic approaches to induce Hb F expression.

2.2.2 β-Globin Gene Mutations

In 1949, Linus Pauling published the seminal paper declaring sickle cell anemia a molecular disorder (Pauling et al. 1949), and the causative amino acid substitution in Hb S (β codon 6 GAG → GTG; Glu → Val; β^S) was demonstrated in 1957 by Ingram and associates (Ingram 1957). Subsequently, using restriction endonucleases to identify SNPs in the β-globin locus, inherited chromosome structures (haplotypes) were defined. In Africa, the β^S gene was associated with four haplotypes representing regions where independent mutations occurred including Benin, Senegal, Central African Republic (Bantu), and Cameroon (Pagnier et al. 1984; Nagel et al. 1985). A fifth Asian β-locus haplotype was reported in Saudi Arabia and India (Nagel and Fabry 1985). The regions in Africa where the spontaneous β^S-globin mutation arose to produce hemoglobin S (Hb S) are endemic for malarial infestation. This observation is consistent with the notion that the high incidence of the β^S-globin mutation is derived from natural selection (Carlson et al. 1994).

Hemoglobin C (β codon 6 GAG → AAG; Glu → Lys) also protects against malaria. Modiano and associates studied 4,000 Hb C trait patients (Modiano et al. 2007) and demonstrated they had fewer episodes of malaria infections than did Hb A controls, and even lower rates of infection were observed in Hb CC individuals. Hemoglobin E produced by a β codon 26 GAG → AAG (Glu → Lys) mutation is one of the world's most common hemoglobin variants. Individuals homozygous for the hemoglobin E allele (Hb EE) have a mild hemolytic anemia and mild splenomegaly; however, Hb E trait is asymptomatic. In combination with certain thalassemia mutations, it provides resistance to malaria infection. Hb E is most prevalent in Southeast Asia (Thailand, Indonesia, Bangladesh, and Vietnam) and North East India, where in certain areas carrier rates reach 60% of the population. Other hemoglobin mutations including Hb D (β codon 121 GAA → CAA; Glu → Gln) and Hb O can combine with Hb S to produce rarer forms of sickle cell disease (SCD).

2.2.2.1 Sickle Cell Disease

Sickle cell anemia (Hb SS) is an autosomal recessive genetic disorder which is inherited in a Mendelian pattern (Smith-Whitley and Pace 2007). About 8% of African Americans are carriers, and 100,000 individuals have some form of SCD. The most recent prevalence statistics published by the General Services Administration demonstrates that the distribution of SCD among different ethnic groups has become fluid

due to the high rate of racial admixture in the United States (http://www.pueblo.gsa.gov/cictext/health/sicklecell/496_sick.html). The data showed an 85% prevalence rate for all forms of SCD in individuals of African descent. The second most commonly affected group is Native Americans at a 10.6% prevalence rate; the remaining 4.4% of SCD occurs in Asians, Hispanics, and Caucasians. Sixty-five percent of SCD is caused by homozygous Hb SS disease, 25% by the compound heterozygote state Hb S and Hb C (Hb SC), 9% Hb Sβ-thalassemia, and the remaining 1% consists of other variants (Bond 2005).

Infants with Hb SS have a delay in the γ- to β-globin switch and Hb F levels average 9% at 24 months of age. This observation provided the impetus for widespread research efforts to understand mechanisms for γ-gene silencing and to develop strategies to reverse this process in SCD. The efficacy of Hb F is due to its ability to dilute Hb S concentrations below the threshold required for polymerization in erythrocytes; furthermore, Hb F has a direct influence on Hb S polymer stability (Bookchin et al. 1975). Asymmetric hybrid molecules Hb F/S ($\alpha_2,\gamma\beta^S$) are produced when Hb F levels remain elevated to produce an observed clinical benefit (Bookchin et al. 1977; Nagel et al. 1979).

The pathophysiology of SCD is based on Hb S polymerization that leads to the characteristic sickle-shaped red blood cells and oxidative membrane damage (Browne et al. 1998). The hallmark symptoms are chronic hemolysis, anemia, and complications related to vascular (vaso)-occlusion. SCD is characterized by recurrent vaso-occlusive episodes caused by sickle-shaped red blood cells that obstruct capillaries and restrict blood flow to organs, resulting in ischemia, pain, necrosis, and often organ damage. However, symptoms vary in frequency and severity between subpopulations of SCD patients in part due to variable Hb F levels (Platt et al. 1991) and other undefined genetic modifiers.

The Multicenter Study of Hydroxyurea was initiated in 1992 to establish the first drug treatment for SCD. The major outcome of this randomized clinical trial was a significant reduction in vaso-occlusive episodes in the majority of patients treated with hydroxyurea (Charache et al. 1995). Limitations to using this agent in adults, such as bone marrow suppression, concerns over long-term potential carcinogenic complications, and a 30% nonresponse rate (Steinberg et al. 1997), make the development of alternative therapies desirable. Subsequent studies in children showed that hydroxyurea induces Hb F levels to twice the average level achieved in adults (Zimmerman et al. 2004; Hankins et al. 2005; Wang et al. 2011), suggesting that the γ-globin genes may not be completely silenced in young subjects. This finding argues that early institution of treatment will permit greater efficacy and prevention of complications in SCD (Wang et al. 2011). Other Hb F-inducing agents including arginine butyrate (Perrine et al. 1993; Atweh et al. 1999), decitabine (Saunthararajah et al. 2003), and novel short-chain fatty acid derivatives (Perrine et al. 2011) are being investigated in humans as potential Hb F inducers for the treatment of SCD. Hematopoietic stem cell transplantation is currently the only cure for SCD (Walters and Sullivan 2010); however, gene therapy approaches are under development, which hold promise (Yannaki and Stamatoyannopoulos 2010).

2.2.2.2 β-Thalassemia

The thalassemias are the most common single-gene disorders in the world population. β-thalassemia is produced by mutations in the β-globin gene, inherited in an autosomal recessive fashion. The deficiency or absence of β-globin chains that characterizes β-thalassemia reflects the action of mutations that affect every level of β-globin gene function, including transcription, mRNA processing, translation, and posttranslational stability of the β-globin chain. The mutations are grouped according to the mechanism by which they affect β-globin expression (reviewed in Huisman 1997; Thein 1993). As the molecular pathology is worked out, a more accurate approach to the classification of the different types of β-thalassemia has become feasible.

Common causes of β-thalassemia include, but are not limited to, gene deletions (Huisman 1997; Thein 1993), mutations in the promoter regions of the β-globin gene (Antonarakis et al. 1984; Orkin et al. 1983, 1984; Gonzalez-Redondo et al. 1988; 1989), cap site mutations (Huisman 1997; Thein 1993), mutations of the 5′-untranslated region of the β-globin gene (Cai et al. 1992; Cheng et al. 1984; Thein 1993; Gonzalez-Redondo et al. 1988), and splice site mutations involving intron/exon boundaries (Antonarakis et al. 1984; Kazazian et al. 1984) among others. Over 200 mutations have been identified, but only 20 mutations account for 80% of β-thalassemia genes worldwide.

The severity of β-thalassemia depends on the nature of the defect. Mutations are characterized as β^o if they prevent any formation of β-globin chains or β^+ if they allow some β-globin chain formation to occur. In either case, there is a relative excess of α-globin chains which do not form tetramers; rather, they bind to the red blood cell membrane, producing membrane damage and toxic aggregates.

Several mutations have been identified in the proximal β-globin promoter including substitution in the TATA box around −30 bp from the cap site or in the CACACC elements at −90 and −105 bp (Huisman 1997; Orkin et al. 1984). These mutations produce decreased β-globin transcription from 10% to 25% of normal levels clinically manifesting at β^+-thalassemia. Another mutation at −101 (C/T) produces a mild deficit of β-globin mRNA (Gonzalez-Redondo et al. 1989); a mutation in the β-globin cap site at +1 bp (A/C) (Wong et al. 1987) and 5′-untranslated region at +33 (C/G) also produce a mild effect on β-globin transcription.

The boundaries between exons and introns are characterized by the invariant dinucleotides G-T at the donor (5′) site and A-G at the acceptor (3′) site. Mutations that affect either of these sites can abolish or alter normal splicing and give rise to β-thalassemia (Huisman 1997; Kazazian et al. 1984). Such mutations within the 5′ donor site at position five of the intervening sequence 1 (IVS-1) (G/C or G/T) can lead to alternative splicing and a reduction in β-globin chains (Orkin et al. 1982). A mutation at position 110 in IVS-1 (G/A) creates a cryptic 3′ donor site which produced 90% abnormal β-globin mRNA and severe β^+-thalassemia (Spritz et al. 1981). Finally, a mutation in the AAUAAA sequence in the 3′-untranslated region of β-globin mRNA leads to a 90% loss of normal transcript (Orkin et al. 1985).

If both alleles have β-thalassemia mutations (thalassemia major or Cooley's anemia), a severe microcytic, hypochromic anemia is observed. Untreated, it will lead to splenomegaly, severe bone deformities, and death before the age of 20 (Cunningham 2010). Treatment consists of periodic blood transfusion, splenectomy if hypersplenism is present, and treatment of transfusion-caused iron overload. Due to recent advances in iron chelation treatments, patients with thalassemia major can live long lives if they have access to proper treatment. Popular iron chelators include deferoxamine and deferiprone. Cure is possible by hematopoietic stem cell transplantation.

Thalassemia intermedia results when mutations in the β-globin gene lead to the synthesis of lower than normal levels of hemoglobin A; it is a milder form of β-thalassemia. These patients vary in their treatment needs, depending on the severity of their anemia. All thalassemia patients are susceptible to health complications that involve the spleen (which is often enlarged and frequently removed) and gall stones. These complications are mostly prevalent to thalassemia major and intermedia patients.

2.2.3 α-Thalassemia

α-Thalassemia is typically caused by gene deletions in the α-globin locus (Hardison et al. 2002), with one gene deletion leading to a silent carrier state (/−α) and two gene deletions (−α/−α) producing an α-thalassemia trait. α-Thalassemia results in decreased α-globin chain production, resulting in an excess of β-globin chains in adults and excess γ-globin chains in newborns (Higgs and Gibbons 2010). The excess β-chains form unstable tetramers called hemoglobin H comprised of four β-chains which have abnormal oxygen dissociation curves. The excess γ-chains form tetramers which are poor carriers of oxygen. Homozygote α^0 thalassemia (−/−, −/−), where only γ_4 hemoglobin molecules (Hb Barts) are produced, often result in a still birth or *hydrops fetalis*.

To interpret the molecular pathology of α-thalassemia, it is important to appreciate the structural variability in the α-globin cluster not associated with clinical abnormalities. There are numerous point mutations, rearrangements, and gene conversions that have no effect on α-globin gene expression (Lauer et al. 1980; Higgs et al. 1989) called nondeletion α^+-thalassemia. These disorders result from single or oligonucleotide mutations; they are much less common than deletion forms of α^+-thalassemia. For example, two cases of mutations that inactivate the initiation codon (ATG to ACG or GTG) interfere with translation of α-globin mRNA (Pirastu et al. 1984). Another group of mutations involve substitution in the α_2-globin termination codon TAA (Weatherall and Clegg 1975), leading to the insertion of an amino acid instead of chain termination. Several variant hemoglobin including Constant Spring, Icaria, and others are produced by this mechanism (Clegg et al. 1971, 1974). Finally, substitutions in the poly(A) signal have been demonstrated that interfere with termination of transcription (Higgs et al. 1983). Due to the wide variety of DNA mutations associates with α^+-thalassemia, direct sequence analysis is often required to define the abnormality at the molecular level.

About a third of patients with Hb SS have coincidental α-thalassemia (Steinberg and Embury 1986). These individuals have less hemolysis, higher packed cell volume, lower mean corpuscular volume, and lower reticulocyte counts (Embury et al. 1982; Steinberg et al. 1984). Coinheritance of α-thalassemia results in relatively longer erythrocyte life span because of the reduction of dense and rigid sickle red cells. The resulting increased blood viscosity may increase the incidence of certain vaso-occlusive complications such as painful episodes, acute chest syndrome, and osteonecrosis.

2.3 Summary

For many decades, globin gene expression has been the focus of intensive research efforts because of its value to enlighten the biology of developmental gene regulation. Analysis of the human globin genes in transgenic mice has provided many insights into mechanisms of hemoglobin switching. Moreover, numerous hemoglobinopathies resulting from genetic changes in coding and noncoding portions of the globin genes have been defined at the molecular level. Specifically, critical knowledge has been acquired through the study of naturally occurring HPFH mutations that have shed light on mechanisms of the γ- to β-globin switch. As genomic techniques advance, our appreciation of the impact of these changes in globin gene expression and DNA–protein interactions will be expanded. These data will provide a basis for strategies to induce Hb F expression and to reduce disease severity in individuals with SCD and β-thalassemia. Understanding the regulation of hemoglobin synthesis could potentially lead to novel gene-based therapeutic approaches or a cure for the hemoglobinopathies.

Abbreviations

CDP	CCAAT displacement protein
CBP	CREB-binding protein
DRED	Direct repeat erythroid-definitive
Hb F	Fetal hemoglobin
HPFH	Hereditary persistence of fetal hemoglobin
HS	Hypersensitive site
LCR	Locus control region
SSE	Stage selector element
SSP	Stage selector protein
STAT3	Signal transducers and activators of transcription
Hb SS	Sickle cell anemia
SCD	Sickle cell disease
SNP	Single nucleotide polymorphism

References

Antonarakis SE, Orkin SH, Cheng TC, Scott AF, Sexton JP, Trusko SP, Charache S, Kazazian HH Jr (1984) Beta-Thalassemia in American Blacks: novel mutations in the "TATA" box and an acceptor splice site. Proc Natl Acad Sci USA 81:1154–1158

Antoniou M, deBoer E, Habets G, Grosveld F (1988) The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. EMBO J 7:377–384

Antoniou M, Grosveld F (1990) Beta-globin dominant control region interacts differently with distal and proximal promoter elements. Genes Dev 4:1007–1013

Antoniou M, de Boer E, Spanopoulou E, Imam A, Grosveld F (1995) TBP binding and the rate of transcription initiation from the human beta-globin gene. Nucleic Acids Res 23:3473–3480

Asano H, Li XS, Stamatoyannopoulos G (1999) FKLF, a novel Kruppel-like factor that activates human embryonic and fetal β-like globin genes. Mol Cell Biol 19:3571–3579

Asano H, Li XS, Stamatoyannopoulos G (2000) FKLF-2: a novel Kruppel like transcriptional factor that activates globin and other erythroid lineage genes. Blood 95:3578–3584

Atweh GF, Sutton M, Nassif I, Boosalis V, Dover GJ, Wallenstein S, Wright E, Mc-Mahon L, Stamatoyannopoulos G, Faller DV, Perrine SP (1999) Sustained induction of fetal hemoglobin by pulse butyrate therapy in sickle cell disease. Blood 93:1790–1797

Aufiero B, Neufeld EJ, Orkin SH (1994) Sequence-specific DNA binding of individual cut repeats of the human CCAAT displacement/cut homeodomain protein. Proc Natl Acad Sci USA 91:7757–7761

Barberis A, Superti-Furga G, Busslinger M (1987) Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. Cell 50:347–359

Bacolla A, Ulrich MJ, Larson JE, Ley TJ, Wells RD (1995) An intramolecular triplex in the human gamma-globin 5′-flanking region is altered by point mutations associated with hereditary persistence of fetal hemoglobin. J Biol Chem 270:24556–24563

Behringer RR, Hammer RE, Brinster RL, Palmiter RD, Townes TM (1987) Two 3′ sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. Proc Natl Acad Sci USA 8:7056–7060

Bond DR (2005) Three decades of innovation in the management of sickle cell disease: the road to understanding the sickle cell disease clinical phenotype. Blood Rev 19:99–110

Bookchin RM, Nagel RL, Balaza T (1975) Role of hybrid tetramer formation in gelation of haemoglobin S. Nature 256:667–668

Bookchin RM, Balazs T, Nagel RL, Tellez I (1977) Polymerisation of haemoglobin SA hybrid tetramers. Nature 269:526–527

Borg J, Papadopoulos P, Georfitsi M, Gutierrez L, Grech G, Franis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, Cassar W, Galdies R, van Licken W, Ozqur Z, Gillemans N, Hou J, Bugeja M, Grosveld FG, von Lindern M, Felice AE, Patronis GP, Philipsen S (2010) Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat Genet 42:801–805

Browne P, Shalev O, Hebbel RP (1998) The molecular pathobiology of cell membrane iron: the sickle red cell as a model. Free Radic Biol Med 24:1040–1048

Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci USA 96:5129–5134

Cai SP, Eng B, Francombe WH, Olivieri NF, Kendall AG, Waye JS, Chui DH (1992) Two novel beta-thalassemia mutations in the 5′ and 3′ noncoding regions of the beta-globin gene. Blood 79:1342–1346

Carlson J, Nash GB, Gabutti V, al-Yaman F, Wahlgren M (1994) Natural protection against severe *Plasmodium falciparum* malaria due to impaired rosette formation. Blood 84:3909

Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, McMahon RP, DR B (1995) Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the multicenter study of hydroxyurea in sickle cell anemia. N Eng J Med 332:1317–1322

Chen Z, Luo HY, Steinberg MH, Chui DH (2009) BCL11A represses HBG transcription in K562 cells. Blood Cells Mol Dis 42:144–149

Cheng TC, Orkin SH, Antonarakis SE, Potter MJ, Sexton JP, Markham AF, Giardina PJ, Li A, Kazazian HH Jr (1984) Beta-thalassemia in Chinese: use of in vivo RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects. Proc Natl Acad Sci USA 81:2821–2825

Clegg JB, Weatherall DJ, Milner PF (1971) Haemoglobin constant spring – a chain termination mutant? Nature 234:337–340

Clegg JB, Weatherall DJ, Contopoou-Griva I, Caroutsos K, Poungouras P, Tsevrenis H (1974) Haemoglobin Icaria, a new chain-termination mutant with causes alpha thalassaemia. Nature 251:245–247

Crossley M, Whitelaw E, Perkins A, Williams G, Fujiwara Y, Orkin SH (1996) Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. Mol Cell Biol 16:1695–1705

Cunningham MJ (2010) Update on thalassemia: clinical care and complications. Hematol Oncol Clin N Am 24:215–227

deBoer E, Antoniou M, Mignotte V, Wall L, Grosveld F (1988) The human beta-globin promoter: nuclear protein factors and erythroid specific induction of transcription. EMBO J 7:4203–4212

Embury SH, Dozy AM, Miller J, Davis JR Jr, Kleman KM, Preisler H, Vichinsky E, Lande WN, Lubin BH, Kan YW, Mentzer WC (1982) Concurrent sickle-cell anemia and alpha-thalassemia: effect on severity of anemia. N Engl J Med 306:270–274

Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipsen S, Grosveld F (1996) A dominant chromatin-opening activity in 5′ hypersensitive site 3 of the human beta-globin locus control region. EMBO J 15:562–568

Enver T, Zhang J-W, Papayannopoulou T, Stammatoyannopoulos G (1988) DNA methylation: a secondary event in globin gene switching? Genes Dev 2:698–706

Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G (1990) Developmental regulation of human fetal to-adult globin gene switching in transgenic mice. Nature 344:309–313

Ferry A, Baliga S, Monterio C, Chen Y, Pace BS (1997) γ-Globin gene silencing in primary erythroid cultures: an inhibitory role for interleukin-6. J Biol Chem 272:20030–20037

Flint J, Harding RM, Boyce AJ, Clegg JB (1998) The population genetics of the haemoglobinopathies. Bailieres Clin Haematol 11:1–51

Foley H, Ofori-Acquah S, Baliga BS, Pace BS (2002) STAT3 mediates globin repression by interleukin-6 in K562 cells. J Biol Chem 77:16211–16219

Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin expressing hybrids. Nucleic Acids Res 15:10159–10177

Fraser P, Grosveld F (1998) Locus control regions, chromatin activation and transcription. Curr Opin Cell Biol 10:361–365

Fucharoen S, Shimizu K, Fukumaki Y (1990) A novel C-T transition within the distal CCAAT motif of the G gamma-globin gene in the Japanese HPFH: implication of factor binding in elevated fetal globin expression. Nucleic Acids Res 18:5245–5253

Gardiner MR, Daggett DF, Zon LI, Perkins AC (2005) Zebrafish KLF4 is essential for anterior mesendoderm/pre-polster differentiation and hatching. Dev Dyn 234:992–996

Gardiner MR, Gongora MM, Grimmone SM, Perkins AC (2007) A global role for zebrafish klf4 in embryonic erythropoiesis. Mech Dev 124:762–774

Gong QH, Stern J, Dean A (1991) Transcriptional role of a conserved GATA-1 site in the human epsilon-globin gene promoter. Mol Cell Biol 11:2558–2566

Gonzalez-Redondo JM, Stoming TA, Kutlar A, Kutlar F, Lanclos KD, Howard EF, Fei YJ, Aksoy M, Altay C, Gurgey A et al (1989) A C–T substitution at nt–101 in a conserved DNA sequence of the promotor region of the beta-globin gene is associated with "silent" beta-thalassemia. Blood 73:1705–1711

Gonzalez-Redondo JM, Stoming TA, Lanclos KD, Gu YC, Kutlar A, Kutlar F, Nakatsuji T, Deng B, Han IS, McKie VC et al (1988) Clinical and genetic heterogeneity in black patients with homozygous beta-thalassemia from the southeastern United States. Blood 72:1007–1014

Goodwin AJ, McInerney JM, Glander MA, Pomerantz O, Lowrey CH (2001) In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1. J Biol Chem 276:26883–26892

Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51:975–985

Gumucio DL, Rood KL, Gray TA, Riordan MF, Sartor CI, Collins FS (1988) Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. Mol Cell Biol 8:5310–5322

Hankins JS, Ware RE, Rogers ZR et al (2005) Long-term hydroxyurea therapy for infants with sickle cell anemia – the HUSOFT extension study. Blood 106:2269–2275

Hardison RC, Chui DH, Giardine B, Riemer C, Patrinos GP, Anagnou N, Miller W, Wajcman H (2002) HbVar: a relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. Hum Mutat 19:225–233

Harju-Baker S, Costa FC, Fedosyuk H, Neades R, Peterson KR (2008) Silencing of $^{A}\gamma$-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the −566 GATA site. Mol Cell Biol 28:3101–3113

Higgs DR, Wood WG, Jarman AP, Sharpe J, Lida J, Pretorius IM, Ayyub H (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev 4:1588–1601

Higgs DR, Vickers MA, Wilkie AO, Pretorius IM, Jarman AP, Weatherall DJ (1989) A review of the molecular genetics of the human alpha-globin gene cluster. Blood 73:1081–1104

Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. Nature 306:398–400

Higgs DR, Gibbons RJ (2010) The molecular basis of α-thalassemia: a model for understanding human molecular genetics. Hematol Oncol Clin N Am 24:1033–1054

Huisman TH (1997) Combinations of beta chain abnormal hemoglobins with each other or with beta-thalassemia determinants with known mutations: influence on phenotype. Clin Chem 43:1850–1856

Ingram VM (1957) Gene mutations in human haemoglobins: the chemical difference between normal and sickle cell haemoglobin. Nature 180:326

Jackson DA, McDowell JC, Dean A (2003) Beta-globin locus control region HS2 and HS3 interact structurally and functionally. Nucleic Acids Res 31:1180–1190

Jane SM, Gumucio DL, Ney PA, Cunningham JM, Nienhuis AW (1993) Methylation-enhanced binding of Sp1 to the stage selector element of the human gamma-globin gene promoter may regulate development specificity of expression. Mol Cell Biol 13:3272–3281

Jane SM, Nienhuis AW, Cunningham JM (1995) Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. EMBO J 14:97–105

Kalra IS, Alam MM, Choudhary PK, Pace BS (2011) Krüppel-like factor 4 activates HBG gene expression in primary erythroid cells. Br J Hematol 154:248–259

Kazazian HH Jr, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, Waber PG (1984) Molecular characterization of seven beta-thalassemia mutations in Asian Indians. EMBO J 3:593–596

Kollias G, Wrighton N, Hurst J, Grosveld F (1986) Regulated expression of human Aγ-, β-, and hybrid β γ-globin genes in transgenic mice: manipulation of the developmental expression patterns. Cell 46:89–94

Kollias G, Hurst J, deBoer E, Grosveld F (1987) The human beta-globin gene contains a downstream developmental specific enhancer. Nucleic Acids Res 15:5739–5747

Lauer J, Shen CK, Maniatis T (1980) The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. Cell 20:119–130

Levings PP, Bungert J (2002) The human beta-globin locus control region. Eur J Biochem 269:1589–1599

Lewis BA, Orkin SH (1995) A functional initiator element in the human beta-globin promoter. J Biol Chem 270:28139–28144

Li S, Moy L, Pittman N, Shue G, Aufiero B, Neufeld EJ, LeLeiko NS, Walsh MJ (1999) Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/*cut* homolog, is associated with histone deacetylation. J Biol Chem 274:7803–7815

Li S, Aufiero B, Schiltz RL, Walsh MJ (2000) Regulation of the homeodomain CCAAT displacement/cut protein function by histone acetyltransferases p300/CREB-binding protein (CBP)-associated factor and CBP. Proc Natl Acad Sci USA 97:7166–7171

Liu Q, Bungert J, Engel JD (1997) Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. Proc Natl Acad Sci USA 94:169–174

Liu LR, Du ZW, Zhao HL, Liu XL, Huang XD, Shen J, Ju LM, Fang FD, Zhang JW (2005) T to C substitution at −175 or −173 of the gamma-globin promoter affects GATA-1 and OCT-1 binding in vitro differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. J Biol Chem 280:7452–7459

Magram J, Niederreither K, Costantini F (1989) Beta-globin enhancers target expression of a heterologous gene to erythroid tissues of transgenic mice. Mol Cell Biol 9:4581–4584

Maier-Redelsperger M, de Noguchi CT, Montalembert M, Rodgers GP, Schechter AN, Gourbil A, Blanchard D, Jais JP, Ducrocq R, Peltier JY (1994) Variation in fetal hemoglobin parameters and predicted hemoglobin S polymerization in sickle cell children in the first two years of life: Parisian prospective study on sickle cell disease. Blood 84:3182–3188

Mantovani R, Malgaretti N, Nicolis S, Ronchi A, Giglioni B, Ottolenghi S (1988) The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors. Nucleic Acids Res 16:7783–7797

Mantovani R (1998) A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res 26: 1135–1143

Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239:15–27

Marini MG, Procu L, Asunis I, Loi MG, Ristaldi MS, Procu S, Ikuta T, Cao A, Moi P (2010) Regulation of the human HBA genes by KLF4 in erythroid cell lines. Br J Haematol 149: 748–758

Martin DI, Orkin SH (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev 4:1886–1898

Mavilio F, Giampaolo A, Care A, Migliaccio G, Calandrini M, Russo G, Pagliardi GL, Mastroberardino G, Marinucci M, Peschle C (1983) Molecular mechanisms of human hemoglobin switching: selective undermethylation and expression of globin genes in embryonic, fetal, and adult erythroblasts. Proc Natl Acad Sci USA 80:6907–6911

Miller IJ, Bieker JJ (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol Cell Biol 13:2776–2786

Modiano D, Bancone G, Ciminelli BM, Pompei F, Blot I, Simpore J, Modiano G (2007) Haemoglobin S and haemoglobin C: 'quick but costly' versus 'slow but gratis' genetic adaptations to plasmodium falciparum malaria. Hum Mol Genet 17:789–799

Nagel RL, Bookchin RM, Johnson J, Labie D, Wajcman H, Isaac-Sodeye WA, Honig GR, Schiliro G, Crookston JH, Matsutomo K (1979) Structural bases of the inhibitory effects of Hb F and A2 on the polymerization of Hb S. Proc Natl Acad Sci USA 76:670–672

Nagel RL, Fabry ME, Pagnier J, Zohoun I, Wajcman H, Baudin V, Labie D (1985) Hematologically and genetically distinct forms of sickle cell anemia in Africa. N Engl J Med 312:880

Nagel RL, Fabry ME (1985) The many pathophysiologies of sickle cell anemia. Am J Hematol 20:195

Neufeld EJ, Skalnik DG, Lievens PM, Orkin SH (1992) Human CCAAT displacement protein is homologous to the *Drosophila* homeoprotein, cut. Nat Genet 1:50–55

Ney PA, Sorrentino BP, Lowrey CH, Nienhuis AW (1990) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. Nucleic Acids Res 18:6011–6017

Nishio H, Walsh MJ (2004) CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription. Proc Natl Acad Sci USA 101:11257–11262

Orkin SH, Antonarkis SE, Kaxazian JJJR (1984) Base substitution at position −88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. J Biol Chem 259:8679–8681

Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH Jr (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. EMBO J 4:453–456

Orkin SH, Kazazian HH Jr, Antonarakis SE, Ostrer H, Goff SC, Sexton JP (1982) Abnormal RNA processing due to the exon mutation of beta E-globin gene. Nature 300:768–769

Orkin SH, Sexton JP, Cheng TC, Goff SC, Giardina PJ, Lee JI, Kazazian HH Jr.(1983) ATA box transcription mutation in beta-thalassemia. Nucleic Acids Res. 11:4727–34

Osada S, Yamamoto H, Nishihara T, Imagawa M (1996) DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. J Biol Chem 271:3891–3896

Pagnier J, Mears JG, Dunda-Belkhodia O, Schaefer-Rego KE, Beldjord C, Nagel RL, Labie D (1984) Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. Proc Natl Acad Sci USA 81:1771–1773

Pauling L, Itano HA, Singer SJ and Wells IC (1949) Sickle cell anemia a molecular disease. Science 110:543–8

Perrine SP, Ginder GD, Faller DV, Dover GH, Ikuta T, Witkowska HE, Cai SP, Vichinsky EP, Olivieri NF (1993) A short-term trial of butyrate to stimulate fetal-globin-gene expression in the beta-globin disorders. N Engl J Med 328:81–86

Perrine SP, Wargin WA, Boosalis MS, Wallis WJ, Case S, Keefer JR, Faller DV, Welch WC, Berenson RJ (2011) Evaluation of safety and pharmacokinetics of sodium 2,2 dimethylbutyrate, a novel short chain fatty acid derivative, in a phase 1, double-blind, placebo-controlled, single-dose, and repeat-dose studies in healthy volunteers. J Clin Pharmacol 51:1186–1194

Peters B, Merezhinskaya N, Diffley JF, Noguchi CT (1993) Protein-DNA interactions in the epsilon-globin gene silencer. J Biol Chem 268:3430–3437

Pirastu M, Saglio G, Chang JC, Cao A, Kan YW (1984) Initiation codon mutation as a cause of alpha thalassemia. J Biol Chem 259:12315–12317

Platt OS, Thorington BD, Brambilla DJ, Milner PF, Rosse WF, Vichinsky E, Kinney TR (1991) Pain in sickle cell disease. Rates and risk factors. N Engl J Med 325:11–16

Raich N, Papayannopoulou T, Stamatoyonnopoulos G, Enver T (1992) Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. Blood 79:861–864

Raich N, Clegg CH, Grofti J, Romeo PH, Stamatoyannopoulos G (1995) GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. EMBO J 14:801–809

Ristaldi MS, Drabek D, Gribnau J, Poddie D, Yannoutsous N, Cao A, Grosveld F, Imam AM (2001) The role of the −50 region of the human gamma-globin gene in switching. EMBO J 20:5242–5249

Ronchi A, Nicolis S, Santoro C, Ottolenghi S (1989) Increased Sp1 binding mediates erythroid-specific overexpression of a mutated (HPFH) gamma-globulin promoter. Nucleic Acids Res 17:10231–10241

Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO J 24:2354–2366

Sampietro M, Thein SL, Contreras M, Pazmany L (1992) Variation of Hb F and F-cell number with the G-gamma Xmn I (C-T) polymorphism in normal individuals. Blood 79:832–833

Sankaran VG, Menne TF, Xu J, Akie TE, Letter G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322:1839–1842

Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudin M, Bender MA, Tucker PW, Orkin SH (2009) Developmental and species-divergent globin switching are driven by BCL11A. Nature 460:1093–1097

Sauntharajah Y, Hillery CA, Lavelle D, Molokie R, Dorn L, Bressler L, Gavazova S, Chen YH, Hoffman R, DeSimone J (2003) Effects of 5-aza-2′-deoxycytidine on fetal hemoglobin levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. Blood 102:3865–3870

Skalnik DG, Strauss EC, Orkin SH (1991) CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter. J Biol Chem 266:16736–16744

Smith-Whitley K, Pace BS (2007) Sickle cell disease: a phenotypic patchwork. In: Pace BS (ed) Renaissance of sickle cell disease research in the genome era. Imperial College Press, London, pp 45–63

Spritz RA, Jagadeeswaran P, Choudary PV, Biro PA, Elder JT, deRiel JK, Manley JL, Gdfter ML, Forget BG, Weissman SM (1981) Base substitution in an intervening sequence of a beta+−thalassemic human globin gene. Proc Natl Acad Sci USA 78:2455–2459

Stamatoyannopoulos G, Josephson B, Zhang JW, Li Q (1993) Developmental regulation of human gamma-globin genes in transgenic mice. Mol Cell Biol 13:7636–7644

Stamatoyannopoulos G, Grosveld F (2001) Hemoglobin switching. In: Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H (eds) The molecular basis of blood disease, vol 3. Saunders, Philadelphia

Steinberg MH, Rosenstock W, Coleman MB, Adams JG, Platica O, Cedeno M, Rieder RF, Wilson JT, Milner P, West S (1984) Effects of thalassemia and microcytosis on the hematologic and vasoocclusive severity of sickle cell anemia. Blood 63:1353–1360

Steinberg MH, Embury SH (1986) Alpha-thalassemia in blacks: genetic and clinical aspects and interactions with the sickle hemoglobin gene. Blood 68:985–990

Steinberg MH, Lu ZH, Barton FB, Terrin ML, Charache S, Dover GJ (1997) Fetal hemoglobin in sickle cell anemia: determinants of response to hydroxyurea. Multicenter study of hydroxyurea. Blood 89:1078–1088

Stoming TA, Stoming GS, Lanclos KD, Fei YJ, Altay C, Kutlar F, Huisman TH (1989) An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T→C mutation at position −175 to the cap site of the A gamma globin gene. Blood 73:329–333

Strouboulis J, Dillon N, Grosveld F (1992) Developmental regulation of a complete 70-kb human β-globin locus in transgenic mice. Genes Dev 6:1857–1864

Stuve LL, Myers RM (1990) A directly repeated sequence in the beta-globin promoter regulates transcription in murine erythroleukemia cells. Mol Cell Biol 10:972–981

Superti-Furga G, Barberis A, Schaffner G, Busslinger M (1988) The −117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. EMBO J 7:3099–3107

Tanabe O, Katsuoka F, Campbell AD, Song W, Yamamoto M, Tanimoto K, Engel JD (2002) An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer. EMBO J 21:3434–3442

Thein SL (1993) Beta-thalassaemia. Baillieres Clin Haematol 6:151–175

Townes TM, Behringer RR (1990) Human globin locus activation region (LAR): role in temporal control. Trends Genet 6:219–223

Trudel M, Constantini F (1987) A 3′ enhancer contributes to the stage-specific expression of the human beta-globin gene. Genes Dev 1:954–961

Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell 90:109–119

Tuan DY, Solomon WB, Cavallesco R, Huang G, London IM (1989) Characterization of a human globin enhancer element. Prog Clin Biol Res 316A:63–72

Ulrich MJ, Gray WJ, Ley TJ (1992) An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. J Biol Chem 267:18649–18658

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell 17:453–462

Walters MC, Sullivan KM (2010) Stem-cell transplantation for sickle cell disease. N Engl J Med 362:955–956

Wood WG, Bunch C, Kelly S, Gunn Y, Breckon G (1985) Control of haemoglobin switching by a developmental clock? Nature 313:320–323

Wood WG (1993) Increased HbF in adult life. Baillieres Clin Haematol 6:177–213

Wang WC, Ware RE, Miller ST, Iyer RV, Casella JF, Minniti CP, Raba S, Thornburg CD, Roger ZR, Kalpatthi RV, Barredo JC, Brown RC, Sarnaik SA, Howard TH, Wynn LW, Kutlar A, Armstron FD, Files BA, Goldsmith JC, Waclawiw MA, Huang X, Thompson BW, BABY HUG investigators (2011) Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). Lancet 377:1663–1672

Weatherall DJ, Clegg JB (1975) The alpha-chain-termination mutants and their relation to the alpha-thalassaemias. Philos Trans R Soc Lond B Biol Sci 271:411–455

Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazizn JJJR (1987) Characterization of beta-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. Nature 330:384–386

Yannaki E, Stamatoyannopoulos G (2010) Hematopoietic stem cell mobilization strategies for gene therapy of beta thalassemia and sickle cell disease. Ann N Y Acad Sci 1202:59–63

Yu CY, Motamed K, Chen J, Bailey AD, Shen CK (1991) The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element in vitro and in vivo. J Biol Chem 266:8907–8915

Zhang P, Basu P, Redmond LC, Morris PE, Rupon JW, Ginder GD, Lloyd JA (2005) A functional screen for Krüppel-like factors that regulate the human gamma-globin gene through the CACCC promoter element. Blood Cells Mol Dis 35:227–235

Zhao Q, Cumming H, Cerruti L, Cunningham JM, Jane SM (2004) Site-specific acetylation of the fetal globin activator NF-E4 prevents its ubiquitination and regulates its interaction with the histone deacetylase, HDAC1. J Biol Chem 279:41477–41486

Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM (2010) KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat Genetic 42:742–744

Zimmerman SA, Schultz WH, Davis JS, Pickens CV, Mortier NA, Howard TA (2004) Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. Blood 103:2039–2045

Chapter 3
Regulatory Polymorphisms and Osteoporosis

Huilin Jin and Stuart H. Ralston

Abstract Osteoporosis is a common disease characterized by low bone mass and micro-architectural deterioration of bone tissue which leads to an increased risk of fragility fracture. Genetic factors play an important role in regulating bone mineral density (BMD) and other phenotypes relevant to the pathogenesis of osteoporosis. It is currently believed that a large number of susceptibility alleles contribute to the risk of osteoporosis each with a small effect size. Very little is known about the molecular mechanisms by which these variants predispose to osteoporosis, but it is likely that many affect regulatory elements and act by altering gene expression. Here, we review the molecular mechanisms by which common variants at the *ERS1*, *COL1A1* and *VDR* loci regulate gene expression and predispose to osteoporosis. The most extensively studied locus is *COL1A1*, where a specific haplotype encompassing polymorphisms in the promoter and intron 1 leads to over-expression of *COL1A1* mRNA and an imbalance in production of the collagen type 1 α1 chain relative to the α2 chain. Polymorphisms in the regulatory regions of *ESR1* and *VDR* have also been described which modulate gene expression, but the mechanisms by which these predispose to osteoporosis have not been fully investigated. Genome-wide association studies have identified several variants that are associated with the expression of these genes, but further work will be required to define the responsible mechanisms.

Keywords Osteoporosis • Bone mineral density • Estrogen receptor • Type 1 collagen • Vitamin D receptor

H. Jin
Kennedy Institute of Rheumatology, University of Oxford,
65 Aspentea Road, London W6 8LH
e-mail: huilin.jin@kennedy.ox.ac.uk

S.H. Ralston (✉)
Molecular Medicine Centre, Institute of Genetics and Molecular Medicine,
Western General Hospital, Edinburgh EH4 2XU, UK
e-mail: Stuart.Ralston@ed.ac.uk

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_3,

3.1 Introduction

Osteoporosis is a common disease with a strong genetic component characterized by low bone mass, architectural deterioration of bone tissue and an increased risk of low trauma fractures. It is a major public health problem worldwide with enormous social and economic impact, with annual treatment costs estimated at $20 billion in the USA and about $30 billion in the European Union (Cummings and Melton 2002).

Osteoporosis is diagnosed when bone mineral density levels fall more than 2.5 standard deviations below those observed in young healthy individuals (T-score less than −2.5) (Kanis et al. 1994). According to this definition, osteoporosis is predominantly a disease of older people; it is uncommon below the age of 50 but increases in incidence thereafter to affect about 50% of women and 12% of men at some point in life. Fractures also increase in incidence with age in both men and women although this is only partly due to the reduction in BMD; a more important factor is the increased risk of falling with age due to a deterioration in muscle power, balance and cognitive function (De Laet et al. 1997).

3.1.1 Pathophysiology

Osteoporosis occurs because of an imbalance in the amount of bone that is removed by osteoclasts during the bone remodeling cycle and the amount that is replaced by bone formation. Many factors influence this process including circulating hormones such as estrogen, parathyroid hormone and calcitriol and locally produced regulatory molecules such as receptor activator of nuclear factor kappa B ligand (RANKL), osteoprotegerin (OPG), sclerostin (SOST) and members of the Wnt family of proteins. Estrogen plays a key role in protecting against osteoporosis in both men and women by inhibiting bone resorption and coupling resorption to new bone formation. Deficiency of estrogen such as occurs after menopause results in increased bone turnover with relative uncoupling of bone resorption and bone formation, leading to bone loss.

There are many risk factors for osteoporosis including early menopause, inflammatory diseases, poor diet, excessive alcohol intake, smoking and immobility. In addition, many drug treatments predispose to osteoporosis including corticosteroids which suppress bone formation and aromatase inhibitors which lower estrogen levels by inhibiting aromatization of adrenal androgens. One of the most important risk factors for osteoporosis is family history, emphasizing the importance of genetic factors in the pathogenesis of the disease.

3.1.2 Genetic Architecture of Osteoporosis

Twin and family studies have shown that BMD and other osteoporosis-related phenotypes such as biochemical markers of bone turnover, skeletal geometry, ultrasound

Table 3.1 Heritability of BMD and other osteoporosis phenotypes

Phenotype	Heritability	References
BMD	0.5–0.85	Pocock et al. (1987), Arden et al. (1996)
Fracture	0.1–0.68	Michaelsson et al. (2005)
Biochemical markers of bone turnover	0.59–0.75	Hunter et al. (2001)
Skeleton geometry	0.62	Arden et al. (1996)
Quantitative ultrasound	0.53	Arden et al. (1996)
Lean body mass and muscle strength	0.63–0.82	Arden and Spector (1997), Kaprio et al. (1995)
Age at menarche	0.37	Kaprio et al. (1995)
Age at menopause	0.59–0.63	Snieder et al. (1998)

properties of bone, body mass index (BMI) and muscle strength all have a heritable component (Table 3.1).

Heritability studies have indicated that genetic influences on BMD and the other phenotypes mentioned above are polygenic in nature and are mediated by a large number of variants of modest effect sizes and their interactions with environmental factors (Gueguen et al. 1995). This has been borne out by the results of genome-wide association studies (GWAS) which have identified a large number of susceptibility loci for BMD and fracture which have small effect sizes (Rivadeneira et al. 2009). It is unclear to what extent rare variants of medium to large effect size also contribute to the pathogenesis of osteoporosis, but this is likely to become apparent as genome-wide sequencing is performed in patients with the disease. It should be noted that several rare diseases have been identified where osteoporosis, fragility fractures or unusually high bone mass are caused by mutations in single genes with large effects, including osteogenesis imperfecta (OI), osteoporosis-pseudoglioma syndrome (OPS) and high bone mass syndromes (Balemans et al. 2005; Janssens and Van Hul 2002). Although these diseases are caused by rare mutations, some of the common polymorphic variations in the disease-causing genes also contribute to the regulation of BMD in the general population.

A large number of genes and loci have been identified that are associated with BMD, but at the current time, only 20 loci have attained the threshold for genome-wide significance (Rivadeneira et al. 2009), and none have been identified where genome-wide significance has been attained for the phenotype of fracture.

3.2 Regulatory Variants and Osteoporosis

Little work has been done to characterize the functional mechanisms by which susceptibility alleles for osteoporosis regulate gene expression or function. Here, we review specific examples of candidate genes where polymorphisms have been identified that are associated with BMD and where functional assays have been

conducted. We also briefly review the evidence emerging from GWAS which suggest that additional polymorphisms affecting the regulatory regions of candidate genes predispose to osteoporosis.

3.2.1 Estrogen Receptor

Estrogen plays an important role in many physiological processes, one of them being the regulation of bone mass and turnover. There are two estrogen receptors (alpha and beta) encoded by the *ESR1* and *ESR2* genes, respectively. Preclinical studies indicate that *ESR1* plays the predominant role in regulating bone mass and bone turnover in males, whereas in females, both receptors are important (Sims et al. 2002). The *ESR1* locus has been implicated as a genetic determinant of susceptibility to osteoporosis by candidate gene studies (Albagha et al. 2005; Ioannidis et al. 2002) and GWAS (Rivadeneira et al. 2009; Styrkarsdottir et al. 2008). Polymorphisms in the coding regions of *ESR1* do not appear to be responsible for these associations; instead, it seems likely that variants affecting the regulatory region of the gene are responsible. Most work has focused on two polymorphisms, rs2234693 and rs9340799, which are situated within intron 1 and recognized by the restriction enzymes *PvuII* and *XbaI*, respectively. These polymorphisms are in strong linkage disequilibrium with each other and with a TA repeat polymorphism in the *ESR1* promoter (Albagha et al. 2001). While the region surrounding the intron 1 polymorphisms is highly conserved across species, the region surrounding the TA repeat polymorphism is not, suggesting that the intron 1 variants may be responsible for the associations observed (Albagha et al. 2001). Bioinformatic analysis has shown that the rs2234693 (*PvuII*) polymorphism is situated at a consensus binding site for the AP-4 (*TFAP4*) and v-myb myeloblastosis viral oncogene homolog (*MYB*) transcription factors (Albagha et al. 2001; Herrington et al. 2002). In addition, promoter-reporter assays have shown that the C allelic variant at the *PvuII* site gives significantly greater reporter gene expression when compared with the T-variant in the presence of *MYB* (Herrington et al. 2002). Other researchers have also confirmed that haplotypes in this region regulate transcription in reporter assays (Maruyama et al. 2000). At present, it is unclear whether the associations with BMD noted above are primarily driven by variation at the polymorphic sites discussed above or whether other polymorphic sites within this region also contribute to the phenotype.

3.2.2 Type 1 Collagen

Type 1 collagen is the major bone protein. It is a triple-helical protein comprising two $\alpha 1$ polypeptide chains and $\alpha 2$ polypeptide chain, which are encoded by the collagen, type 1, alpha 1 (*COL1A1*) and collagen, type 1, alpha 2 (*COL1A2*) genes, respectively. Mutations affecting the protein coding regions of both genes cause

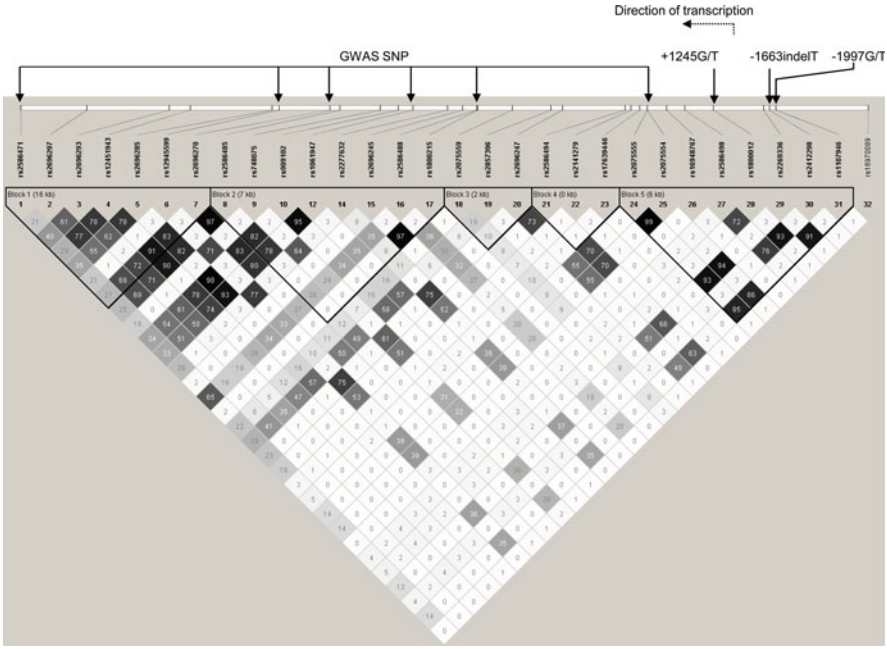

Fig. 3.1 Linkage disequilibrium across the COL1A1 locus as assessed by the R2 value. Dark grey or black boxes indicate high LD whereas light grey or white boxes indicate low LD. The six SNPs included in GWAS panels are highlighted along with the intron 1 and promoter polymorphisms

osteogenesis imperfecta, a rare disorder characterized by reduced BMD and fragility fractures (Byers 2000). Over recent years, evidence has emerged to suggest that common variants in the regulatory region of the *COL1A1* gene predispose to osteoporosis (Grant et al. 1996; Jin et al. 2011).

The first study to be reported was that of Grant and colleagues who identified a G/T polymorphism (rs1800012; at position +1245) in the first intron of the *COL1A1* gene (Fig. 3.1) that was associated with low bone mass and an increased risk of fracture (Grant et al. 1996). This polymorphism was shown to lie within an Sp1 transcription factor (Sp1) binding site. It has long been established that intron 1 of the human *COL1A1* gene contains regulatory elements including several Sp1 binding sites that play a role in regulating gene transcription (Bornstein et al. 1987). Positive associations between the Sp1 polymorphism and BMD, fractures and postmenopausal bone loss were subsequently reported in several studies (Keen et al. 1999; Uitterlinden et al. 1998; Langdahl et al. 1998; Garnero et al. 1998) and in meta-analyses of published studies (Mann and Ralston 2003; Mann et al. 2001; Jin et al. 2011; Ralston et al. 2006). The *COL1A1* locus has not emerged as a genome-wide significant determinant of BMD in recent GWAS studies, although it should be noted that the rs1800012 polymorphism shows no significant linkage disequilibrium with the SNPs used in the marker panels for these studies (Fig. 3.1) (Jin et al. 2011).

However, even in the largest meta-analysis which included 24,511 participants, the association between the rs1800012 SNP and BMD did not attain genome-wide significance (Ralston et al. 2006). Association studies between the *COL1A1* Sp1 polymorphism and osteoporosis suggested that the osteoporosis-associated "T" variant reduced BMD in a dominant fashion with an allele-dose effect (Grant et al. 1996; Uitterlinden et al. 1998). More recent meta-analyses have suggested that the effect on BMD is only observed in homozygotes for the T-variant, whereas the association with vertebral fracture is allele-dose dependent (Ralston et al. 2006). In many studies, the association between *COL1A1* variants and vertebral fracture has not been fully explained on the basis of the association with BMD, which indicates that variation at this site may influence bone quality (Ralston et al. 2006). In addition to BMD and vertebral fracture, genetic variation at rs1800012 has been associated with other phenotypes relevant to osteoporosis, such as femoral neck geometry (Qureshi et al. 2001), bone mineralization in vitro and in vivo (Stewart et al. 2005) and the therapeutic response to etidronate therapy (Qureshi et al. 2002).

The mechanisms by which rs1800012 predisposes to osteoporosis have been extensively studied. The rs1800012 polymorphism was shown to be at a binding site for Sp1 by Grant and colleagues (Grant et al. 1996), and it was subsequently shown that binding affinity for the Sp1 protein was greater for the osteoporosis-associated T allele as compared with the G allele. These differences in DNA-protein binding were accompanied by increased allele-specific transcription of *COL1A1* mRNA relative to *COL1A2* mRNA in cultured osteoblasts from patients heterozygous for rs1800012 (Fig. 3.2) (Mann et al. 2001). This was accompanied by a relative increase in the amount collagen type 1 α1 protein produced relative to the collagen type 1 α2 chain such that in G/T heterozygotes, the ratio of collagen type 1 α1 to collagen type 1 α2 chains was increased to 2.3:1 instead of the expected 2:1 observed in G/G homozygotes. This suggests that some of the collagen produced by osteoblasts in G/T heterozygotes is in the form of collagen α1 homotrimers which are mechanically weaker than α1/α2 heterotrimers because of altered inter-molecular cross-linking (Misof et al. 1997). In keeping with the hypothesis that rs1800012 alleles may influence bone quality, bone cores from G/T heterozygotes were shown to have impaired bone strength when compared with G/G homozygotes by biomechanical testing. In addition, a subtle reduction in bone mineralization was also detected by quantitative backscatter electron imaging (Stewart et al. 2005; Jin et al. 2009a). The mineralisation potential of cultured osteoblasts was also found to be reduced in G/T heterozygotes (Stewart et al. 2005). Taken together, these data suggest that the rs1800012 polymorphism is a functional variant that influences Sp1 DNA binding, *COL1A1* transcription and protein production which has adverse effects on bone composition and biomechanical strength.

Two other polymorphisms have been identified in the promoter of the *COL1A1* gene that are in linkage disequilibrium with each other and with rs1800012 (Fig. 3.1). These are a G/T polymorphism at position −1997 relative to the transcription start site (rs1107946) and an insertion/deletion polymorphism of a T-residue at position −1663 (rs2412298) (Garcia-Giralt et al. 2002). Variants at these two polymorphic sites have been found to interact with each other and with rs1800012 to regulate

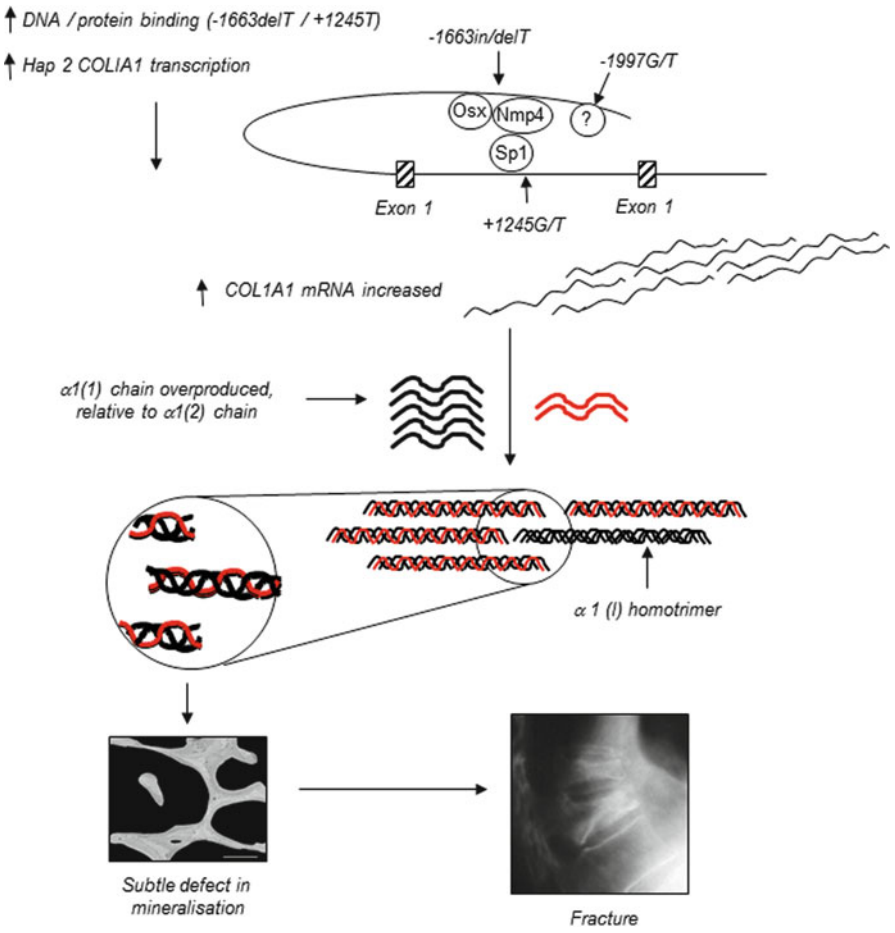

Fig. 3.2 *Proposed mechanism by which COL1A1 alleles predispose to osteoporosis.* Polymorphisms in the 5′ flanking region of *COL1A1* regulate binding affinity of several critical nuclear binding proteins including Nmp4, Osterix and Sp1. This increases transcription of haplotype 2 which causes increased expression of *COL1A1* mRNA and over production of the alpha 1 chain relative to alpha 2, such that a proportion of collagen produced by osteoblasts is in the form of alpha 1 homotrimers. This adversely affects mineralization of bone, resulting in reduced BMD and an increased risk of fracture

BMD and predict fracture risk in several populations (Jin et al. 2009a; Garcia-Giralt et al. 2002; Bustamante et al. 2007; Stewart et al. 2006). The promoter polymorphisms at positions −1997 and −1663 have also been shown to have effects on regulation of *COL1A1* transcription. The region surrounding the −1663indelT polymorphism is homologous to the rat *COL1A1* promoter site B, and previous studies in the rat have shown that the polyT tract in this region binds the transcription factor zinc finger protein 384 (ZNF384; NMP4) (Alvarez et al. 1997). The region surrounding the human −1663indelT site has also been found to bind NMP4

with greater affinity of binding for the −1663delT allele (Garcia-Giralt et al. 2002, 2005; Jin et al. 2009b). This is of interest in light of the fact that NMP4 negatively regulates collagen transcription (Robling et al. 2009) and that the −1663delT allele is associated with reduced BMD (Stewart et al. 2006). The region upstream of −1663indelT site also binds the SP7 transcription factor (SP7; Osterix) (Jin et al. 2009b), which plays an essential role in osteoblast differentiation (Nakashima et al. 2002). The region surrounding the −1997 site contains a consensus sequence for Sp1 binding (Garcia-Giralt et al. 2005). Although this site recognizes osteoblast-derived nuclear proteins in vitro (Garcia-Giralt et al. 2002), competition assays indicate that Sp1 does not appear to be one of the proteins responsible for DNA binding in nuclear extracts from osteoblast-like cells (Jin et al. 2009b). Promoter-reporter assays and chromatin immunoprecipitation assays (ChIP) described below have shown that the polymorphic sites in the promoter and intron 1 interact together to form a haplotype that regulates *COL1A1* transcription (Garcia-Giralt et al. 2005; Jin et al. 2009b). When examined individually, higher levels of transcription have been observed with the G allele at the −1997 site, with the delT allele at the −1663 site and with the T allele at the +1245 (Sp1) site (Garcia-Giralt et al. 2005; Jin et al. 2009b). In accord with these results, the highest overall transcription has been observed with the −1997G/−1663delT/+1245T haplotype (Jin et al. 2009b), which has also been associated with reduced BMD in clinical studies (Stewart et al. 2006; Jin et al. 2009b).

Analysis of the *COL1A1* 5′ flanking region by ChIP assays using antibodies for NMP4, Osterix and Sp1 has shown evidence that the promoter region and intron 1 interact to regulate transcription and that Nmp4 is recruited to the Sp1 binding site within this intron (Jin et al. 2009b). Taken together, these observations are consistent with a model whereby increased *COL1A1* transcription driven by an interaction between the three polymorphic sites in the regulatory region of the gene predisposes to osteoporosis, probably by increasing production of the alpha 1 chain and disrupting the normal ratio of collagen type 1 alpha 1 and alpha 2 chains.

3.2.3 Vitamin D Receptor

The active metabolites of vitamin D play an important role in regulating bone cell function and maintenance of calcium homeostasis by binding to the vitamin D receptor (VDR). The *VDR* gene was one of the first candidates to be studied in relation to osteoporosis. The first study was that of Morrison and colleagues who found an association between polymorphisms affecting the 3′ region of *VDR* and circulating osteocalcin levels (Morrison et al. 1992). In a subsequent study, the same group reported a significant association between a *BsmI* polymorphisms in intron 8 of *VDR* (rs1544410) and BMD in a twin study and a population-based study (Morrison et al. 1997). Further association studies of these polymorphisms were performed in relation to BMD and osteoporotic fractures with conflicting results although most of these studies were underpowered. In the GENOMOS study, which

involved almost 26,000 subjects who had been phenotyped for spine and hip BMD and for the presence of fractures, no association was observed between the *BsmI*, *ApaI* (rs1787935) or *TaqI* (rs1788009) 3′ polymorphisms and BMD or fracture (Uitterlinden et al. 2006). Other polymorphisms at the VDR locus have also been studied in relation to osteoporosis phenotypes, most notably a polymorphism in exon 2 of VDR recognized by the *Fok*I restriction enzyme (rs17881966) which introduces an alternative translational start site yielding two isoforms of VDR, one slightly shorter than the other (Arai et al. 1997; Gross et al. 1998a). The other is rs11568820 which is located in the *VDR* promoter and is thought to affect a binding site for the transcription factor caudal type homeobox 2 (*CDX2*) (Arai et al. 2001). Studies of these polymorphisms in relation to BMD and fracture have yielded inconsistent results although large-scale studies have yielded some evidence for an association between the *CDX2* polymorphism and osteoporosis-related phenotypes (Uitterlinden et al. 2006).

Many investigators have conducted functional analysis of individual *VDR* polymorphisms and haplotypes. Reporter gene constructs prepared from the 3′ region of the VDR gene from different individuals have shown evidence of haplotype-specific differences in gene transcription, raising the possibility that polymorphisms in this region may be involved in regulating mRNA stability (Morrison et al. 1994). In support of this view, cell lines which were heterozygous for the *TaqI* polymorphism showed differences in allele-specific transcription of the *VDR* gene (Verbeek et al. 1997). In this study however, transcripts from the "t" allele were 30% more abundant than the "T" allele which is the opposite from the result expected on the basis of Morrison's results (Morrison et al. 1994). In another study, bone samples from subjects in the MrOS study of community dwelling men aged >65 from the USA showed differences in allele-specific transcription in association with 3′ *VDR* haplotypes (Grundberg et al. 2007). Specifically, carriage of haplotype 1 (baT) was associated with increased *VDR* mRNA abundance, and this haplotype was also associated with an increased risk of fracture in men. In a comprehensive analysis of several cell lines, Fang and colleagues also demonstrated that the baT haplotype was associated with decreased *VDR* mRNA levels (Fang et al. 2005). Other in vitro studies have shown no differences in allele-specific transcription, mRNA stability or ligand binding in relation to the *BsmI* polymorphism (Mocharla et al. 1997; Gross et al. 1998b; Durrin et al. 1999).

Additional functional studies have been carried out on other VDR 3′ variants. In vitro studies have shown that different *VDR FokI* alleles differ in their ability to drive reporter gene expression (Arai et al. 1997; Jurutka et al. 2000), and the polymorphic variant lacking three amino acids ("F") has also been found to interact with human basal transcription factor IIB more efficiently than the longer isoform ("f"). However, other researchers have found no differences between *FokI* alleles in terms of VDR ligand binding, DNA binding or transactivation activity (Gross et al. 1998a). There is good evidence that the *CDX2* polymorphism within the promoter of the *VDR* gene is functional. Arai and colleagues noted that the G allele had reduced affinity for CDX2 protein binding and also had a 70% reduced ability to drive reporter gene expression compared with the A allele (Arai et al. 2001).

In summary, the studies which have been performed to date do not support the hypothesis that allelic variation at the *VDR* locus plays a major role in regulating bone mass or osteoporotic fracture. However, there is evidence to support that some of these *VDR* polymorphisms have functional effects.

3.2.4 Other Regulatory Variants

Emerging data from GWAS suggests that some common variants which predispose to osteoporosis do so by affecting transcriptional regulation of the target genes. Recent GWAS studies found associations between SNPs at the wntless homolog (*WLS*; *GPR177*), myocyte enhancer factor 2C (*MEF2C*), forkhead box C1 (*FOXC1*) and tumor necrosis factor receptor superfamily member 11b (*TNFRSF11B*) loci that were associated with BMD and *cis*-allelic expression of variants at the same loci (Rivadeneira et al. 2009; Richards et al. 2008). At the present time, the molecular mechanisms underlining these associations have not been explored, although it seems likely that polymorphic variations affecting common regulatory elements will be found to underlie the effects mentioned above.

3.3 Conclusions

Current evidence suggests that common genetic variants affecting regulatory elements in the human genome contribute to the pathogenesis of osteoporosis by affecting expression levels of genes that regulate bone cell function and matrix production. Further work will be required to investigate the molecular mechanisms by which susceptibility alleles for osteoporosis exert their effects and to define the extent to which regulatory variants and protein coding variants regulate bone mass, bone structure and other phenotypes relevant to the pathogenesis of osteoporosis.

Abbreviations

BMD	Bone mineral density
BMI	Body mass index
ER	Endoplasmic reticulum
GWAS	Genome-wide association studies
HAL	Skeleton genometry
LRP5	Low-density lipoprotein-related receptor-5 gene
OI	Osteogenesis imperfecta
SD	Standard deviation
SSc	Systemic sclerosis
OPS	Osteoporosis-pseudoglioma syndrome

References

Albagha OM, McGuigan FEA, Reid DM, Ralston SH (2001) Estrogen receptor alpha gene polymorphisms and bone mineral density: haplotype analysis in women from the United Kingdom. J Bone Miner Res 16:128–134

Albagha OM, Pettersson U, Stewart A, McGuigan FE, MacDonald HM, Reid DM, Ralston SH (2005) Association of oestrogen receptor alpha gene polymorphisms with postmenopausal bone loss, bone mass, and quantitative ultrasound properties of bone. J Med Genet 42:240–246

Alvarez M, Long H, Onyia J, Hock J, Xu W, Bidwell J (1997) Rat osteoblast and osteosarcoma nuclear matrix proteins bind with sequence specificity to the rat type I collagen promoter. Endocrinology 138:482–489

Arai H, Miyamoto K-I, Taketani Y, Yamamoto H, Iemori Y, Morita K, Tonai T, Nishisho T, Mori S, Takeda E (1997) A vitamin D receptor gene polymorphism in the translation initiation codon: effect on protein activity and relation to bone mineral density in Japanese women. J Bone Miner Res 12:915–921

Arai H, Miyamoto KI, Yoshida M, Yamamoto H, Taketani Y, Morita K, Kubota M, Yoshida S, Ikeda M, Watabe F, Kanemasa Y, Takeda E (2001) The polymorphism in the caudal-related homeodomain protein Cdx-2 binding element in the human vitamin D receptor gene. J Bone Miner Res 16:1256–1264

Arden NK, Spector TD (1997) Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. J Bone Miner Res 12:2076–2081

Arden NK, Baker J, Hogg C, Baan K, Spector TD (1996) The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. J Bone Miner Res 11:530–534

Balemans W, Van WL, Van HW (2005) A clinical and molecular overview of the human osteopetroses. Calcif Tissue Int 77:263–274

Bornstein P, McKay J, Morishima JK, Devarayalu S, Gelinas RE (1987) Regulatory elements in the first intron contribute to transcriptional control of the human collagen alpha 1 (I) collagen gene. Proc Natl Acad Sci USA 84:8869–8873

Bustamante M, Nogues X, Enjuanes A, Elosua R, Garcia-Giralt N, Perez-Edo L, Caceres E, Carreras R, Mellibovsky L, Balcells S, ez-Perez A, Grinberg D (2007) COL1A1, ESR1, VDR and TGFB1 polymorphisms and haplotypes in relation to BMD in Spanish postmenopausal women. Osteoporos Int 18:235–243

Byers PH (2000) Osteogenesis imperfecta: perspectives and opportunities. Curr Opin Pediatr 12:603–609

Cummings SR, Melton LJ (2002) Epidemiology and outcomes of osteoporotic fractures. Lancet 359:1761–1767

De Laet CE, van Hout BA, Burger H, Hofman A, Pols HA (1997) Bone density and risk of hip fracture in men and women: cross sectional analysis [published erratum appears in BMJ 1997 Oct 11; 315(7113):916]. Br Med J 315:221–225

Durrin LK, Haile RW, Ingles SA, Coetzee GA (1999) Vitamin D receptor 3′-untranslated region polymorphisms: lack of effect on mRNA stability. Biochim Biophys Acta 1453:311–320

Fang Y, van Meurs JB, d'Alesio A, Jhamai M, Zhao H, Rivadeneira F, Hofman A, van Leeuwen JP, Jehan F, Pols HA, Uitterlinden AG (2005) Promoter and 3′-untranslated-region haplotypes in the vitamin D receptor gene predispose to osteoporotic fracture: the Rotterdam study. Am J Hum Genet 77:807–823

Garcia-Giralt N, Nogues X, Enjuanes A, Puig J, Mellibovsky L, Bay-Jensen A, Carreras R, Balcells S, Diez-Perez A, Grinberg D (2002) Two new single nucleotide polymorphisms in the COLIA1 upstream regulatory region and their relationship with bone mineral density. J Bone Miner Res 17:384–393

Garcia-Giralt N, Enjuanes A, Bustamante M, Mellibovsky L, Nogues X, Carreras R, ez-Perez A, Grinberg D, Balcells S (2005) In vitro functional assay of alleles and haplotypes of two COL1A1-promoter SNPs. Bone 36:902–908

Garnero P, Borel O, Grant SFA, Ralston SH, Delmas PD (1998) Collagen I a 1 polymorphism, bone mass and bone turnover in healthy French pre-menopausal women: The OFELY study. J Bone Miner Res 13:813–818

Grant SFA, Reid DM, Blake G, Herd R, Fogelman I, Ralston SH (1996) Reduced bone density and osteoporosis associated with a polymorphic Sp1 site in the collagen type I alpha 1 gene. Nat Genet 14:203–205

Gross C, Krishnan AV, Malloy PJ, Eccleshall TR, Zhao XY, Feldman D (1998a) The vitamin D receptor gene start codon polymorphism: a functional analysis of FokI variants. J Bone Miner Res 13:1691–1699

Gross C, Musiol IM, Eccleshall TR, Malloy PJ, Feldman D (1998b) Vitamin D receptor gene polymorphisms: analysis of ligand binding and hormone responsiveness in cultured skin fibroblasts. Biochem Biophys Res Commun 242:467–473

Grundberg E, Lau EM, Pastinen T, Kindmark A, Nilsson O, Ljunggren O, Mellstrom D, Orwoll E, Redlund-Johnell I, Holmberg A, Gurd S, Leung PC, Kwok T, Ohlsson C, Mallmin H, Brandstrom H (2007) Vitamin D receptor 3′ haplotypes are unequally expressed in primary human bone cells and associated with increased fracture risk: the MrOS Study in Sweden and Hong Kong. J Bone Miner Res 22:832–840

Gueguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, Siest G (1995) Segregation analysis and variance components analysis of bone mineral density in healthy families. J Bone Miner Res 12:2017–2022

Herrington DM, Howard TD, Brosnihan KB, McDonnell DP, Li X, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Meyers DA, Bleecker ER (2002) Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. Circulation 105:1879–1882

Hunter D, de Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker RV, Spector TD (2001) Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J Bone Miner Res 16:371–378

Ioannidis JP, Stavrou I, Trikalinos TA, Zois C, Brandi ML, Gennari L, Albagha O, Ralston SH, Tsatsoulis A (2002) Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis. J Bone Miner Res 17:2048–2060

Janssens K, Van Hul W (2002) Molecular genetics of too much bone. Hum Mol Genet 11:2385–2393

Jin H, Stewart TL, Hof RV, Reid DM, Aspden RM, Ralston S (2009a) A rare haplotype in the upstream regulatory region of COL1A1 is associated with reduced bone quality and hip fracture. J Bone Miner Res 24:448–454

Jin H, van't Hof RJ, Albagha OM, Ralston SH (2009b) Promoter and intron 1 polymorphisms of COL1A1 interact to regulate transcription and susceptibility to osteoporosis. Hum Mol Genet 18:2729–2738

Jin H, Evangelou E, Ioannidis JP, Ralston SH (2011) Polymorphisms in the 5′ flank of COL1A1 gene and osteoporosis: meta-analysis of published studies. Osteoporos Int 22:911–921

Jurutka PW, Remus LS, Whitfield GK, Thompson PD, Hsieh JC, Zitzer H, Tavakkoli P, Galligan MA, Dang HT, Haussler CA, Haussler MR (2000) The polymorphic N terminus in human vitamin D receptor isoforms influences transcriptional activity by modulating interaction with transcription factor IIB. Mol Endocrinol 14:401–420

Kanis JA, Melton LJ III, Christiansen C, Johnston CC, Khaltaev N (1994) The diagnosis of osteoporosis. J Bone Miner Res 9:1137–1141

Kaprio J, Rimpela A, Winter T, Viken RJ, Rimpela M, Rose RJ (1995) Common genetic influences on BMI and age at menarche. Hum Biol 67:739–753

Keen RW, Woodford-Richens KL, Grant SF, Ralston SH, Lanchbury JS, Spector TD (1999) Association of polymorphism at the type I collagen (COL1A1) locus with reduced bone mineral density, increased fracture risk, and increased collagen turnover. Arthritis Rheum 42:285–290

Langdahl BL, Ralston SH, Grant SFA, Eriksen EF (1998) An Sp1 binding site polymorphism in the COLIA1 gene predicts osteoporotic fractures in men and women. J Bone Miner Res 13:1384–1389

Mann V, Ralston SH (2003) Meta-analysis of COL1A1 Sp1 polymorphism in relation to bone mineral density and osteoporotic fracture. Bone 32:711–717

Mann V, Hobson EE, Li B, Stewart TL, Grant SF, Robins SP, Aspden RM, Ralston SH (2001) A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. J Clin Invest 107:899–907

Maruyama H, Toji H, Harrington CR, Sasaki K, Izumi Y, Ohnuma T, Arai H, Yasuda M, Tanaka C, Emson PC, Nakamura S, Kawakami H (2000) Lack of an association of estrogen receptor alpha gene polymorphisms and transcriptional activity with Alzheimer disease. Arch Neurol 57:236–240

Michaelsson K, Melhus H, Ferm H, Ahlbom A, Pedersen NL (2005) Genetic liability to fractures in the elderly. Arch Intern Med 165:1825–1830

Misof K, Landis WJ, Klaushofer K, Fratzl P (1997) Collagen from the osteogenesis imperfecta mouse model (oim) shows reduced resistance against tensile stress. J Clin Invest 100:40–45

Mocharla H, Butch AW, Pappas AA, Flick JT, Weinstein RS, De Togni P, Jilka RL, Roberson PK, Parfitt AM, Manolagas SC (1997) Quantification of vitamin D receptor mRNA by competitive polymerase chain reaction in PBMC: lack of correspondence with common allelic variants. J Bone Miner Res 12:726–733

Morrison NA, Yeoman R, Kelly PJ, Eisman JA (1992) Contribution of trans-acting factor alleles to normal physiological variability: vitamin D receptor gene polymorphisms and circulating osteocalcin. Proc Natl Acad Sci USA 89:6665–6669

Morrison NA, Qi JC, Tokita A, Kelly P, Crofts L, Nguyen TV, Sambrook PN, Eisman JA (1994) Prediction of bone density from vitamin D receptor alleles. Nature 367:284–287

Morrison NA, Qi JC, Tokita A, Kelly P, Crofts L, Nguyen TV, Sambrook PN, Eisman JA (1997) Prediction of bone density from vitamin D receptor alleles (Erratum). Nature 387:106

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B (2002) The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. Cell 108:17–29

Pocock NA, Eisman JA, Hopper JL, Yeates MG, Sambrook PN, Eberl S (1987) Genetic determinants of bone mass in adults: a twin study. J Clin Invest 80:706–710

Qureshi AM, McGuigan FEA, Seymour DG, Hutchison JD, Reid DM, Ralston SH (2001) Association between COLIA1 Sp1 alleles and femoral neck geometry. Calcified Tissue Int 69:67–72

Qureshi AM, Herd RJ, Blake GM, Fogelman I, Ralston SH (2002) COLIA1 Sp1 polymorphism predicts response of femoral neck bone density to cyclical etidronate therapy. Calcified Tissue Int 70:158–163

Ralston SH, Uitterlinden AG, Brandi ML, Balcells S, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Scollen S, Bustamante M, Husted LB, Carey AH, Diez-Perez A, Dunning AM, Falchetti A, Karczmarewicz E, Kruk M, van Leeuwen JPTM, Meurs JB, Mangion J, McGuigan FE, Mellibovsky L, Monte FD, Pols HA, Reeve J, Reid DM, Renner W, Rivadeneira F, Schoor NM, Sherlock RE, Ioannidis JP (2006) Large-scale evidence for the effect of the COLIA1 Sp1 polymorphism on osteoporosis outcomes: The GENOMOS study. PLoS Med 3:e90

Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, Andrew T, Falchi M, Gwilliam R, Ahmadi KR, Valdes AM, Arp P, Whittaker P, Verlaan DJ, Jhamai M, Kumanduri V, Moorhouse M, van Meurs JB, Hofman A, Pols HA, Hart D, Zhai G, Kato BS, Mullin BH, Zhang F, Deloukas P, Uitterlinden AG, Spector TD (2008) Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 371:1505–1512

Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, Zillikens MC, Kavvoura FK, Amin N, Aulchenko YS, Cupples LA, Deloukas P, Demissie S, Grundberg E, Hofman A, Kong A, Karasik D, van Meurs JB, Oostra B, Pastinen T, Pols HA, Sigurdsson G, Soranzo N, Thorleifsson G, Thorsteinsdottir U, Williams FM, Wilson SG, Zhou Y, Ralston SH, van Duijn CM, Spector T, Kiel DP, Stefansson K, Ioannidis JP, Uitterlinden AG (2009) Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat Genet 41:1119–1206

Robling AG, Childress P, Yu J, Cotte J, Heller A, Philip BK, Bidwell JP (2009) Nmp4/CIZ suppresses parathyroid hormone-induced increases in trabecular bone. J Cell Physiol 219:734–743

Sims NA, Dupont S, Krust A, Clement-Lacroix P, Minet D, Resche-Rigon M, Gaillard-Kelly M, Baron R (2002) Deletion of estrogen receptors reveals a regulatory role for estrogen receptors-beta in bone remodeling in females but not in males. Bone 30:18–25

Snieder H, MacGregor AJ, Spector TD (1998) Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. J Clin Endocrinol Metab 83:1875–1880

Stewart TL, Roschger P, Misof BM, Mann V, Fratzl P, Klaushofer K, Aspden RM, Ralston SH (2005) Association of *COLIA1* Sp1 alleles with defective bone nodule formation *in vitro* and abnormal bone mineralisation *in vivo*. Calcified Tissue Int 77:113–118

Stewart TL, Jin H, McGuigan FE, Albagha OM, Garcia-Giralt N, Bassiti A, Grinberg D, Balcells S, Reid DM, Ralston SH (2006) Haplotypes defined by promoter and intron 1 polymorphisms of the COLIA1 gene regulate bone mineral density in women. J Clin Endocrinol Metab 91:3575–3583

Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, Jonsdottir T, Saemundsdottir J, Center JR, Nguyen TV, Bagger Y, Gulcher JR, Eisman JA, Christiansen C, Sigurdsson G, Kong A, Thorsteinsdottir U, Stefansson K (2008) Multiple genetic loci for bone mineral density and fractures. N Engl J Med 358:2355–2365

Uitterlinden AG, Burger H, Huang Q, Yue F, McGuigan FEA, Grant SFA, Hofman A, van Leeuwen JPTM, Pols HAP, Ralston SH (1998) Relation of alleles of the collagen type I a 1 gene to bone density and risk of osteoporotic fractures in postmenopausal women. N Engl J Med 338:1016–1022

Uitterlinden AG, Ralston SH, Brandi ML, Carey AH, Grinberg D, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Reeve J, Reid DM, Amidei A, Bassiti A, Bustamante M, Husted LB, ez-Perez A, Dobnig H, Dunning AM, Enjuanes A, Fahrleitner-Pammer A, Fang Y, Karczmarewicz E, Kruk M, van Leeuwen JP, Mavilia C, van Meurs JB, Mangion J, McGuigan FE, Pols HA, Renner W, Rivadeneira F, van Schoor NM, Scollen S, Sherlock RE, Ioannidis JP (2006) The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. Ann Intern Med 145:255–264

Verbeek W, Gombart AF, Shiohara M, Campbell M, Koeffler HP (1997) Vitamin D receptor: no evidence for allele-specific mRNA stability in cells which are heterozygous for the Taq I restriction enzyme polymorphism. Biochem Biophys Res Commun 238:77–80

Chapter 4
Gene Regulation in Van Buchem Disease

Gabriela G. Loots

Abstract Van Buchem disease (VB) is a rare autosomal recessive disorder in which progressive bone overgrowth leads to very dense bones, distortion of the face, and entrapment of cranial nerves. It uniquely stands out as a congenital disorder likely to be caused by noncoding mutations for several reasons: (1) it maps to the same locus on human chromosome 17q12–21 as a highly similar disorder, sclerosteosis; (2) several single specific mutations have been identified in sclerosteosis patients that all predict null alleles in the determinant gene, sclerostin or *SOST*; (3) no coding mutations in *SOST* have been identified in VB patients; and (4) all VB patients carry a homozygous 52-kb noncoding deletion downstream of the *SOST* transcript. Here, we describe how by using comparative sequence analysis, BAC recombination, and enhancer assays, in combination with the generation of transgenic and knockout mice, it has been shown that human *SOST* is essential for maintaining healthy bone metabolism and that VB disease is caused by a noncoding deletion that removes a *SOST*-specific regulatory element, ECR5.

Keywords High bone mass • Osteopetrosis • SOST • Sclerostin • Van Buchem disease • Sclerosteosis • Van Buchem deletion • ECR5 enhancer

G.G. Loots (✉)
Biology and Biotechnology Division, Lawrence Livermore National Laboratory, 7000 East Avenue, L-452, Livermore, CA 94550, USA

School of Natural Sciences, University of California Merced, Merced, CA 95343, USA
e-mail: loots1@llnl.gov; gloots@ucmerced.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_4,

4.1 Introduction

4.1.1 Clinical and Radiological Features of Sclerosteosis and Van Buchem Disease

Genetic disorders affecting the skeleton are rare, and they comprise a large group of clinically distinct and genetically heterogeneous conditions. Clinically, they can range from neonatal lethal to only mild growth retardation, and their clinical diversity makes them difficult to diagnose (Kornak and Mundlos 2003). In general, skeletal disorders have been subdivided into dysostoses, defined by the type of malformations observed for individual groups of bones or as osteochondrodysplasias, defined as developmental disorders of chondro-osseous tissues.

One category of skeletal dysplasias is represented by disorders of skeletal homeostasis. From the moment the bone is formed, two types of cells, osteoblasts and osteoclasts, simultaneously contribute to maintaining the integrity of the skeleton by forming or resorbing bone, respectively. This processes termed bone remodeling functions both to preserve the overall bone mass as well as to restructure the bone in response to metabolic and mechanical needs of the organism (Kornak and Mundlos 2003). In general, these disorders interfere with bone balance such that low (osteoporosis) or high (osteopetrosis) bone mass phenotypes can arise. Here, we introduce two conditions mediated by osteoblast dysfunction: Van Buchem disease (VB) and sclerosteosis which result in generalized high bone mass (HBM) due to overactive osteoblast activity. In particular, we will review recent data that establishes that VB is due to the removal of a transcriptional regulatory element, ECR5 (evolutionary conserved region 5), that is essential for the transcriptional activation of the causative gene, *SOST*, in the skeleton.

4.1.2 Van Buchem Disease

Van Buchem (VB) disease or hyperostosis corticalis generalisata (MIM 239100) is a rare autosomal recessive bone dysplasia first described in 1955 by Van Buchem et al. (1955). Clinically and radiographically, the disorder manifests itself as massive hyperostosis of the calvarium and mandible, mild sclerosis of the spine, and increased radiographic density and cortical thickening of the long bones of the arms and legs. On average, bone overgrowth in VB results in very high bone mineral density that leads to a skeleton that is 3–4 times heavier than normal (Fig. 4.1a) (Janssens and Van Hul 2002). As a consequence to bone overgrowth, VB patients display facial distortion (Fig. 4.1b, c) accompanied by deafness and facial palsy, directly due to bone entrapment of the seventh and eighth cranial nerves. The prevalence of the disorder is very low; in 2002, it was estimated that only about 30 individuals have been diagnosed worldwide, predominantly from Dutch ancestry (Staehling-Hampton et al. 2002).

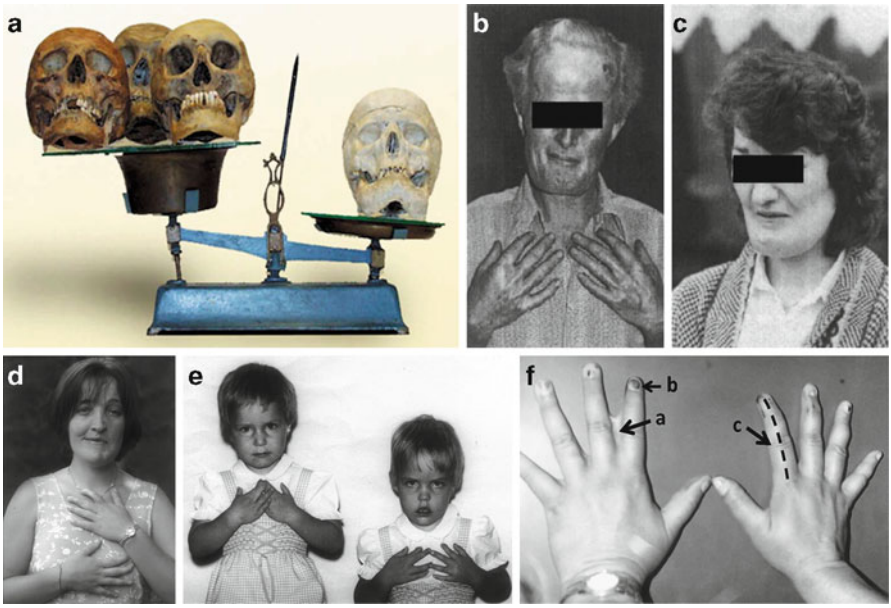

Fig. 4.1 *Clinical representation of patients with Van Buchem and Sclerosteosis disease.* Frontal views (**b**–**e**) and hands (**b**, **d**–**f**). To illustrate the impressive gain in weight by adding mineral density, Janssens and Van Hul have previously shown a VB skull which weighs close to 4 times that of a normal skull (**a**) (Janssens and Van Hul 2002). Facial characteristics of VB include a high forehead, protruding large chin, and partial face paralysis on the *left side* (**b**–**c**). No syndactyly has ever been documented for VB patients (**b**). The adult in this picture is a 30-year-old female that shows facial characteristics of sclerosteosis with a high forehead, protruding large chin, and partial face paralysis on the *left side*. Her syndactyly has been corrected, and she also underwent craniectomy and mandibular reduction. She is 5 ft 9 in. tall and she weighs 187 lb (**d**). The children are 4- and 3-year-old siblings, who have unilateral digits 2/3 syndactyly. Both have already experienced intermittent facial palsy, and already at this young age, they are distinguished by large foreheads and protruding mandibles (**e**). In addition to syndactyly (a), sclerosteosis patients also display nail dysplasia (b) and radial deviation of the phalanges (c) (Hamersma et al. 2003)

4.1.3 Sclerosteosis

Similar to VB, sclerosteosis (MIM 269500) is also an autosomal recessive skeletal dysplasia characterized by the presence of generalized skeletal overgrowth which is more severe than VB. Patients have been described to display enlarged skulls and mandibles (Fig. 4.1d, e) and to be very tall. The average height for affected men was 6 ft 4 in. (194 cm; range 178–207 cm) and 5 ft 10 in. (180 cm; range 168–190 cm) for women (Hamersma et al. 2003). These individuals have a normal body fat index but have excessive weight, averaging 185 lb (lb) (83–85 kg) for men and women, due only to high skeletal weight. Intracranial pressure due to bone overgrowth is more serious in sclerosteosis than in VB patients, where sudden death frequently occurs (Hamersma et al. 2003; Balemans et al. 2001). This condition is also very

rare, with less than 100 documented cases, and most of the affected individuals are among the Afrikaner community in South Africa, who are also of Dutch ancestry. A recent study by Hamersma et al. in 2003 described data collected for 63 affected individuals (29 females and 34 males) corresponding to 38 families, 8 of which were consanguineous. These individuals were observed for the course of 38 years, during which 34 of the 63 patients died primarily as a direct result of elevated intracranial pressure (Hamersma et al. 2003). Facial palsy and deafness were evident as early as 5 years of age, and 82% of the affected individuals developed these symptoms in childhood.

4.1.4 Differences Between Van Buchem Disease and Sclerosteosis

One characteristic that sets VB apart from sclerosteosis is the absence of hand abnormalities. Among the 63 sclerosteosis individuals described by Hamersma et al., 48 had syndactyly of the digits, 41 of which represented a fusion of the index and middle fingers (digits 2/3; Fig. 4.1e, f). The extent of syndactyly ranges from minor soft tissue webbing to bone fusions. Radiographs of these individuals reveal marked cortical hyperostosis, where most of the tubular bones are widened and irregular. Another observed anomaly is radial deviation of phalanges, where the bones curve away from the axis of the body. While no digit abnormalities are noted on the toes, these patients do have dysplastic or absent nails on both hands and toes. Soft tissue syndactyly and abnormal nails suggest that the hand defects in sclerosteosis are in part a syndrome that affects derivatives of the ectoderm. In contrast to sclerosteosis, syndactyly has never been documented for VB patients (Staehling-Hampton et al. 2002).

4.2 Genetics of VB and Sclerosteosis

Despite the radiographic and clinical similarities between VB and sclerosteosis, these two craniotubular hyperostoses were originally classified as two distinct sclerosteoses. It was only in 1984 that Beighton et al. first suggested that they are allelic because VB displays milder characteristics of sclerosteosis, and hence they may potentially be caused by hyper- and hypomorphic allelic versions of the same gene (Beighton et al. 1984). A genome-wide linkage study with 391 highly polymorphic microsatellite markers in 11 Van Buchem patients from a highly inbred family, in a small ethnic isolate of the Netherlands, suggested that VB is linked to a chromosomal region around marker D17S1299 with a LOD score of 8.82 at a recombination frequency of 0.01 (Van Hul et al. 1998). This region was further refined to a 1 centimorgan (cM) segment encompassing many genes on human chromosome 17q12–21, as the

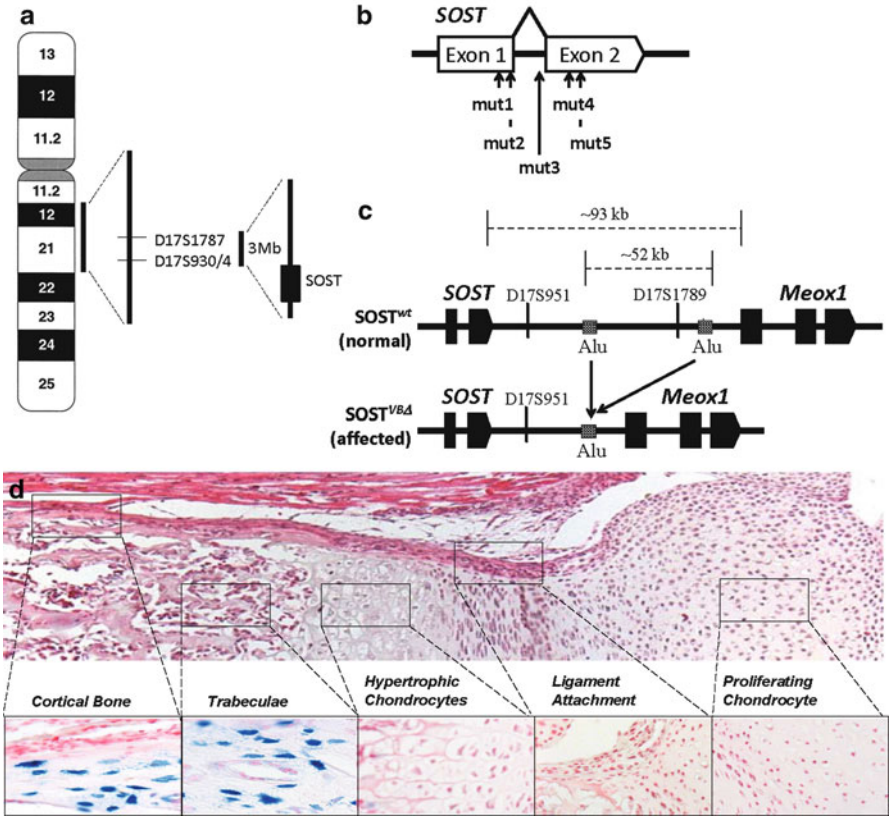

Fig. 4.2 *Mapping of Van Buchem disease and sclerosteosis to the same region on human chromosome 17q12–21 and genetic mutations specific for the two disorders.* Genetic map of human chromosome 17 showing the initial sclerosteosis region mapped on 17q12–21 and the subsequent narrowing of the region between D17S1787 and D17S930/4 interval that contains the *SOST* transcript (**a**). Mutation analysis of the *SOST* gene identified several mutations (mut1–5) that either interfere with the splice site (mut2–3) or cause premature termination (mut1; mut4–5) that cause sclerosteosis (**b**). A homozygous 52-kb deletion was identified downstream of *SOST* in all Van Buchem patients examined, within the *SOST-MEOX1* intergenic region. The deletion was flanked by Alu repeats, suggesting that the deletion was derived by intrachromosomal recombination (**c**). A 158-kb human BAC ($SOST^{wt}$) spanning *SOST* and *MEOX1* was modified by in vitro BAC recombination to delete the 52-kb noncoding region absent in VB patients ($SOST^{VB\Delta}$) (**c**) (Loots et al. 2005). SOST-LacZ knockin mice were used to determine *SOST* expression based on β-galactosidase activity. The mouse humerus was sectioned and processed for LacZ expression. Expression was detected only in osteocytes (**d**)

candidate region for VB (Van Hul et al. 1998). Shortly thereafter, two unrelated families from Brazil and the United States were used to map the sclerosteosis causative gene to the same genomic interval, strengthening the argument that these two disorders are due to mutations in the same gene (Balemans et al. 1999) (Fig. 4.2a).

The VB gene was later mapped to a 0.7-cM region between markers D17S1787 and D17S934 (Van Hul et al. 1998), and the sclerosteosis candidate gene was subsequently mapped to ~1-Mb region between markers D17S1326 and D17S1860 by Balemans et al. (2001) and to an ~3-Mb region between markers D17S1787 and D17S930 by Brunkow et al. (2001) (Fig. 4.2a). Through random shotgun sequencing and mapping to bacterial artificial chromosome (BAC) clones across the D17S1787/D17S930 interval, Brunkow et al. predicted the presence of ~170 genes within the 3-Mb region, ~100 of which were not previously known. Using PCR amplification of ~1,000 predicted exons in search of sclerosteosis-specific mutations, they compared DNA sequences from an affected individual, an obligate carrier, and an unrelated, unaffected control. A polymorphism in a novel gene later designated as *SOST* or sclerostin displayed 100% concordance with the sclerosteosis chromosome when 29 affected individuals, 33 obligate carriers, and 24 unaffected siblings were analyzed, strongly suggesting that *SOST* is the likely determinant gene (Brunkow et al. 2001). Several causative mutations were subsequently identified: A single base substitution (Fig. 4.2b; mut1; C69T) located 69 base pairs (bp) downstream of the predicted translation initiation site was founded in affected Afrikaners; isolated Senegal patients displayed a homozygous splice site mutation (Fig. 4.2b; mut2; IVS1+3 A→T); two substitutions within the intronic sequence were found in another affected Senegalese individual (Fig. 4.2b; mut3); two nonsense mutations in exon two were found in a Brazilian (Fig. 4.2b; mut4; Trp124X) and an American family (Fig. 4.2b; mut5; Arg126X). All sclerosteosis mutations identified to date result in premature termination of the *SOST* transcript (nonsense mutations) or fail to splice the second exon properly resulting in what are believed to be "null" *SOST* alleles.

Several VB patients were examined for sequence variation in the *SOST* gene. The entire 5-kb locus of *SOST* (including 1 kb upstream of the gene, the 3′ and 5′ UTRs, the two exons and the intron) was examined by Balemans et al. in three Dutch VB patients (Balemans et al. 2002) and by Brunkow et al. in 15 family members of a VB family in the Netherlands (Staehling-Hampton et al. 2002), all not finding any candidate *SOST* mutations. To date, no *SOST* mutations have ever been identified in any Van Buchem family member (Staehling-Hampton et al. 2002; Balemans et al. 2001, 1999; Brunkow et al. 2001). However, what both groups identified was the presence of a homozygous noncoding deletion (Fig. 4.2c) downstream of the *SOST* transcript (Staehling-Hampton et al. 2002; Balemans et al. 2002). This discovery came about through the lack of amplification of genetic markers along the 1-Mb region of human chromosome 17q12–21 from the VB DNA samples (Fig. 4.2a), while normal amplification was observed in healthy controls as well as from a human BAC clone (RP11–209M4). Using the BAC clone, it was reasonable to estimate the location of the two genes, *SOST* and *MEOX1* ~93 kb apart as well as to determine that the missing region in VB spanned ~52 kb (Fig. 4.2c). Through an amplicon walking strategy, the endpoints of the deletion were identified and found to contain identical 16-bp sequences representative of Alu repetitive sequences. The presence of homologous sequences flanking the Van Buchem deletion hinted that the possible mechanism that generated the deletion is through homologous recombination

(Staehling-Hampton et al. 2002; Balemans et al. 2002) (Fig. 4.2c). All VB patients to date have been shown to be homozygous for this 52-kb noncoding deletion (approximate location on hg19 chr17:41796600–41744700).

4.3 Sclerostin Protein and Its Expression Pattern

The *SOST* gene includes two exons that encompass a 639-bp coding sequence in addition to 47-bp 5′ and a 1,615-bp 3′ UTR. Amino acid sequence analysis of *SOST* identified a putative secretion signal and two N-glycosylation sites similar to secreted proteins. Comparisons to other known protein domains revealed weak but significant *SOST* similarity to gastric mucin, the beta subunit of luteinizing hormone, DAN and PRDC (*P*rotein *R*elated to *D*an and *C*erberus) corresponding to a cysteine-rich region between *SOST* residues 80–167 (Brunkow et al. 2001). Further analysis of this domain concluded that sclerostin likely belongs to a family of secreted proteins that contain this structural cysteine knot motif and include the TGF-β superfamily, the Norrie disease protein (NDP), the mucins, and the von Willebrand factor and that it is most closely related to DAN, cerberus, gremlin, PRDC, and caronte proteins.

SOST expression in adult human samples was found to be in whole long bones, cartilage, kidney, and liver. In embryonic/fetal human samples, *SOST* was detected in the placenta, fetal skin, aorta, and fetal kidney. Semiquantitative reverse transcriptase PCR (rtPCR) of mouse tissues detected significant *SOST* levels in whole fetus, liver, heart, kidney, brain, thymus, and whole long bone (Brunkow et al. 2001). Despite the fact that *SOST* has been detected in non-bone tissues, the last decade of research has primarily focused on *SOST* bone expression and its function in the skeleton. Within bone, *SOST* is robustly expressed in osteocytes (Fig. 4.2d) of the cortical and endochondral bone (Winkler et al. 2003).

4.4 Characterizing the Van Buchem Deletion Region

The discovery of the VB deletion prompted several hypotheses in regards to the underlying genetic causes of the disease: VB is caused by the disruption of a novel (non-*SOST*) gene or the VB deletion dysregulates the transcription of *SOST* or other nearby gene. Extensive computational and molecular characterization of the complete 93-kb *SOST-MEOX1* intergenic region through gene prediction, exon trapping, Northern blot analysis, cDNA library screening, and rtPCR approaches were all unsuccessful at identifying a novel causative gene. In general, most of the short expressed sequences within this region represented nonspecific, low-abundance repetitive elements, and no experimental approach was able to authenticate the predicted transcripts as bona fide novel protein encoding transcripts (Staehling-Hampton et al. 2002; Balemans et al. 2002).

The shared clinical similarities between VB and sclerosteosis along with their strong genetic linkage to the same locus on chromosome 17q12–21, now accompanied by the lack of evidence that another transcript within the VB region may cause the disease, further strengthened the alternative hypothesis that the two hyperostoses are allelic. Loots et al. proposed that the deletion in VB patients removes an enhancer element essential for directing the expression of human *SOST* in the adult skeleton (Loots et al. 2005). To characterize the transcriptional regulation of *SOST*, they proceeded to characterize the expression of human *SOST* in transgenic mice carrying either a normal ($SOST^{wt}$) or a VB allele ($SOST^{VB\Delta}$) (Fig. 4.2c). They were able to show that only the $SOST^{wt}$ allele faithfully expressed human *SOST* in the adult bone and impacted bone metabolism, consistent with the model that the VB deletion (VBΔ) removes a *SOST*-specific regulatory element.

4.4.1 Molecular and Phenotypic Characterization of Van Buchem Transgenic Mouse Models

An ~158-kb human BAC (RP11-209M4) ($SOST^{wt}$), encompassing the 3′ end of the *DUSP3* gene, *SOST, MEOX1,* and the ~93-kb noncoding intergenic interval separating *SOST* from its neighbor, was engineered using homologous recombination in bacteria (Lee et al. 2001) to delete the 52-kb region missing in VB patients and to create a construct that resembles the allele present in VB patients ($SOST^{VB\Delta}$) (Fig. 4.2c). Transgenic mice were generated using these two BACs ($SOST^{wt}$ and $SOST^{VB\Delta}$), and the *SOST* expression from the human BAC was compared to the endogenous mouse *SOST* and the reported human *SOST* expression (Balemans et al. 2001; Brunkow et al. 2001). $SOST^{wt}$ transgenic animals reliably expressed human *SOST* in the mineralized bone of neonatal and adult mice (skull, rib, and femur), while $SOST^{VB\Delta}$ mice had dramatically reduced levels of human *SOST* mRNA expression, as determined by rtPCR and qPCR. These results demonstrated that in vivo, the VB allele confers dramatically reduced *SOST* expression in the adult bone strengthening the argument that the VBΔ region contains an essential *SOST*-specific regulatory element or elements required for *SOST* bone expression (Loots et al. 2005).

Consistent with the differential *SOST* expression observed between $SOST^{wt}$ and $SOST^{VB\Delta}$ mice in the adult skeleton, $SOST^{wt}$ transgenic animals developed osteopenia due to decreased bone mineral density (BMD) in the appendicular and axial skeleton. Micro-computed tomography (μCT) analysis of three-dimensional cancellous bone structures revealed that the mice had decreased bone volume, trabecular number, thickness, and increased trabecular separation (Loots et al. 2005). In contrast, the bone parameters of $SOST^{VB\Delta}$ transgenics were indistinguishable from non-transgenic, age-matched littermate controls.

It was also determined that the observed osteopenia was gene dose dependent. $SOST^{wt}$ transgenic mice bred to homozygosity revealed a further dramatic decrease

in tibial cancellous bone volume. The bone formation rates at skeletal maturity were also much lower, as reflected by a significant decrease in fluorochrome marker uptake into the mineralizing bone, both in cancellous and cortical bone in the appendicular and the axial skeleton (Loots et al. 2005). In contrast to SOST^{wt} transgenics, SOST$^{VB\Delta}$ animals bred to homozygosity and continued to maintain bone parameters indistinguishable from wild-type controls. These data demonstrated that overexpression of human *SOST* under the control of its own proximal promoter elements, in concert with the downstream VB region, negatively modulates adult bone mass. In contrast, bone mass was unaffected in transgenic animals lacking the 52-kb VB region consistent with the model that Van Buchem disease is caused by the removal of a *SOST*-specific regulatory element required for *SOST* skeletal expression (Loots et al. 2005).

4.4.2 Comparative Sequence Analysis of VB Region and Enhancer Assays

It has previously been shown that transcriptional regulatory sequences tend to be highly conserved across species and was therefore suggested that comparative sequence analysis can be employed as an effective strategy for prioritizing candidate regulatory elements (Loots et al. 2000). When the ~140-kb human *SOST* region was aligned to the corresponding orthologous mouse region from chromosome 11, seven evolutionarily conserved regions (ECR2–8; ECR1 was immediately disregarded because it overlapped with a repetitive element) were identified within the VBΔ genomic interval (conservation criteria ≥80% identify for at least 200bp). Testing of ECR2–8 for their ability to stimulate a heterologous (SV40) as well as the endogenous human *SOST* promoter in osteoblastic (UMR-106) and kidney (293)-derived cell lines highlighted one conserved element, ECR5 (Fig. 4.3a), as the candidate *SOST*-specific enhancer responsible for dictating the osteoblastic lineage-specific *SOST* transcription. Consistently, this element boosted transcription in vitro ~3–4-fold above the background level of the endogenous or heterologous promoter only in the osteoblastic UMR-106 cell line (Loots et al. 2005).

This element has also been tested for enhancer activity in vivo, in transgenic mice. Initially, the ECR5-*hsp68-LacZ* transgene was examined in E14.5 mouse embryos using transient transgenic mice (Nobrega et al. 2003). *LacZ* expression was detected in the cartilage of the ribs, vertebrae, and skull plates, and the expression was identical in all positive transgenic embryos obtained from two independent injections (Loots et al. 2005). Two additional transgenic constructs were also tested in mouse transgenic lines. One construct included a larger 2-kb ECR5 fragment in combination with the beta-globin promoter driving the *LacZ* reporter (Fig. 4.3b). A second transgenic construct included three copies of the most conserved 255-bp region of ECR5 in combination with the human *SOST* promoter driving a green

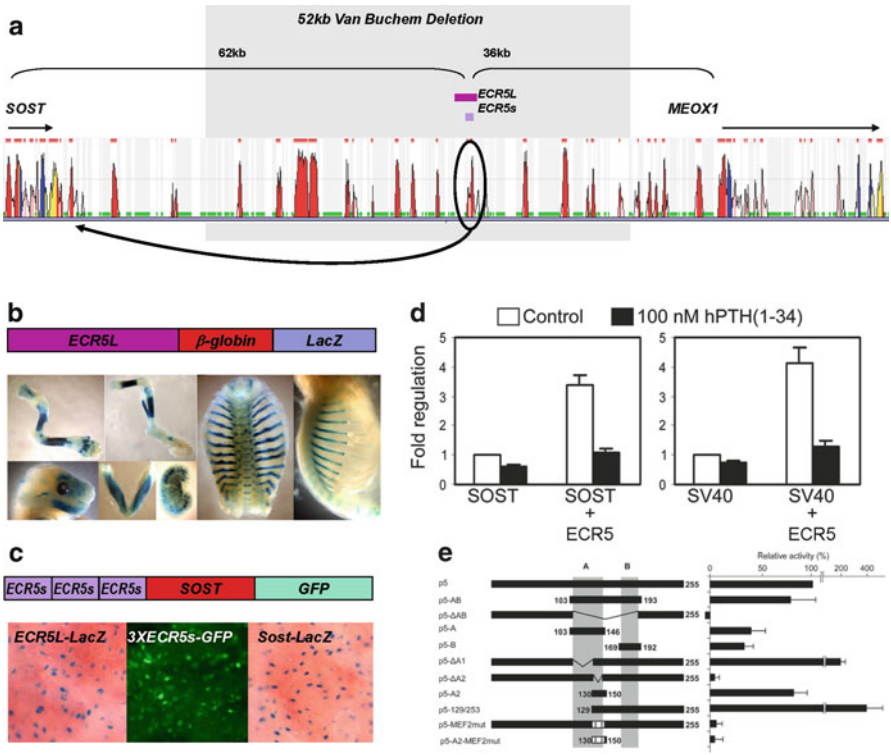

Fig. 4.3 *ECR5 activates transgenic expression in the mouse skeleton and is controlled by PTH and Mef2 transcription factors.* Using comparative sequence analysis, several conserved elements present in the VB deletion region (**a**) were initially tested in vitro using transient transgenic transfections in UMR 106 cells. ECR5 consistently enhanced expression ~3–4-fold. A 2-kb ECR5 (ECR5L) fragment was sufficient to drive *LacZ* reporter expression in the mouse skeleton from the β-globin promoter (**b**). The tissue-specific expression was reproduced when three copies of a shorter, 255-bp ECR5 fragment (ECR5s) was used in combination with the 2-kb human *SOST* promoter and GFP (**c**). Comparing the *ECR5-LacZ*, *3XECR5s-GFP* with the *LacZ* knocked into the mouse *SOST* locus showed expression in osteocytes in all three samples (**c**). UMR-106 cells transfected with reporter plasmids with or without ECR5 [2-kb *SOST* proximal promoter (*left*) or SV40 promoter (*right*) constructs] were treated with 100-nM PTH (1–34) (*filled bars*) or solvent control (*open bars*) 8 h post-transfection, and luciferase activity was measured 16 h posttreatment (**d**). Constructs depicted in (**e**) were tested for transcriptional activation of luciferase reporter in UMR-106 cells. Percent luciferase activity was described relative to the level of luciferase activity obtained with a wild-type construct of ECR5 (p5). The two protein-binding regions *A* and *B* identified by footprint analysis are highlighted in *gray* (**e**) (Leupin et al. 2007). Data in **d**–**e** represents the mean and standard deviations from five independent experiments

fluorescent protein (GFP) (Fig. 4.3c). Both these constructs expressed the reporter gene in the neonate and adult mouse skeleton (Fig. 4.3b), and the expression was primarily in osteoblasts and osteocytes of the calvarium and long bones, consistent with the endogenous mouse *SOST* expression (Fig. 4.3c).

4.4.3 Regulation of ECR5 by Parathyroid Hormone (PTH) and Mef2 Transcription Factors

Intermittent parathyroid hormone (PTH) treatment has an anabolic effect on bone, and daily injections of PTH are currently used clinically for the treatment of osteoporosis (Kraenzlin and Meier 2011). Since a decrease in *SOST* expression increases bone formation, it was critical to determine if the anabolic effect of PTH is mediated by *SOST*. Keller and Kneissel were able to show that PTH suppresses *SOST* expression in vitro, in UMR-106 cells, and that PTH directly regulates the transcript levels of *SOST*, in mouse and rat bones (Keller and Kneissel 2005). *SOST* expression has been examined in several osteoblast-like cell lines, and it was discovered that rat UMR-106 osteoblast-like cells express high levels of *SOST* comparable to those found in vivo, in bone (Keller and Kneissel 2005). Therefore, most in vitro research on *SOST* transcriptional regulation has been carried out in this cell line (Loots et al. 2005; Leupin et al. 2007).

To determine whether PTH regulates *SOST* expression through the distal bone enhancer element ECR5 or through the *SOST* proximal promoter, the PTH effects on luciferase reporter constructs containing ECR5 upstream of the 2-kb human *SOST* proximal promoter or upstream of the *SV40* heterologous promoter were tested in UMR-106 transient transfections. It was found that the ~3–4-fold activity that ECR5 normally exerts on these promoters is completely suppressed by PTH treatment (Fig. 4.3d). Thus, it was concluded that the *SOST* bone enhancer activity is negatively regulated by PTH, independent of the endogenous *SOST* promoter. To identify functional elements within the 255-bp enhancer sequence, a DNase I footprinting experiment was performed to localize DNA regions likely to physically interact with transcription factors. Two regions were identified by this approach, designated as region A and region B (Fig. 4.3e; gray shading). The first region extended approximately from nucleotides 106–146, and the second region extended from nucleotides 169–192 (ECR5 approximate location in hg19: chr17: 41773918–41774104) (Leupin et al. 2007).

Next, deletion analyses were carried out to determine whether either one of these regions were functionally important for transcription activation mediated by ECR5. A truncated ECR5 fragment comprising only the two footprint regions A and B from base pairs 103–193 (p5-AB) was sufficient to recapitulate most of ECR5 enhancer activity in UMR-106 cells, whereas deleting both these regions resulted in a complete loss of enhancer activity (p5-ΔAB). Transcription factor binding site (TFBS) analysis of the two footprint regions predicted two putative TFBS in region A: sites A1 and A2. Deleting putative TFBS A1 boosted the enhancer activity ~2-fold above the full-length enhancement of ECR5, whereas deleting putative TFBS A2 inactivated ECR5 (Fig. 4.3e) (Leupin et al. 2007). These data showed that ECR5 harbors a putative repressor (A1) and two putative activator elements (A2 and B). The A2 sequence matched the consensus binding site for myocyte enhancer factor 2 (MEF2) regulatory proteins; therefore, it was hypothesized that one or more members of the MEF2 family of transcription factors is the likely upstream

regulatory protein controlling ECR5-mediated *SOST* transcription, in the skeleton (Leupin et al. 2007).

MEF2 transcription factors were not intuitive candidate regulatory proteins for controlling transcription in bone, since traditionally *Mef2* genes have been primarily linked to muscle phenotypes (Wang et al. 2001; Potthoff et al. 2007; Lin et al. 1998), and none of the four *Mef2* genes encoded by *Mef2A*, *B*, *C*, and *D* were previously shown to be expressed in bone. Quantitative mRNA expression analysis of rat cortical bone and UMR-106 cells showed strong expression of *Mef2s* comparable to the two known *Mef2* target tissues, heart and brain. In femur and UMR-106 cells, *Mef2C* showed the highest expression level among the four different *Mef2* genes, followed by *Mef2A* and *Mef2D* with about one half as much; no significant expression of *Mef2B* was detected (Leupin et al. 2007; Arnold et al. 2007).

To determine the functional impact of *Mef2* transcription factors on ECR5, Leupin et al. co-transfected human *Mef2C* with various ECR5 constructs and observed a ~300% increase in transcription, independent of the human *SOST* promoter (Leupin et al. 2007). In parallel, they tested a dominant-negative MEF2 construct and observed a 55% reduction in ECR5 transgenic activity. They also examined the potential contribution of individual members of the *Mef2* gene family using RNA interference. Through transfections of specific siRNA directed against all members of the *Mef2* family, they determined that downregulation of *Mef2A*, *C*, and *D* expression leads to a decrease in *SOST* expression by 65%, 59%, and 84%, respectively; *Mef2B* inhibition had no significant effect. To assess functional redundancy among *Mef2* transcription factors, they also examined the synergistic effects of *Mef2A*, *C*, and *D* and noted that downregulation of *Mef2C* and *D* resulted in the most dramatic reduction in *SOST* expression (Leupin et al. 2007).

The findings by Leupin et al. highlight a new role for *Mef2* transcription factors in controlling the transcription of the bone formation inhibitor *SOST* in osteocytes and thereby potentially regulating adult bone mass (Leupin et al. 2007). While these findings will have to be confirmed in vivo, in animal models, it is likely that deleting one or more *Mef2* transcription factors, in the bone, would phenocopy Van Buchem disease. It is also likely that downregulating *SOST* or *Mef2* transcription factors could potentially represent novel therapeutic venues for stimulating bone formation in patients affected by severe bone loss.

4.5 Animal Models of Sclerosteosis and Van Buchem Disease

4.5.1 Targeted Deletion of SOST Causes High Bone Mass

Li et al. used a knockout (KO) targeting vector to replace 80% of the coding region of mouse *Sost*, including exons 1–2, and the intron by a neomycin resistance gene cassette (Li et al. 2008). The generated allele was confirmed to be a null since no *Sost* transcripts were detected by Northern blot analysis of RNA isolated from adult mouse bones. Physically, the knockout mice were indistinguishable from

littermate wild-type (WT) control mice, lacking observable digit abnormalities or facial distortion/paralysis. Li et al. also noted no elevated mortality in the KO mice compared with WT controls observed up to 10 months of age and older. These observations were in contrast to the clinical characteristics of sclerosteosis patients, where a high incidence of syndactyly, facial distortion/palsy, and early death (mid-30s caused by elevated intracranial pressure) has been documented (Li et al. 2008). While no digit defects were originally documented by Li et al., 2/3 digit syndactyly, nail dysplasia, and radial deviation have since been observed in homozygous KO mice, at very low frequency (Loots GG unpublished results).

These homozygous KO mice did however display increased radiodensity, indicative of high bone mineral density (BMD) throughout the skeleton (skull, axial skeleton, ribs, pelvis, long bones) and increased bone volume in vertebrae, long bone, and calvaria as determined qualitatively by histology (Li et al. 2008). Similar to the human characteristics, the skeletal morphology of the KO mice appeared to be normal but with abnormally high bone mass. Gender and heterozygosity were also examined, and it was determined that heterozygous animals were indistinguishable from wild-type littermate controls and that there were no significant differences in BMD between age-matched males and females (examined at 4–6.5 months of age) (Li et al. 2008).

In *Sost* homozygous KO mice, the areal BMD was increased by 62% for lumbar vertebrae and by 55% for whole leg compared with wild-type littermate control mice. High-resolution computed tomography (μCT) analysis of the metaphyseal region of the distal femur showed increased trabecular bone for KO mice compared with wild-type control mice. Both the volumetric BMD (+146% in males, +224% in females) and the bone volume fraction (+149% in males, +281% in females) were significantly increased in KO mice. The cortical thickness was also increased with a reciprocal decrease in the bone marrow cavity area of KO mice compared with sex-matched wild-type control mice. The periosteal perimeter in the KO mice was also significantly increased (+15% in males, +22% in females) as was cortical area (+93% in males, +117% in females). Dynamic histomorphometric analysis of trabecular and cortical bone performed by Li et al. confirmed that the HBM phenotype is mediated by an increase in bone formation rates and hence sclerostin functions primarily to antagonize an important bone formation pathway (Li et al. 2008).

One other important criterion that is critical for the future clinical evaluation of anti-osteoporosis treatments is whether the quality of the newly formed bone is sufficient to sustain loads associated with the normal function of the skeleton. Li et al. subjected *Sost* homozygous KOs to compression testing of lumbar vertebrae and determined that KO mice had significantly higher values for maximum load (+243% in males and +188% in females), stiffness (+93% in males), and energy to failure (+415% in males and +556% in females) compared with sex-matched wild-type controls. They also conducted three-point bend testing of femoral diaphyses and determined that the *SOST* KO mice had significantly higher values for maximum load (+132% in males and +154% in females), stiffness (+87% in males and +119% in females), and energy to failure (+83% in males and +198% in females) compared with sex-matched WT controls (Li et al. 2008).

This impressive increase in bone density and strength combined with the observation that sclerosteosis patients have never been reported to suffer bone fractures, even in accidents in which substantial physical trauma occurred (Hamersma et al. 2003), positions sclerostin as an ideal pharmacological target which should be further exploited for the development of anabolic agents that could potentially treat disorders of bone loss.

4.5.2 Targeted Deletion of ECR5 Causes High Bone Mass

Recent work from our laboratory has shown that a targeted deletion of ECR5 in mice, results in osteoblast-/osteocyte-specific downregulation of *Sost* and subsequent high bone mass phenotype. The bone mass, bone architecture, and histomorphometric analysis of these ECR5 KO mice are consistent with phenotypes that are milder, but highly similar to those documented for *Sost* knockout mice, suggesting that removing the ECR5 regulatory element causes phenotypes similar to Van Buchem disease. Furthermore, the authors presented preliminary in vivo evidence that the *Mef2C* transcription factor is required for ECR5-dependent *Sost* transcription, where osteoblast- and osteocyte-specific ablation of *Mef2C* in conditional knockout mice results in ECR5 dependent loss of *SOST* transcription (Loots GG, unpublished results). These results are highly suggestive that a small, 255-bp, long-range transcriptional regulatory element is an important modulator of *SOST* transcription in bone and its removal is sufficient to inactivate *SOST* transcription and cause Van Buchem disease phenotypes.

4.6 Concluding Remarks

Sclerosing bone dysplasias are rare genetic disorders in which excessive bone formation occurs due to defects in bone remodeling (Van Hul et al. 2001). Identifying the responsible genes, their regulation, and mechanisms of action will provide useful insights into bone physiology and potentially benefit the treatment of these disorders, as well as facilitate the development of therapies for replenishing bone loss in osteoporosis. Genetic studies of Van Buchem disease in animal models has demonstrated that removing a distant *SOST*-specific regulatory element, termed ECR5, causes high bone mass, suggesting that Van Buchem disease is caused by a homozygous hypomorphic allele of the sclerosteosis disease-causing gene, *SOST*.

These findings provide evidence that noncoding regions can be critical to generating disease-causing alleles. In the case of VB disease, the removal of a distant regulatory element affects the transcription levels of *SOST* in a tissue-specific manner and modulates bone mineral density in humans and rodents. An important question remains whether variation in BMD in the general population could also be directly impacted by sequence variants in key noncoding regions of the VB deletion, or

other genetic loci that contribute to bone metabolism. In fact, a recent study investigating the association between common polymorphisms in the *SOST* gene region with BMD in elderly Caucasians identified a polymorphic variant (SRP9) from the VB deletion region that is highly associated with increased BMD, in men (Uitterlinden et al. 2004). Whereas this SRP9 does not map to any human-mouse conserved region in the VB deletion, an important question for future studies is whether this SNP is in linkage disequilibrium with ECR5 or if additional functional SNPs could be identified in this or other *SOST*-specific regulatory elements.

The genetic factors that contribute to susceptibility to bone loss are extremely heterogeneous. Therefore, murine models that affect bone development and growth can provide invaluable insights into the molecular mechanisms of progressive bone loss in humans. Human genetic diseases of the skeleton such as sclerosteosis and Van Buchem disease provide a starting point for understanding the modulation of anabolic bone formation and ultimately have the potential to identify key molecular components that can be used as new therapeutic agents to treat individuals suffering from bone loss disorders. The genomic era has changed the landscape for disease-causing candidate regions, expanding it to include putative transcriptional regulatory elements. Comparative sequence analysis and techniques such as ChIP-seq can now be employed to prioritize candidate regions and to enhance the discovery of noncoding disease-causing mutations in discrete enhancer elements or in large noncoding deletions. While Van Buchem disease may represent a relatively unambiguous case where altering noncoding genomic content deleteriously impacts gene expression and causes a congenital disorder, demonstrating that mutations in distant regulatory elements contribute to inborn errors or susceptibility to disease will continue to remain a great challenge in human genetics.

Abbreviations

BAC	Bacterial artificial chromosome
BMD	Bone mineral density
bp	Base pair
chr	Chromosome
cm	Centimeter
cM	Centimorgan
ECR	Evolutionary conserved region
GFP	Green fluorescent protein
HBM	High bone mass
Hsp68	Heat shock protein 68
Kb	Kilobase
KO	Knockout
lacZ	Beta-galactosidase
lb	Pound
LOD	Logarithm (base 10) of odds

Mb	Megabase
MEF2	Myocyte enhancer factor 2
microCT	Micro-computed tomography
mut	Mutation
NDP	Norrie disease protein
PCR	Polymerase chain reaction
PRDC	Protein related to dan and cerberus
PTH	Parathyroid hormone
qPCR	Quantitative polymerase chain reaction
rtPCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SOST	Sclerostin
SV40	Simian vacuolating virus 40
TFBS	Transcription factor binding site
TGF-β	Transforming growth factor beta
VB	Van Buchem disease
VBΔ	Van Buchem deletion
WT	Wild type

References

Arnold MA et al (2007) MEF2C transcription factor controls chondrocyte hypertrophy and bone development. Dev Cell 12(3):377–389

Balemans W et al (1999) Localization of the gene for sclerosteosis to the van Buchem disease-gene region on chromosome 17q12–q21. Am J Hum Genet 64(6):1661–1669

Balemans W et al (2001) Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). Hum Mol Genet 10(5):537–543

Balemans W et al (2002) Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. J Med Genet 39(2):91–97

Beighton P et al (1984) The syndromic status of sclerosteosis and van Buchem disease. Clin Genet 25(2):175–181

Brunkow ME et al (2001) Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. Am J Hum Genet 68(3):577–589

Hamersma H, Gardner J, Beighton P (2003) The natural history of sclerosteosis. Clin Genet 63(3): 192–197

Janssens K, Van Hul W (2002) Molecular genetics of too much bone. Hum Mol Genet 11(20): 2385–2393

Keller H, Kneissel M (2005) SOST is a target gene for PTH in bone. Bone 37(2):148–158

Kornak U, Mundlos S (2003) Genetic disorders of the skeleton: a developmental approach. Am J Hum Genet 73(3):447–474

Kraenzlin ME, Meier C (2011) Parathyroid hormone analogues in the treatment of osteoporosis. Nat Rev Endocrinol 7(11):647–656

Lee EC et al (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73(1):56–65

Leupin O et al (2007) Control of the SOST bone enhancer by PTH using MEF2 transcription factors. J Bone Miner Res 22(12):1957–1967

Li X et al (2008) Targeted deletion of the sclerostin gene in mice results in increased bone formation and bone strength. J Bone Miner Res 23(6):860–869
Lin Q et al (1998) Requirement of the MADS-box transcription factor MEF2C for vascular development. Development 125(22):4565–4574
Loots GG et al (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288(5463):136–140
Loots GG et al (2005) Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. Genome Res 15(7):928–935
Nobrega MA et al (2003) Scanning human gene deserts for long-range enhancers. Science 302(5644):413
Potthoff MJ et al (2007) Regulation of skeletal muscle sarcomere integrity and postnatal muscle function by Mef2c. Mol Cell Biol 27(23):8143–8151
Staehling-Hampton K et al (2002) A 52-kb deletion in the SOST-MEOX1 intergenic region on 17q12–q21 is associated with van Buchem disease in the Dutch population. Am J Med Genet 110(2):144–152
Uitterlinden AG et al (2004) Polymorphisms in the sclerosteosis/van Buchem disease gene (SOST) region are associated with bone-mineral density in elderly whites. Am J Hum Genet 75(6): 1032–1045
Van Buchem FS, Hadders HN, Ubbens R (1955) An uncommon familial systemic disease of the skeleton: hyperostosis corticalis generalisata familiaris. Acta Radiol 44(2):109–120
Van Hul W et al (1998) Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12–q21. Am J Hum Genet 62(2):391–399
Van Hul W et al (2001) Molecular and radiological diagnosis of sclerosing bone dysplasias. Eur J Radiol 40(3):198–207
Wang DZ et al (2001) The Mef2c gene is a direct transcriptional target of myogenic bHLH and MEF2 proteins during skeletal muscle development. Development 128(22):4623–4633
Winkler DG et al (2003) Osteocyte control of bone formation via sclerostin, a novel BMP antagonist. EMBO J 22(23):6267–6276

Chapter 5
Cis-Regulatory Enhancer Mutations are a Cause of Human Limb Malformations

Julia E. VanderMeer and Nadav Ahituv

Abstract Congenital limb malformations are the second most common class of human birth defects and can be caused both by environmental and genetic factors. While it is known that some limb malformations are the result of coding mutations that disrupt genes, identifying the causal mutation in a patient with an isolated limb malformation is often difficult. This may be due in part to the growing number of cases with isolated limb malformations that are shown to be the result of nucleotide changes in gene regulatory elements. These regulatory mutations affect gene expression in the developing limb and can cause dramatic changes to patterning, leading to congenital limb malformations. In this chapter, we will review characterized gene regulatory mutations leading to human limb malformations and also provide evidence that additional limb enhancers could be the cause of other human limb malformations.

Keywords Limb • *Sonic hedgehog* • ZPA • ZRS • *BMP2* • *SOX9* • Polydactyly • Brachydactyly

5.1 Human Limb Malformations

Human congenital limb malformations occur in as many as 1 in 500 births and, collectively, represent the second most common form of congenital defect (Moore and Persaud 1998). These malformations display a wide range of severities from small

J.E. VanderMeer • N. Ahituv (✉)
Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, 513 Parnassus Ave., Health Sciences East building, Room HSE901H, Box 0794, San Francisco, CA 94143, USA
e-mail: Julia.Vandermeer@ucsf.edu; nadav.ahituv@ucsf.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_5,

changes to digit morphology that do not impair function to more detrimental malformations such as digit fusion or limb truncations that cause severe functional impairment. In some cases, surgery is necessary to restore or improve function. Some types of limb malformations are caused by environmental factors such as exposure to teratogens or by physical constraints like amniotic bands and vascular disruptions in developing embryos, but there can also be genetic causes. The genetic etiology of congenital limb malformations is important to understand for purposes of genetic counseling. Discovering the various genetic mechanisms that cause human limb malformations also provides insight to genes and pathways that are important for tetrapod limb development.

While human limb malformations have been studied since the early nineteenth century (Farabee 1903), the identification of causal mutations is difficult. Though gene mutations have been shown to be the cause of some limb malformations, these are often in the context of a syndrome with phenotypes that affect other tissues and organs (Schwabe and Mundlos 2004). The genetic causes of isolated limb malformations – those that occur in a patient who has no other tissues or organs affected – have proven more difficult to discover. This may be because some of the mutations that cause congenital limb malformations are in regulatory DNA regions that affect the expression of genes important in limb development.

5.2 *Cis*-regulatory Enhancers and Gene Regulation

With the initial analysis of the complete human genome sequence, it quickly became apparent that the human genome contains fewer protein-coding genes than originally expected. This relatively small number of genes reinforced the idea that many genes serve multiple functions, a fact that is especially evident in genes with roles in tissue patterning and embryonic development. The regulation of these genes is very important in order for these roles to be executed at the proper time, place, and expression levels. It appears that some of the regulation for many developmental genes occurs through the actions of *cis*-regulatory elements – sequences of noncoding DNA that control the expression of nearby genes. There are multiple types of *cis*-regulatory elements that can affect gene expression, explored in more detail in Chap. 1 of this book, entitled “Gene Regulatory Elements.”

5.2.1 Identifying and Studying Enhancers

Enhancers are *cis*-regulatory elements that upregulate gene expression. Enhancers are frequently active in only a limited tissue type, even when the gene regulated by the enhancer is expressed in multiple tissues (Visel et al. 2009a). The expression of a gene in multiple tissues is controlled by multiple enhancers, each with its own pattern of activity, where the gene expression pattern is the sum of the enhancers’

activity patterns. Enhancers are not located at fixed positions relative to the genes that they regulate. They can be 3′ or 5′ of the gene and have been found up to 1 megabase away from the gene promoter. Enhancers can be found in intergenic DNA or within introns of the regulated gene or unrelated genes. They are thought to function through the recruitment of transcription factors (TFs) and subsequent physical interactions with the gene promoter.

Because enhancers can be located at great distances relative to the genes that they regulate, it can be difficult to identify them. Traditionally, comparative genomics has been used to identify enhancers. A high degree of conservation can imply that a sequence is functional, and it is thought that many noncoding conserved regions may have *cis*-regulatory roles (Dermitzakis et al. 2005; Pennacchio et al. 2006). Other tactics for identifying enhancers are based on the epigenetic changes that are located at regulatory sequences. Chromatin immunoprecipitation (ChIP) technologies can identify regions of the genome that are enriched for these epigenetic marks or regions that are bound by proteins that are related to enhancer function (Visel et al. 2009b). These approaches and specific proteins and epigenetic enhancer signatures are explored further in Chap. 1.

The gold standard experiment for establishing enhancer function of a particular sequence is based on a simple premise. The potential enhancer sequence is placed in a vector with a minimal promoter and a reporter gene. The minimal promoter is not sufficient to express the reporter gene without the presence of an active enhancer. This construct can be tested in in vivo or in vitro models, but the most commonly used assay system in the context of limb development uses a *LacZ* reporter and is tested in mouse embryos. In this system, *LacZ* is expressed only in the tissues of the embryo where the enhancer is active. Mutations in enhancer sequence can produce patterns of *LacZ* expression that are different from the pattern driven by the normal enhancer, similar to what might happen to the gene normally regulated by the enhancer. The primary drawback to this system is that the expression patterns are qualitative and can only show changes to the expression domains rather than more subtle changes to the degree of expression because of varying numbers of enhancer construct integrations per mouse.

5.2.2 Modular Enhancers and Human Disease

Many developmental genes play key roles in multiple tissues and at different developmental stages. A coding mutation in one of these genes would cause a syndromic phenotype that consists of the sum of the effects to these different tissues and stages. On the other hand, a mutation in a *cis*-regulatory element that controls one aspect of the gene's expression would only cause a phenotype due to the change to the particular tissue where the *cis*-regulatory element is active. This model implies that malformations which occur in both syndromic and isolated forms could represent the results of mutations in coding genes and in the *cis*-regulatory elements that control these genes, respectively. Human limb malformations occur in both syndromic and

isolated forms, making this class of malformation an informative field for studying the effect of *cis*-regulatory mutations on development. In order to study the effect of these mutations, we can use the rich knowledge of tetrapod limb development. Many years of research on limb development have led to detailed characterization of gene expression patterns and phenotypes that result from changes to gene expression.

5.3 Limb Development: Tissue Patterning Along Three Axes

Many limb malformations can arise from changes in patterning that occur in the early stages of limb development. Normal limb development requires the coordinated establishment of the anterior-posterior (AP), proximal-distal (PD), and dorsal-ventral (DV) axes (Fig. 5.1b). Through classical developmental biology studies on developing mouse and chicken limbs, a lot is known about the genes that control these axes and the networks through which they interact.

5.3.1 Early Development and Axis Specification

The limb bud begins as a thickening of mesenchyme cells from the lateral plate mesoderm and somites at the flank of the embryo. This bulge of cells is called the limb bud. Limb bud outgrowth begins with expression of fibroblast growth factor 10 (*FGF10*) in the lateral plate mesoderm which signals through Wnt proteins to induce fibroblast growth factor 8 (*FGF8*) in the ectoderm. *FGF8* in turn stimulates further expression of *FGF10* in the mesoderm, creating a feedback loop that causes proliferation. This signaling also induces the overlying ectoderm cells to form a structure called the apical ectodermal ridge (AER; Fig. 5.1a) at the edge of the limb bud. The AER becomes the primary signaling center that determines the outgrowth of the limb and has a role in establishing the PD axis. Signals from the AER sustain mitotic proliferation in the underlying cells, and if AER signaling is removed, physically or through genetic manipulations, further development of the distal limb ceases (Summerbell 1974).

The AP axis is also controlled by a signaling center. A small region of mesodermal cells at the posterior of the limb bud is required for determining the AP polarity of the limb, and this region is called the zone of polarizing activity (ZPA; Fig. 5.1a). Many experiments have shown that the ZPA defines the AP axis (Saunders and Gasseling 1968) and that it does so by expressing Sonic hedgehog (*SHH*). The SHH protein undergoes an autocatalytic cleavage to generate an active N-terminal fragment that is covalently bound to cholesterol and acts as a morphogen to directly signal to other cells. *SHH* also acts indirectly through bone morphogenetic protein 2 (*BMP2*) and GLI family zinc finger 3 (*GLI3*) in establishing AP gradients and regulating digit identity. The expression of *SHH* defines the posterior portion of the limb, and grafting a secondary ZPA or an alternative source of SHH signal on the anterior side of a chick limb bud causes the development of supernumerary preaxial digits that develop as a mirror image to the normal digits (Riddle et al. 1993).

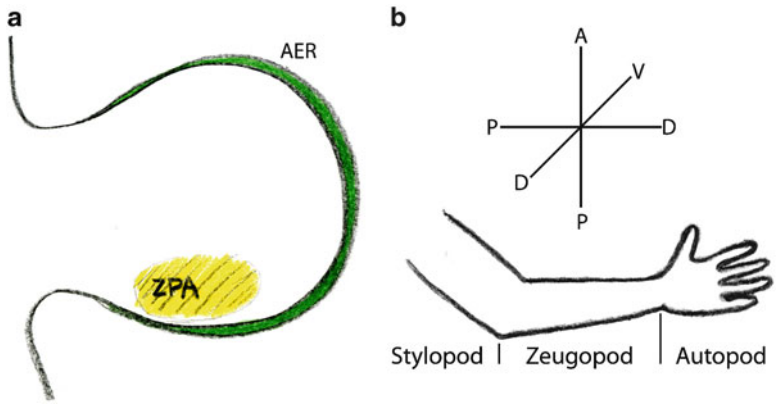

Fig. 5.1 *Limb development and signaling centers.* (**a**) The mouse limb bud at embryonic day 11.5 shows the two major signaling centers, the zone of polarizing activity (*ZPA, in yellow*), and the apical ectodermal ridge (*AER, in green*). (**b**) Signals from these two regions control the three axes of the limb: anterior-posterior (*AP*), proximal-distal (*PD*), and dorsal-ventral (*DV*) and determine proper limb patterning. The axes and their relationship to the limb are shown with labels on the three main limb regions

Cross talk between the ZPA and AER coordinates limb growth. *SHH* from the ZPA induces gremlin 1 (*GREM1*), which in turn induces fibroblast growth factor 4 (*FGF4*) in the posterior AER, a gene that is required for the maintenance of the ZPA. These interactions control the growth and patterning of the limb and illustrate how interference with these pathways could affect development on multiple axes.

5.3.2 Limb Structures and Development

The mature tetrapod limb is divided into three parts from proximal to distal: the stylopod, zeugopod, and autopod (Fig. 5.1b). After the three axes have been established in the limb bud, the next phase of limb development consists of the development of the components of the limb such as the muscles, tendons, skeleton, and nerves. The skeletal elements of the limb form through chondrogenic differentiation where some cells of the limb bud turn into chondrocytes and begin to produce cartilage. This process involves the bone morphogenic protein (BMP) pathway as well as signals from HOX and SOX family transcription factors. Mesenchymal condensations form in the distal limb bud and eventually develop into the bones of the autopod.

Studies from classical embryology show that disruption of genes expressed by the ZPA and AER can cause changes in limb morphology and that disruption of signals at later time points can cause problems with bone development. Given our understanding of the basic roles of these signaling centers in normal limb development, it is clear that disruption of the primary patterning of limb axes can cause limb

malformations. Because many of the genes involved in these pathways are regulated by *cis*-regulatory enhancers, mutations that affect these enhancers could change gene expression patterns and lead to problems in development that result in congenital limb malformations. This has been shown to be the case with mutations that change enhancers that control genes involved in early limb patterning and later morphological changes like bone and cartilage development.

5.4 The ZRS Enhancer in Limb Development

One of the most studied developmental enhancers is the enhancer that controls the expression of *SHH* in the limb bud. Because this enhancer activates *SHH* in the posterior ZPA, it is known as the ZPA regulatory sequence (ZRS). This enhancer was discovered through a combination of mouse models and the study of human patients with preaxial polydactyly.

Developmental studies have shown that inducing ectopic *Shh* expression in the anterior side of the limb bud induces ectopic digits. There are multiple mouse models of preaxial polydactyly where the extra digits appear similar to what was seen in experiments where an ectopic ZPA was grafted to the anterior of a chick limb bud. Because these mouse models were generated or discovered by different screens, they have mutations in different genes, but many were found to have embryonic limb *Shh* expression that extended far beyond the normal posterior ZPA and, in some cases, was even considered a second "anterior ZPA" (Masuya et al. 1995; Chan et al. 1995; Blanc et al. 2002).

Some of these mice were discovered to have defects in genes that function upstream of *Shh* and would normally restrict its expression to the posterior ZPA (*xt* mutant; Hui and Joyner 1998) (*lst* mutant; Qu et al. 1998). Due to mutations in these genes, *Shh* could now be expressed in anterior tissues. In other mice, the phenotype was mapped to the *Shh* locus, but no *Shh* coding mutations were found (Sharpe et al. 1999), and the ectopic *Shh* expression indicates that the gene is functional. The Sasquatch mutant (*ssq*) is an example where the insertion of a transgene caused preaxial polydactyly and the mutation was mapped to the *Shh* locus. This transgenic insertion created a 20 kilobase (kb) duplication within intron 5 of the limb region 1 (*Lmbr1*) gene, which is about 1 megabase away from *Shh* (Lettice et al. 2002). The *ssq* mutant showed not only ectopic *Shh* expression but also expression of a similar pattern for the transgene in both the posterior ZPA and the anterior limb bud, suggesting that the transgene integrated into a region of the genome that is able to regulate the spatial expression of genes – that is to say, a region with a *cis*-regulatory enhancer.

5.4.1 Identifying the ZRS

The homologous human region including *SHH* and *LMBR1* was also recognized to be important in limb patterning through studies of human patients with preaxial

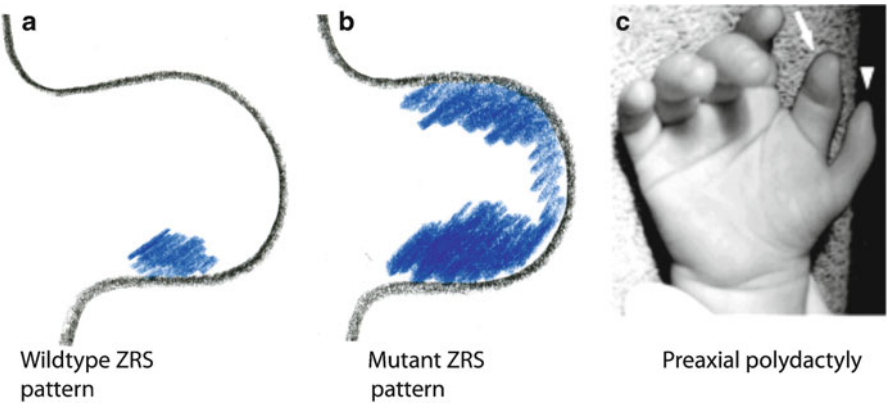

Fig. 5.2 *ZRS expression and preaxial polydactyly.* (**a**) The expression of a LacZ reporter gene, indicated in *blue*, controlled by the normal ZRS enhancer is limited to the posterior mesenchyme that is the normal ZPA. (**b**) Mutations in the ZRS cause the reporter to be expressed more extensively in the posterior and in an ectopic region in the anterior part of the limb bud. Mutations like this are thought to have the same impact on *SHH* expression, and the patterning defect causes preaxial polydactyly (**c**) (Reproduced with permission from Lettice et al. 2003)

polydactyly (PPD). PPD patients have an extra digit on the anterior (thumb) side of the hand or foot, or triphalangeal thumb (TPT), a thumb with a third bone that has the appearance of a second index finger (Fig. 5.2). This phenotype is relatively common and can occur either as an isolated phenotype or as part of a syndrome with other associated phenotypes outside of the limb. Through linkage analysis, isolated PPD was mapped in several families to a region of approximately 500 kb on chromosome seven (Heus et al. 1999). A patient with isolated PPD that had a *de novo* chromosomal translocation in this region led to the fine mapping of the PPD locus to a region within intron 5 of the gene *LMBR1*, the same region disrupted by the mouse *ssq* transgene insertion (Lettice et al. 2002). This intron contains a highly conserved sequence of about 800 base pairs (bp) that was found to have enhancer activity in the posterior limb bud, where it normally controls *SHH* expression, and was called the ZRS. The ZRS has since been studied in many organisms where it was shown to be required for normal *Shh* expression and limb development (Sagai et al. 2005) and to harbor mutations that cause polydactyly in mice (Sagai et al. 2004; Masuya et al. 2007; Lettice et al. 2003), chickens (Maas and Fallon 2004; Maas et al. 2011), dogs (Park et al. 2008), and cats (Lettice et al. 2008). Other human ZRS mutations that have been identified since the 2003 study have been named based on the patients where they were identified as well as the ZRS site where they are located, in accordance with the numbering system assigned by Lettice et al. (2003). Another conserved region next to the ZRS has also been identified as the proximal ZRS (pZRS) and has been found to be the site of mutations that cause polydactyly in dogs, extending the length of the regulatory region where polydactyly mutations can reside to nearly 2 kb (Park et al. 2008).

5.4.2 *Point Mutations Within the ZRS Region*

Over a dozen single nucleotide mutations and one small (13 bp) insertion mutation within the ZRS have been shown to cause human limb malformations. These mutations result in preaxial polydactyly (PPD) and triphalangeal thumb (TPT) with or without supernumerary digits. The phenotype can affect hands and or feet and can be unilateral or bilateral. The known human ZRS mutations are distributed throughout a 700-bp region (see Table 5.1 for a list of human ZRS mutations) within the conserved ZRS. These mutations are thought to change the enhancer activity of the ZRS. Support for this comes from the observation that when some of these mutations were tested in a mouse enhancer assay, the LacZ reporter was shown to be expressed in the anterior of the limb bud (Fig. 5.2a, b). These mutations do not all affect predicted transcription factor binding sites. In addition, the fact that ZRS-related phenotypes differ between the various reported mutations is further evidence that particular mutations in the ZRS can affect the enhancer's function in subtly different ways. Currently, it is not possible to predict phenotypic severity based on sequence alone.

The severity of the human phenotypes does appear to be related to the extent of *SHH* misexpression. While only a few ZRS mutations have been tested in mice for enhancer activity, some do show a correlation between a higher level of reporter gene expression – in the ectopic anterior region and/or the normal posterior region of the limb bud – and more severe human phenotypes. Two of the first reported mutations, referred to as Cuban (ZRS 404G>A) and Belgian1 (ZRS 305A>T), illustrate this correlation. The more severe Cuban patient phenotype includes polydactyly and tibial malformations, and the reporter assay showed a very strong anterior and posterior expression pattern (Zguricas et al. 1999; Lettice et al. 2003, 2008). The Belgian1 mutation that causes PPD2 (OMIM#174500), an extra thumb anterior to a triphalangeal thumb with no abnormalities in the long bones of the arms, shows only weak anterior reporter expression in the reporter assay (Lettice et al. 2003, 2008).

Most of the known ZRS mutations have complete penetrance within the affected family and are inherited in a dominant pattern. There is, however, one reported mutation that does not always cause polydactyly. The mutation at ZRS 295 (295T>C) was first reported to be a neutral polymorphism because it was found in 10–30% of unaffected samples (Lettice et al. 2003). Later, this mutation was discovered to be associated with TPT in multiple English families (Furniss et al. 2008). Examination of this mutation in a mouse enhancer assay showed a weak anterior expression of the reporter, suggesting that this incompletely penetrant mutation might not always cause enough ectopic expression to lead to polydactyly (Furniss et al. 2008). The factors that determine whether the 295T>C mutation causes a phenotype are not known, and no other low-penetrance ZRS mutations have been identified to date.

While most point mutations in the ZRS cause preaxial polydactyly that is limited to the autopod, one particular site in the enhancer is thought to be associated with

Table 5.1 Enhancer defects known to cause limb malformations in human patients

Mutation Name	Mutation	Location (hg19)	Phenotype	Heritability and pheotype varaiation	Reference
***BMP2* limb enhancer (~chr20:6,863,168-6,865,339)**					
Family 1	duplication	~chr20:6,860,129-6,866,024	Brachydactyly type A2	complete penetrance, low variability	Dathe et al. 2009
Family 2	duplication	~chr20:6,860,477-6,866,024	Brachydactyly type A2	incomplete penetrance, variable	Dathe et al. 2009
Chinese family	duplication	~chr20:6,861,382-6,866,044	Brachydactyly type A2	complete penetrance, low variability	Su et al. 2011
***DLX5/6* BS1 enhancer (~chr7:96,357,368-96,357,920)**					
Case	deletion	~chr7:95,552,064-96,432,064	Split hand/foot malformation 1, feet only	sporadic case	Kouwenhoven et al. 2010
***SHH* ZRS enhancer (~chr7:156,583,562-156,584,711)**					
739 A>G, Family A	SNP	chr7:156,583,831	Preaxial polydactyly & triphalangeal thumb	incomplete penetrance, low variability	Gurnett et al. 2007
739 A>G, Family C	SNP	chr7:156,583,831	Preaxial polydactyly & triphalangeal thumb	complete penetrance, variable	Gurnett et al. 2007
621 C>G, Family B	SNP	chr7:156,583,949	Preaxial polydactyly & triphalangeal thumb	complete penetrance, variable	Gurnett et al. 2007
463 T>G	SNP	chr7:156,584,107	Preaxial polydactyly & triphalangeal thumb	incomplete penetrance, low variability	Farooq et al. 2010
404 G>C, Family 2	SNP	chr7:156,584,166	Werner mesomelic syndrome	complete penetrance, low variability	Wieczorek et al. 2009
404 G>A, Family 1	SNP	chr7:156,584,166	Werner mesomelic syndrome	complete penetrance, variable	Wieczorek et al. 2009
404 G>A, Cuban	SNP	chr7:156,584,166	Preaxial polydactyly	complete penetrance, variable	Lettice et al. 2003
396 C>T, Turkish 1	SNP	chr7:156,584,174	Preaxial polydactyly & triphalangeal thumb	complete penetrance, not variable	Semerci et al., 2009
334 T>G, French 2	SNP	chr7:156,584,236	Preaxial polydactyly	incomplete penetrance, not variable	Albuisson et al. 2010
323 T>C, Belgian 2	SNP	chr7:156,584,241	Preaxial polydactyly & triphalangeal thumb	complete penetrance, low variability	Lettice et al. 2003

(continued)

Table 5.1 (continued)

Mutation Name	Mutation	Location (hg19)	Phenotype	Heritability and pheotype varaiation	Reference
305 A>T, Belgian 1	SNP	chr7:156,584,266	Preaxial polydactyly & triphalangeal thumb	complete penetrance, low variability	Lettice et al. 2003
297 G>A, French 1	SNP	chr7:156,584,273	Preaxial polydactyly	complete penetrance, not variable	Albuisson et al. 2010
295 T>C	SNP	chr7:156,584,275	Triphalangeal thumb	low penetrance, high variability	Furniss et al. 2008
105 C>G, Dutch	SNP	chr7:156,584,465	Triphalangeal thumb	complete penetrance, not variable	Lettice et al. 2003
603ins13, Swedish	insertion	~chr7:156,583,968-156,583,969	Preaxial polydactyly & triphalangeal thumb	complete penetrance, not variable	Laurell et al. 2012
Case, Lettice	translocation	t(5,7)(q11,q36)	Preaxial polydactyly & triphalangeal thumb	sporadic case	Lettice et al. 2002
Family, Klopocki	duplication	~chr7:156,143,386-156,732,204	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Klopocki et al. 2008
Family 6, Sun	duplication	~chr7:156,241,020-156,699,998	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Sun et al. 2008
Family 2, Sun	duplication	~chr7:156,241,020-156,677,759	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Sun et al. 2008
Family 5, Sun	duplication	~chr7:156,241,020-156,619,399	Syndactyly type IV	complete penetrance, not variable	Sun et al. 2008
Family 4, Sun	duplication	~chr7:156,354,085-156,687,613	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Sun et al. 2008
Family 3, Sun	duplication	~chr7:156,354,085-156,619,399	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Sun et al. 2008
Family 3, Wieczorek	duplication	~chr7:156,368,541-156,661,877	Triphalangeal thumb-polysyndactyly	sporadic case	Wieczorek et al. 2009
Family 1, Sun	duplication	~chr7:156,539,605-156,699,998	Triphalangeal thumb-polysyndactyly	complete penetrance, variable	Sun et al. 2008
Family, Wu	duplication	~chr7:156,547,469-156,644,074	Syndactyly & tibial hypoplasia	complete penetrance, variable	Wu et al. 2009
Family 4, Wieczorek	duplication	~chr7:156,572,751-156,661,877	Triphalangeal thumb-polysyndactyly	complete penetrance, low variability	Wieczorek et al. 2009
***SOX9* limb enhancer (critical region ~chr17:65,642,665-66,847,686)**					
Family 1	duplication	~chr17:67,900,000-69,860,000	Brachydactyly-anonychia	complete penetrance, not variable	Kurth et al. 2009
Family 2	duplication	~chr17:67,800,000-69,710,000	Brachydactyly-anonychia	suspected germline mosaicism	Kurth et al. 2009
Family 3	duplication	~chr17:68,100,000-69,620,000	Brachydactyly-anonychia	sporadic case	Kurth et al. 2009
Family 4	duplication	~chr17:68,100,000-69,310,000	Brachydactyly-anonychia	unknwown penetrance, low variability	Kurth et al. 2009

Werner mesomelic syndrome (OMIM # 188770), a limb phenotype that includes hypoplastic tibia in addition to triphalangeal thumb polydactyly. Unrelated patients with Werner mesomelic syndrome were found to all have different mutations at ZRS 404, leading to the thought that there might be something special about this site that causes the phenotype to extend beyond the hands and feet (Wieczorek et al. 2009). Because the mechanisms that cause ZRS mutations to change gene expression are unknown, it is currently not possible to determine what makes this site different from the others identified throughout the ZRS.

5.4.3 ZRS Duplications and Complex Polysyndactyly

Human limb malformations have also been attributed to duplications that encompass the ZRS and parts of the surrounding sequence. These duplications cause complex polysyndactyly phenotypes that entail fusion of soft tissue or bones of the autopod in addition to supernumerary digits including triphalangeal thumb polysyndactyly (TPTPS) and syndactyly type IV (Sun et al. 2008). Multiple ZRS duplications have been found in different families. These duplications do not have shared breakpoints, and there is no discernable relationship between the size of the duplication and the severity of the phenotype. The smallest shared region between the various duplications is 47 kb and extends from intron 4 of *LBMR1* and continues into intron 5, ending past the 3′ end of the ZRS. The human ZRS duplication phenotype is different than the mouse *ssq* phenotype which has a 20-kb duplication within intron 5 of *Lmbr1* that includes the ZRS (Sharpe et al. 1999) but shows only polydactyly with no fusion of digits, suggesting either human–mouse phenotypic differences or that the duplicated sequence outside of the ZRS itself may have additional important limb regulatory elements.

5.4.4 ZRS and Acheiropodia

In addition to polydactyly and polysyndactyly, there is another human limb malformation phenotype that has been mapped to the region near the ZRS. Acheiropodia (OMIM #200500) is a severe limb malformation consisting of nearly complete truncations of all limbs and aplasia of the hands and feet. Acheiropodia is a very rare malformation caused by a homozygous deletion of a nearly 6-kb region that removes exon 4 from mRNA transcripts of *LMBR1*, but the deletion does not appear to extend as far as the ZRS in intron 5 (Ianakiev et al. 2001). A mouse model of a ZRS knockout has a similar limb phenotype but lacks only the ZRS and does not have disruptions in *Lmbr1* intron 4 or exon 4 (Sagai et al. 2004). So far, no additional *cis*-regulatory elements have been identified in the acheiropodia deletion.

5.4.5 *Difficulties in Linking ZRS Mutations to Phenotypes*

Animal models of ZRS mutations appear to be of little use in predicting the phenotype caused by specific mutations. Human phenotypes from ZRS point mutations predominantly affect the hands, while mouse models of ZRS mutations tend to have a stronger phenotype in the hind limbs (Knudsen and Kochhar 1982; Sharpe et al. 1999). Furthermore, human patients homozygous for ZRS mutations show phenotypes no more severe than heterozygotes, unlike what has been seen in mice (Semerci et al. 2009). It is clear that not all cases of human isolated preaxial polydactyly are caused by ZRS mutations. There are even numerous families with preaxial polydactyly that is genetically linked to the ZRS that appear to have no mutation or duplications in either the ZRS or in any portion of the acheiropodia deletion (Gurnett et al. 2007; Lettice et al. 2003; Li et al. 2009). Whether other mechanisms are behind these patients' malformations or there is yet another *SHH* limb *cis*-regulatory element in this locus remains to be seen.

5.4.6 *ZRS Looping*

The ZRS has been shown to physically interact with the promoter of *Shh* in mouse limb tissue through DNA looping bringing these two regions into contact (Amano et al. 2009). While the exact looping mechanisms remain unclear, it appears to create specific DNA interactions in a location that correlates with gene expression (Kagey et al. 2010). This interaction occurs only in regions of the limb where *Shh* is "poised" for activation – the posterior ZPA and the anterior mesenchyme region where ectopic *Shh* is observed in polydactylous mouse lines (Amano et al. 2009). In addition to this looping interaction, the same study showed that the looped *Shh*–ZRS complex moves out of its normal chromosome territory in the nucleus when *Shh* is transcribed. This chromosome territory shift normally happens only in the ZPA, suggesting that it is related to the activation of *Shh* expression. Other studies show a role for nuclear matrix proteins in the looping and physical interaction of *Shh* and the ZRS (Zhao et al. 2009). These mechanisms are not entirely clear but show that there are multiple levels of control over *Shh* expression and potentially multiple ways this control could be disrupted by mutations.

5.5 Brachydactyly

While the types of limb malformations that result from ZRS mutations are primarily changes to the number of digits, there are also many types of limb malformation that are due to changes in limb or digit morphology. One class of malformations is the brachydactylies, a related set of conditions where some bones in the autopod are underdeveloped or absent (Fig. 5.3). There are five primary forms of brachydactyly

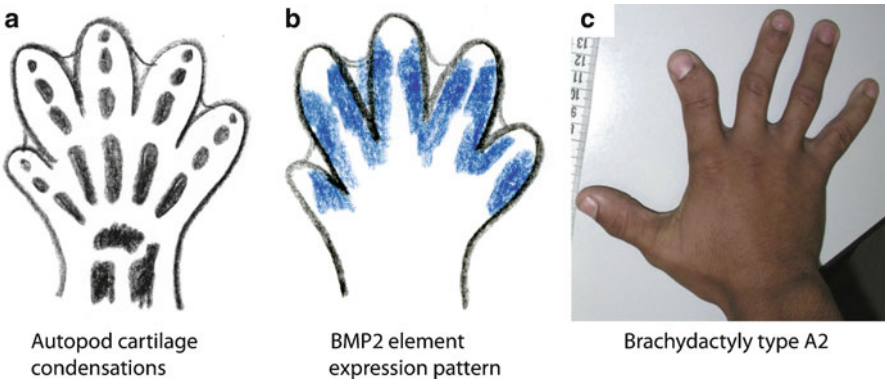

Fig. 5.3 Cartilage condensation and brachydactyly. (**a**) At embryonic day 13.5, mouse limb buds show condensations of cartilage where the bones of the autopod will develop. (**b**) The *BMP2* enhancer shows LacZ reporter expression around the developing phalanges. Changes in gene expression levels that change *BMP2* signaling levels can disrupt development of the digits and cause brachydactyly (**c**) (Reproduced with permission from Dathe et al. 2009)

that are defined by the pattern of affected digits (Mundlos 2009; Stricker and Mundlos 2011). Many are related to mutations in genes associated with growth and differentiation that function through the bone morphogenic protein (BMP) pathway. BMPs are important signaling proteins that are expressed in the condensing mesenchyme of the limb that will form the digits. They were originally thought to be involved only in differentiation of bone but have since then been shown to have multiple important functions in other aspects of growth and patterning.

5.5.1 BMP2 *Limb Enhancer Duplications Cause Brachydactyly Type A2*

Bone morphogenic protein 2 (*BMP2*) is known to play an important role in limb development (reviewed in Robert 2007), and multiple mutations in *BMP2* pathway genes can cause brachydactyly type A2 (BDA2; OMIM# 112600; Fig. 5.3c). A highly conserved region 110 kb 3′ to *BMP2* recapitulates a portion of *BMP2* limb expression (Fig. 5.3b) and is thought to be a *BMP2* limb enhancer (Dathe et al. 2009). The expression pattern of the reporter is in the distal portion of the developing autopod at a time that is critical for digit development. Linkage analysis of a family with autosomal-dominant BDA2 found linkage between the *BMP2* genomic region and this phenotype, but sequencing in two BDA2 families failed to find mutations in the coding region of *BMP2*. Using comparative genomic hybridization (CGH), Dathe et al. found two different, but overlapping, 5.5-kb duplications that include the *BMP2* limb enhancer (Dathe et al. 2009). Another research group later found a third overlapping duplication that caused a similar BDA2 phenotype in an

additional family (Su et al. 2011). These microduplications might increase *BMP2* expression specifically in the limb, disturbing the ratio of signaling factors in the limb. Developing digits and joints are highly sensitive to changes in BMP dosage suggesting that this enhancer-driven increase could cause changes that result in the brachydactyly phenotype.

5.5.2 *The* SOX9 *Enhancer and Brachydactyly*

Another gene whose *cis*-regulatory elements are related to brachydactyly is *SOX9*, a gene that is involved in chondrocyte differentiation. Without expression of *SOX9,* limb skeletal development is severely affected, and limbs are completely absent, though early patterning of the limb bud appears to occur correctly (Akiyama et al. 2002). Mutations in the *SOX9* coding sequence result in a lethal skeletal condition that includes limb malformations, but duplication of a region 5′ of the gene causes only an isolated limb malformation, brachydactyly–anonychia (Kurth et al. 2009). Duplications including this region were found in multiple unrelated families, identifying a "critical region" that likely contains a limb regulatory element. The mechanism here appears to be similar to the *BMP2* enhancer duplication; the increased gene expression due to the duplicated enhancer changes the balance of signaling factors and disrupts development. A transgenic mouse designed to overexpress *SOX9* in the entire limb mesenchyme showed polydactyly as well as short, broad digits (Akiyama et al. 2007), further supporting this proposed mechanism.

5.6 Future Limb Malformation-Associated Enhancers on the Horizon

It is likely that the few enhancers where mutations are confirmed to cause human limb malformations are only the beginning of the discoveries that are still to come. For the ZRS, *BMP2*, and *SOX9* enhancers, the discovery of the mutations in patients with limb malformations came after other indications that these genes and their regulation played roles in development. There are many other genes where there is mounting evidence that expression levels and cues from nearby sequences are important for limb development.

5.6.1 *Chromosomal Rearrangements*

It has long been understood that removing genes from the normal genomic context through chromosomal translocations or inversions can lead to developmental problems. This is now thought to be due in part to the separation of the genes from their

cis-regulatory environment. Chromosomal rearrangements with breakpoints near the homeobox (HOX) gene clusters have been shown to cause limb malformations. In one case, a patient with postaxial polydactyly was found to have a balanced inversion near the HOXA cluster that does not disrupt any of the *HOXA* genes but removes them from a region of several putative *cis*-regulatory elements more than 1 megabase away from the cluster (Lodder et al. 2009). Other chromosomal breakpoints near the HOXD cluster are also associated with limb malformations, also without disrupting genes, but simply removing the normal genomic context (Dlugaszewska et al. 2006). In addition to transcription factors, other genes involved in development have been implicated in this way. A translocation breakpoint near the parathyroid hormone-like hormone (*PTHLH*) gene was shown to downregulate gene expression leading to brachydactyly type E (Maass et al. 2010).

5.6.2 Gene Expression Level Changes

There is also ample evidence that changing the expression level of a certain gene can lead to limb malformations. Clubfoot (also called congenital talipes equinovarus) is a malformation of the legs where the feet are turned inward, disrupting the bones, ankle joints, muscles, and ligaments of the legs. While clubfoot is not painful, it does pose a serious functional problem when a child begins to walk. There are two genes that are thought to be particularly important in specifying the development of tetrapod hind limbs, paired-like homeodomain 1 (*PITX1*), and its downstream target, T-box 4 (*TBX4*). In studies of patients with isolated forms of clubfoot, mutations in the coding regions of *PITX1* that affect gene function and duplications of *TBX4* that would affect expression levels both appear to cause this malformation (Alvarado et al. 2010; Gurnett et al. 2008; Logan and Tabin 1999). Together, these data suggest that expression levels of *TBX4* may be related to this limb malformation. Using a mouse enhancer assay, two hind limb enhancers were discovered in the vicinity of *TBX4* (Menke et al. 2008). In some cases of clubfoot, these hind limb enhancers may have mutations that affect gene expression and cause the limb malformation phenotype. Other limb malformation cases of small duplications or deletions that presumably encompass limb regulatory elements and alter gene expression levels have also been reported (Schluth-Bolard et al. 2008; Tsai et al. 2009; van der Zwaag et al. 2010).

5.6.3 Known Limb Enhancers

There are also known enhancers that are proposed to regulate other important limb developmental genes (Cretekos et al. 2008; Abbasi et al. 2010; Feng et al. 2008; Durand et al. 2009; Sasaki et al. 2002). The study of human limb malformations has also led to the association of particular genomic loci with specific limb phenotypes.

For some of these malformations, no coding mutation can be detected in the associated region. In cases like these, the causal mutation may be in a regulatory element. Split hand–foot malformation (SHFM) is one example for this hypothesis. SHFM is linked to six different genomic loci, and only in two of these loci have coding mutations been found that cause SHFM (SHFM4 is associated with tumor protein p63 (*TP63*) mutations and SHFM6 with wingless-type MMTV integration site family member 10B (*WNT10B*) mutations). An enhancer has been identified within the SHFM1 locus that is thought to control the expression of distal-less homeobox 5 and 6 (*DLX5*/*6*) genes specifically in the limb AER, an expression pattern whose disruption could cause a SHFM-like phenotype (Kouwenhoven et al. 2010). In addition, there are studies in model organisms that have identified *cis*-regulatory element mutations that cause limb malformations in the model organism, though these enhancers have not been studied in detail in human patients (Feng et al. 2008; Liska et al. 2009).

5.7 Summary

Regulatory mutations can affect gene expression and cause dramatic changes in patterning in early development, leading to congenital malformations. Due to their frequency and various phenotypic patterns, limb malformations represent a category of congenital malformations where many cases could be caused by *cis*-regulatory element mutations. As seen with the *SHH* ZRS enhancer, an important limb regulatory element could be a site for many mutations causing related limb malformation phenotypes. In addition to the few known regulatory elements that have been shown to relate to human limb malformations, there is abundant evidence that additional limb-related enhancers exist and that changes to these enhancers could also cause human limb malformation phenotypes. The continued identification of *cis*-regulatory elements that are important in the developing limb will aid in the detection of these sequence changes and increase our understanding of gene regulation and limb development.

Abbreviations

AER	Apical ectodermal ridge
AP	Anterior-posterior [axis]
BDA2	Brachydactyly type A2
BMP	Bone morphogenic protein
BMP2	Bone morphogenic protein 2
bp	Base pairs
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by deep sequencing

DLX5/6	Distal-less homeobox 5 and 6
DV	Dorsal-ventral [axis]
FGF4	Fibroblast growth factor 4
FGF8	Fibroblast growth factor 8
FGF10	Fibroblast growth factor 10
GLI3	GLI family zinc finger 3
GREM1	Gremlin 1
HOX	Homeobox
kb	Kilobase
LMBR1	Limb region 1
PD	Proximal-distal [axis]
PITX1	Paired-like homeodomain 1
PPD	Preaxial polydactyly
PTHLH	Parathyroid hormone-like hormone
SHFM	Split hand-foot malformation
SHH	Sonic hedgehog
SOX9	SRY-box containing gene 9
TBX4	T-box 4
TF	Transcription factor
TPT	Triphalangeal thumb
TPTPS	Triphalangeal thumb polysyndactyly
ZPA	Zone of polarizing activity
ZRS	ZPA regulatory sequence

References

Abbasi AA, Paparidis Z, Malik S, Bangs F, Schmidt A, Koch S, Lopez-Rios J, Grzeschik KH (2010) Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development. BMC Dev Biol 10(44):doi:4410.1186/1471-213x-10-44

Akiyama H, Chaboissier MC, Martin JF, Schedl A, de Crombrugghe B (2002) The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. J Bone Miner Res 17:1071

Akiyama H, Stadler HS, Martin JF, Ishii TM, Beachy PA, Nakamura T, de Crombrugghe B (2007) Misexpression of Sox9 in mouse limb bud mesenchyme induces polydactyly and rescues hypodactyly mice. Matrix Biol 26(4):224–233

Alvarado DM, Aferol H, McCall K, Huang JB, Techy M, Buchan J, Cady J, Gonzales PR, Dobbs MB, Gurnett CA (2010) Familial isolated clubfoot is associated with recurrent chromosome 17q23.1q23.2 microduplications containing TBX4. Am J Hum Genet 87(1):154–160. doi:10.1016/j.ajhg.2010.06.010

Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T (2009) Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell 16(1):47–57. doi:10.1016/j.devcel.2008.11.011

Blanc I, Bach A, Robert B (2002) Unusual pattern of Sonic hedgehog expression in the polydactylous mouse mutant Hemimelic extra-toes. Int J Dev Biol 46(7):969–974

Chan DC, Laufer E, Tabin C, Leder P (1995) Polydactylous limbs in Strongs luxoid mice result from ectopic polarizing activity. Development 121(7):1971–1978

Cretekos CJ, Wang Y, Green ED, Martin JF, Rasweiler JJ, Behringer RR (2008) Regulatory divergence modifies limb length between mammals. Genes Dev 22(2):141–151. doi:10.1101/gad.1620408

Dathe K, Kjaer KW, Brehm A, Meinecke P, Nürnberg P, Neto JC, Brunoni D, Tommerup N, Ott CE, Klopocki E, Seemann P, Mundlos S (2009) Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. Am J Hum Genet 84(4):483–492

Dermitzakis ET, Reymond A, Antonarakis SE (2005) Conserved non-genic sequences – an unexpected feature of mammalian genomes. Nature Rev Genet 6(2):151–157

Dlugaszewska B, Silahtaroglu A, Menzel C, Kubart S, Cohen M, Mundlos S, Tumer Z, Kjaer K, Friedrich U, Ropers HH, Tommerup N, Neitzel H, Kalscheuer VM (2006) Breakpoints around the HOXD cluster result in various limb malformations. J Med Genet 43(2):111–118. doi:10.1136/jmg.2005.033555

Durand C, Bangs F, Signolet J, Decker E, Tickle C, Rappold G (2009) Enhancer elements upstream of the SHOX gene are active in the developing limb. Eur J Hum Genet 18(5):527–532. doi:10.1038/ejhg.2009.216

Farabee W (1903) Hereditary and sexual influence in meristic variation: a study of digital malformations in man. Ph.D. thesis, Harvard University

Feng WG, Huang J, Zhang J, Williams T (2008) Identification and analysis of a conserved Tcfap2a intronic enhancer element required for expression in facial and limb bud mesenchyme. Mol Cell Biol 28(1):315–325. doi:10.1128/mcb.01168-07

Furniss D, Lettice LA, Taylor IB, Critchley PS, Giele H, Hill RE, Wilkie AOM (2008) A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. Hum Mol Genet 17(16):2417–2423. doi:10.1093/hmg/ddn141

Gurnett CA, Alaee F, Kruse LM, Desruisseau DM, Hecht JT, Wise CA, Bowcock AM, Dobbs MB (2008) Asymmetric lower-limb malformations in individuals with homeobox PITX1 gene mutation. Am J Hum Genet 83(5):616–622. doi:10.1016/j.ajhg.2008.10.004

Gurnett CA, Bowcock AM, Dietz FR, Morcuende JA, Murray JC, Dobbs MB (2007) Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. Am J Med Genet 143A(1):27–32. doi:10.1002/ajmg.a.31563

Heus HC, Hing A, van Baren MJ, Joose M, Breedveld GJ, Wang JC, Burgess A, Donnis-Keller H, Berglund C, Zguricas J, Scherer SW, Rommens JM, Oostra BA, Heutink P (1999) A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. Genomics 57(3):342–351

Hui CC, Joyner AL (1998) A mouse model of Greig cephalapolysyndactyly syndrome: the extra-toes' mutation contains an intragenic deletion of the Gli3 gene. Nat Genet 19(4):404–404

Ianakiev P, van Baren MJ, Daly MJ, Toledo SPA, Cavalcanti MG, Neto JC, Silveira EL, Freire-Maia A, Heutink P, Kilpatrick MW, Tsipouras P (2001) Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. Am J Hum Genet 68(1):38–45

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314):430–435. doi:10.1038/nature09380

Knudsen TB, Kochhar DM (1982) The Hemimelic extra toes mouse mutant: historical perspective on unraveling mechanisms of dysmorphogenesis. Birth Defects Res C Embryo Today 90(2):155–162. doi:10.1002/bdrc.20181

Kouwenhoven EN, van Heeringen SJ, Tena JJ, Oti M, Dutilh BE, Alonso ME, de la Calle-Mustienes E, Smeenk L, Rinne T, Parsaulian L, Bolat E, Jurgelenaite R, Huynen MA, Hoischen A, Veltman JA, Brunner HG, Roscioli T, Oates E, Wilson M, Manzanares M, Gomez-Skarmeta JL, Stunnenberg HG, Lohrum M, van Bokhoven H, Zhou HQ (2010) Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. PLoS Genet 6(8):doi:e100106510.1371/journal.pgen.1001065

Kurth I, Klopocki E, Stricker S, van Oosterwijk J, Vanek S, Altmann J, Santos HG, van Harssel JJT, de Ravel T, Wilkie AOM, Gal A, Mundlos S (2009) Duplications of noncoding elements 5′ of SOX9 are associated with brachydactyly-anonychia. Nat Genet 41(8):862–863. doi:10.1038/ng0809-862

Laurel T, VanderMeer JE, Wenger AM, Grigelioniene G, Nordenskjold A, Arner M, Ekblom AG, Bejerano G, Ahituv N, Nordgren A (2012) A novel 13 base pair insertion in the Sonic Hedgehog limb enhancer (LMBR1/ZRS) causes preaxial polydactyly with triphalangeal thumb. Human Mutation. In press

Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12(14):1725–1735. doi:10.1093/hmg/ddg180

Lettice LA, Hill AE, Devenney PS, Hill RE (2008) Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Hum Mol Genet 17(7):978–985. doi:10.1093/hmg/ddm370

Lettice LA, Horikoshi T, Heaney SJH, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci USA 99(11):7548–7553. doi:10.1073/pnas.112212199

Li H, Wang CY, Wang JX, Wu GS, Yu P, Yan XY, Chen YG, Zhao LH, Zhang YP (2009) Mutation analysis of a large Chinese pedigree with congenital preaxial polydactyly. Eur J Hum Genet 17(5):604–610. doi:10.1038/ejhg.2008.240

Liska F, Snajdr P, Sedova L, Seda O, Chylikova B, Slamova P, Krejci E, Sedmera D, Grim M, Krenova D, Kren V (2009) Deletion of a conserved noncoding sequence in Plzf Intron leads to Plzf down-regulation in limb bud and polydactyly in the rat. Dev Dyn 238(3):673–684. doi:10.1002/dvdy.21859

Lodder EM, Eussen BH, van Hassel D, Hoogeboom AJM, Poddighe PJ, Coert JH, Oostra BA, de Klein A, de Graaff E (2009) Implication of long-distance regulation of the HOXA cluster in a patient with postaxial polydactyly. Chromosome Res 17(6):737–744. doi:10.1007/s10577-009-9059-5

Logan M, Tabin CJ (1999) Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. Science 283(5408):1736–1739

Maas SA, Fallon JF (2004) Isolation of the chicken Lmbr1 coding sequence and characterization of its role during chick limb development. Dev Dyn 229(3):520–528. doi:10.1002/dvdy.10502

Maas SA, Suzuki T, Fallon JF (2011) Identification of spontaneous mutations within the long-range limb-specific Sonic hedgehog enhancer (ZRS) that alter Sonic hedgehog expression in the chicken limb mutants oligozeugodactyly and silkie breed. Dev Dyn 240(5):1212–1222. doi:10.1002/dvdy.22634

Maass PG, Wirth J, Aydin A, Rump A, Stricker S, Tinschert S, Otero M, Tsuchimochi K, Goldring MB, Luft FC, Bahring S (2010) A cis-regulatory site downregulates PTHLH in translocation t(8;12)(q13;p11.2) and leads to Brachydactyly type E. Hum Mol Genet 19(5):848–860. doi:10.1093/hmg/ddp553

Masuya H, Sagai T, Wakana S, Moriwaki K, Shiroishi T (1995) A duplicated zone of polarizing activity in polydactylous mouse mutants. Genes Dev 9(13):1645–1653

Masuya H, Sezutsu H, Sakuraba Y, Sagai T, Hosoya M, Kaneda H, Miura I, Kobayashi K, Sumiyama K, Shimizu A, Nagano J, Yokoyama H, Kaneko S, Sakurai N, Okagaki Y, Noda T, Wakana S, Gondo Y, Shiroishi T (2007) A series of ENU-induced single-base substitutions in a long-range cis-element altering Sonic hedgehog expression in the developing mouse limb bud. Genomics 89(2):207–214

Menke DB, Guenther C, Kingsley DM (2008) Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. Development 135(15):2543–2553. doi:10.1242/dev.017384

Moore K, Persaud T (1998) The developing human: clinically oriented embryology, 6th edn. Saunders, Philadelphia

Mundlos S (2009) The brachydactylies: a molecular disease family. Clin Genet 76(2):123–136

Park K, Kang J, Subedi KP, Ha JH, Park C (2008) Canine polydactyl mutations with heterogeneous origin in the conserved intronic sequence of LMBR1. Genetics 179(4):2163–2172. doi:10.1534/genetics.108.087114

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118):499–502

Qu SM, Tucker SC, Ehrlich JS, Levorse JM, Flaherty LA, Wisdom R, Vogt TF (1998) Mutations in mouse Aristaless-like4 cause strong's luxoid polydactyly. Development 125(14):2711–2721

Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic-hedgehog mediates the polarizing activity of the ZPA. Cell 75(7):1401–1416

Robert B (2007) Bone morphogenetic protein signaling in limb outgrowth and patterning. Dev Growth Differ 49(6):455–468. doi:10.1111/j.1440-169X.2007.00946.x

Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T (2005) Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development 132(4):797–803. doi:10.1242/dev.01613

Sagai T, Masuya H, Tamura M, Shimizu K, Yada Y, Wakana S, Gondo Y, Noda T, Shiroishi T (2004) Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). Mamm Genome 15(1):23–34. doi:10.1007/s00335-033-2317-5

Sasaki H, Yamaoka T, Ohuchi H, Yasue A, Nohno T, Kawano H, Kato S, Itakura M, Nagayama M, Noji S (2002) Identification of cis-elements regulating expression of Fgf10 during limb development. Int J Dev Biol 46(7):963–967

Saunders J, Gasseling M (1968) Ectodermal-mesenchymal interactions in the origin of limb symmetry. In: Billingham RFaR (ed) Epithelial-mesenchymal interactions. Williams and Wilkins, Baltimore, pp 78–97

Schluth-Bolard C, Till M, Labalme A, Rey C, Banquart E, Fautrelle A, Martin-Denavit T, Le Lorc'h M, Romana SP, Lazar V, Edery P, Sanlaville D (2008) TWIST microdeletion identified by array CGH in a patient presenting Saethre-Chotzen phenotype and a complex rearrangement involving chromosomes 2 and 7. Eur J Hum Genet 51(2):156–164. doi:10.1016/j.ejmg.2007.12.003

Schwabe G, Mundlos S (2004) Genetics of congenital hand anomalies. Handchir Mikrochir Plast Chir 36:85–97

Semerci CN, Demirkan F, Ozdemir M, Biskin E, Akin B, Bagci H, Akarsu NA (2009) Homozygous feature of isolated triphalangeal thumb-preaxial polydactyly linked to 7q36: no phenotypic difference between homozygotes and heterozygotes. Clin Genet 76(1):85–90. doi:10.1111/j.1399-0004.2009.01192.x

Sharpe J, Lettice L, Hecksher-Sorensen J, Fox M, Hill R, Krumlauf R (1999) Identification of Sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. Curr Biol 9(2):97–101

Stricker S, Mundlos S (2011) Mechanisms of digit formation: human malformation syndromes tell the story. Dev Dyn 240(5):990–1004

Su P, Ding H, Huang D, Zhou Y, Huang W, Zhong L, Vyse TJ, Wang Y (2011) A 4.6 kb genomic duplication on 20p12.2–12.3 is associated with brachydactyly type A2 in a Chinese family. J Med Genet 48(5):312–316. doi:10.1136/jmg.2010.084814

Summerbell D (1974) Quantitative analysis of effect of excision of AER from chick limb-bud. J Embryol Exp Morphol 32:651–660

Sun M, Ma F, Zeng X, Liu Q, Zhao XL, Wu FX, Wu GP, Zhang ZF, Gu B, Zhao YF, Tian SH, Lin B, Kong XY, Zhang XL, Yang W, Lo WHY, Zhang X (2008) Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. J Med Genet 45(9):589–595. doi:10.1136/jmg.2008.057646

Tsai LP, Liao HM, Chen YJ, Fang JS, Chen CH (2009) A novel microdeletion at chromosome 2q31.1–31.2 in a three-generation family presenting duplication of great toes with clinodactyly. Clin Genet 75(5):449–456. doi:10.1111/j.1399-0004.2008.01147.x

van der Zwaag PA, Dijkhuizen T, Gerssen-Schoorl KBJ, Colijn AW, Broens PMA, Flapper BCT, van Ravenswaaij-Arts CMA (2010) An interstitial duplication of chromosome 13q31.3q32.1 further delineates the critical region for postaxial polydactyly type A2. Eur J Hum Genet 53(1):45–49

Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA (2009a) Functional autonomy of distant-acting human enhancers. Genomics 93(6):509–513. doi:10.1016/j.ygeno.2009.02.002

Visel A, Blow MJ, Li ZR, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009b) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–U112. doi:10.1038/nature07730

Wieczorek D, Pawlik B, Li Y, Akarsu NA, Caliebe A, May KJW, Schweiger B, Vargas FR, Balci S, Gillessen-Kaesbach G, Wollnik B (2009) A specific mutation in the distant Sonic Hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. Hum Mutat 31(1):81–89. doi:10.1002/humu.21142

Zguricas J, Heus H, Morales-Peralta E, Breedveld G, Kuyt B, Mumcu EF, Bakker W, Akarsu N, Kay SPJ, Hovius SER, Heredero-Baute L, Oostra BA, Heutink P (1999) Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the gene responsible to a 1.9 cM region on chromosome 7q36. J Med Genet 36(1):33–40. doi:10.1136/jmg.36.1.33

Zhao J, Ding J, Li YQ, Ren KQ, Sha JH, Zhu MS, Gao X (2009) HnRNP U mediates the long-range regulation of Shh expression during limb development. Hum Mol Genet 18(16):3090–3097. doi:10.1093/hmg/ddp250

Chapter 6
Regulatory Mutations Leading to Cleft Lip and Palate

Brian C. Schutte, Walid D. Fakhouri, and Daniel Zemke

Abstract Cleft lip and palate is one of the most common craniofacial birth defects and one of the most common of all birth defects. Its high impact on the affected individual, their families, and society provides strong motivation to understand the causes. Initial genetic studies focused on coding regions of genes that are required for normal development of the lip and palate. However, many individuals with cleft lip and palate do not have mutations in these regions, requiring a broader search for mutations. Recent studies have included conserved noncoding sequences that may harbor regulatory elements. In this chapter, we focus on the discovery and characterization of two noncoding DNA variants in the vicinity of two genes that are associated with cleft lip and palate. First, the minor allele for the SNP rs642961 exemplifies the discovery and validation of a common DNA variant that alters the expression of *IRF6*, a gene that is required for development of both the lip and the palate. Second, a DNA variant in a sequence that is 1.5 Mb away from the *SOX9* gene exemplifies the discovery and validation of a long-range enhancer element. This chapter also contains brief discussions of other examples of DNA variants that affect regulatory elements and contribute to an increased risk for cleft lip and palate. We also discuss approaches and resources available to the craniofacial genetics community to accelerate discovery of additional regulatory elements and DNA variants that affect their activity. We end with a discussion of the tantalizing questions

B.C. Schutte (✉) • W.D. Fakhouri • D. Zemke
Department of Microbiology and Molecular Genetics and Department of Pediatrics and Human Development, Michigan State University, 5162 Biomedical Physical Sciences, East Lansing, MI 48824-4320, USA
e-mail: schutteb@msu.edu; fakhour6@msu.edu; zemkedan@msu.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_6,

that remain to be answered about the regulation of *IRF6* expression and how that may account for missing heritability.

Keywords Cleft lip • Cleft palate • *IRF6* • *MCS9.7* • *TP63* • *FOXE1* • *PDGFC* • *TBX22* • *SOX9* • *SATB2*

6.1 Epidemiology of Cleft Lip and Palate

Worldwide, the incidence of cleft lip and palate is 1–2/1000 live births (Mossey et al. 2009), making it the most common craniofacial birth defect and one of the most common of all birth defects. The incidence of cleft lip and palate varies widely by geographic origin, suggesting regional differences in etiology, which could be due to variation in the environment, but it might also suggest differences in the frequency of DNA variants that contribute risk for cleft lip and palate.

Cases of cleft lip and palate may be divided into two broad categories: syndromic and isolated (non-syndromic). Individuals with a syndromic form of cleft lip and palate not only have an orofacial cleft, but they also have at least one other characteristic abnormality such as a limb or heart defect or developmental delay. Syndromic cases account for about 30% of cleft lip and palate and are highly associated with chromosomal abnormalities, Mendelian disorders, or exposure to teratogens. As we will describe later in this chapter, the chromosomal abnormalities are important for helping to identify long-range regulatory elements, while the Mendelian disorders are important for identifying genes that are key regulators of lip and palate development. We do not discuss the effect of teratogens or other environmental factors, but note that they are extremely important, and a future research challenge is to understand how these exposures alter gene expression or pathway functions that are essential for development of the lip and palate.

Most cases of cleft lip and palate are isolated. That is, the individual with the orofacial cleft lacks any other detectable phenotypic feature. While cases of isolated cleft lip and palate rarely show Mendelian patterns of inheritance, family and twin studies suggest a strong genetic component for its etiology (Lie et al. 1994; Grosen et al. 2011). Although cleft lip and palate is common, studies have shown that cleft lip and palate increases morbidity (Zhu et al. 2002; Bille et al. 2005) and an overall higher risk of mortality that extends into adulthood (Christensen et al. 2004). These observations raise an interesting paradox. How can a genetically caused birth defect, which increases morbidity and mortality, be common? What possible genetic architecture or evolutionary process has allowed this phenomenon to occur? In the context of this chapter, it is important to consider potential genetic models because they inspire hypotheses for the type of mutations that will contribute to cleft lip and palate and where those mutations might be located. In addition, they might suggest effects of DNA variants that extend beyond the development of the lip and palate to contribute risk for or protection from other adult-related health conditions.

At least two models are possible to explain the paradox of a common, genetically caused birth defect. The most intuitive model for a common genetic disease is that it is caused by common variants. This model has been referred to as the "common disease, common variant" hypothesis (Lander 1996). However, how can a variant become common in a human population if it contributes significant risk for a birth defect that increases morbidity and mortality? One possible mechanism is chance, i.e., genetic drift. But, a more interesting model is that the disease-associated DNA variant became common by selection. For example, a DNA variant may affect the function of some other biological process that contributes sufficient positive selection to overcome the negative selection caused by the increased risk for cleft lip and palate. The classic example of positive selection for a common disease-associated DNA variant is sickle cell trait, where the positive selection provided by this allele through resistance to malaria compensates for the negative selection caused by sickle cell anemia. A second model to explain the paradox of the high incidence of cleft lip and palate assumes that there is only negative selection on DNA variants that contribute risk for cleft lip and palate. If so, then the DNA variants may not be common individually, but they may be common collectively. For example, we already have a hint from our discussion above that a very large number of loci (genes) are required for the development of the lip and palate. Thus, with so many loci, it is easy to imagine that the baseline mutation rate in the human genome is sufficient to explain the large number of DNA variants needed to account for a common disorder such as cleft lip and palate. In the next section, we will see that these two models are not mutually exclusive and that both common and rare DNA variants can contribute to the incidence of cleft lip and palate and that these DNA variants have been found in regulatory elements.

6.2 Clinical and Developmental Aspects of Cleft Lip and Palate

Cleft lip and palate are developmental abnormalities that arise in early development. In humans, the lip develops during weeks 6–8 (Carnegie stages 16–23), while the palate develops during weeks 8–12 (Yoon et al. 2000). Thus, normal development of the lip and the palate has both distinct and overlapping time frames. This is important to bear in mind for this chapter because these two developmental processes will likely have distinct mechanisms and regulatory networks, but also, even when common pathways are utilized, they may not be synchronous. Therefore, genes and pathways that are required for development of both the lip and the palate may require distinct or additional regulatory networks.

At week 6, the upper lip begins to take shape as three growth projections begin to merge: the medial nasal prominence, the lateral nasal prominence, and the maxillary process (Yoon et al. 2000). Figure 6.1 shows an example of a developing embryo. Although these are murine embryos, development of the lip and palate in

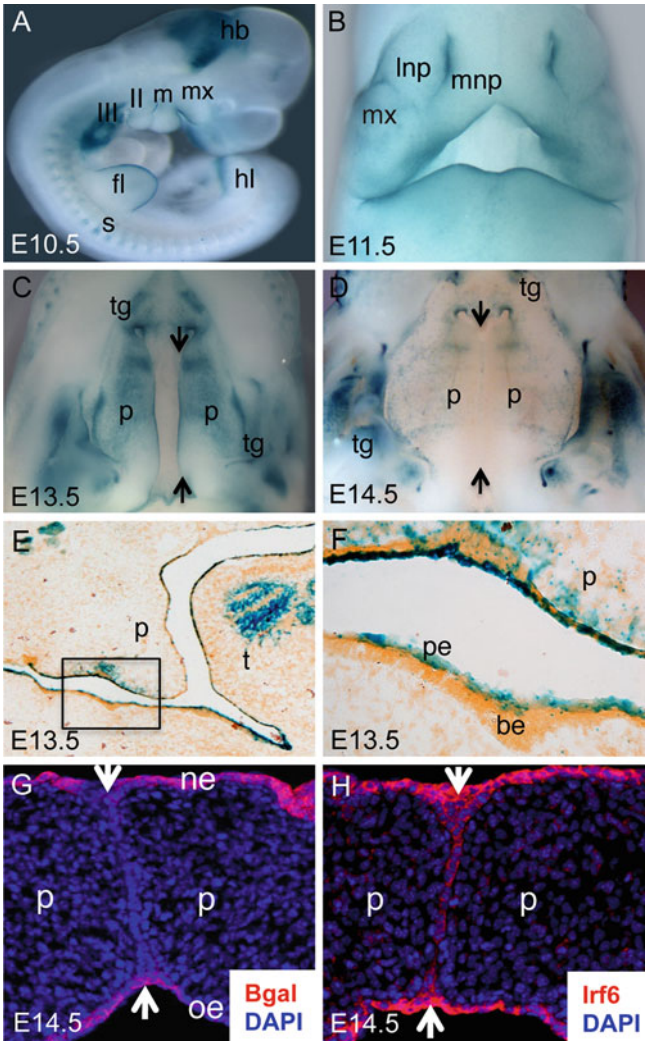

Fig. 6.1 *Mouse MCS9.7 enhancer activity and Irf6 expression during the development of the lip and palate.* (**a**–**f**) *MCS9.7-LacZ* transgenic embryos from various embryonic time points as indicated below each picture. *Blue* staining indicates *MCS9.7* enhancer activity. Staining for Bgal by whole mount (**a**–**d**) or coronal section of head at *low* (**e**) and *high* (**f**) magnification. At *E13.5*, *MCS9.7* is active in periderm (*pe*) and developing muscle in tongue, but not in basal epithelium (*be*) or mesenchyme (mes). Coronal sections of head immunostained for *Bgal* (**g**) or *Irf6* (**h**). At *E14.5*, *MCS9.7* is not active in the medial edge epithelium (MEE), but *Irf6* is highly expressed in these cells. Stained structures include hindbrain (*hb*), maxilla (*mx*), mandible (*m*), second and third pharyngeal arches (*II, III*), somites (s), forelimb (*fl*), hind limb (*hl*), lateral nasal prominence (*lnp*), medial nasal prominence (*mnp*), tooth germ (*tg*), palate (*p*), medial edge epithelium (*arrows*), tongue (*t*), nasal epithelium (*ne*), and oral epithelium (*oe*)

mice is very similar to humans. Therefore, the mouse is an excellent model to study normal development and to determine pathophysiological mechanisms with mutant strains. The relative positions of the three growth projections for the developing lip are shown in Fig. 6.1b. Each of these growth projections is composed of three broadly classified cell types: periderm, basal epithelium, and mesenchyme. The periderm is a highly squamous cell layer that covers the entire external surface of the embryo and the surfaces of the oral cavity (Fig. 6.1f). The basal epithelium is a single, highly ordered layer of cuboidal cells. While the apical surface of the basal epithelium contacts the periderm, the basolateral side is attached to the basement membrane (Fig. 6.1f). Below the basement membrane are the mesenchymal cells. For these three cell types, the critical distinction is that the periderm and basal epithelium are derived from the ectodermal germ cell layer (Byrne et al. 1994) and the mesenchymal cells are derived primarily from cranial neural crest cells. The cranial neural crest cells originate from the lateral ridges of the neural plate and migrate through the pharyngeal arches (Fig. 6.1a) to populate the face. Thus, there is great potential for complex regulation of gene expression. For example, since the epithelial layers and the mesenchymal cells differ in origin, their initial regulatory cues for specification and function will differ. Also, based on their contacts to each other, they will receive different signals from each other. Finally, based on their location in the embryo, these three cell types will differ in their exposure to the environment, whether it is exposures from maternal circulation or from the amniotic fluid.

To complete development of the lip at week 8, the medial and lateral nasal projections fuse with the maxillary process (Yoon et al. 2000). In this context, we define fusion as the formation of a confluent bridge of mesenchymal cells between adjacent tissues. For fusion to occur, the periderm and basal epithelial layers between the fusing tissues must "disappear," and the basement membrane must break down. The mechanism for how the cells between these three growth projections disappear has not been determined. However, in the fusion of the palatal shelves, where more research has been performed, evidence exists for three mechanisms (see below). Thus, the final step in the development of the lip also provides a rich source of regulatory complexity.

Palatal development begins at week 8 with the emergence of the palatal shelves from the maxillary process (Yoon et al. 2000). The analogous event in murine development is shown in Fig. 6.1c, e. Like the growth projections for the lip, the palatal shelves are composed of three similar cell types: periderm, basal epithelium, and mesenchyme (Fig. 6.1f). Over the next few days, the palatal shelves grow downward past the sides of the tongue. By the end of week 8, the tongue drops out of the way, and the vertically oriented palatal shelves elevate into a horizontal position, such that their medial edges appose. Contact between the epithelial cells on the apposing palatal shelves peaks with the formation of the medial edge seam (Fig. 6.1g, h). Once formed, the medial edge seams dissolve, and by week 12, the palatal shelves are fully fused. The confluent bridge of mesenchyme goes on to differentiate into the muscle and cartilage of the mature palate.

While our understanding of development of the lip and palate has been greatly aided by the study of animal models, especially the mouse, it is important to also note

the differences between these two systems. In particular, whereas cleft lip is generally the most common orofacial cleft in humans (Mossey et al. 2009), cleft palate is more common in the mouse (Mouse Genome Informatics; http://www.informatics.jax.org). Also, as we will discuss later, the effect of DNA variants on orthologous genes leads to related but certainly not equivalent effects. The origin of these differences remains elusive and is an important challenge for developmental biologists.

Given this caveat, our current understanding of the cellular functions that are required for palatal development has been aided by the development of an organ culture system from the mouse. Palatal fusion can be divided into three general stages: (1) apoptosis of the periderm, (2) adhesion and intercalation of the basal epithelium to form a single cell layer called the medial edge seam, and (3) dissolution of the medial edge seam to form the confluent bridge of mesenchyme (Nawshad 2008). In order for the basal epithelial cells from apposed palatal shelves to adhere, the superficial layer of periderm cells must disappear, most likely through apoptosis. Based on their location in the oral cavity and at the tip of the palatal shelves, the basal epithelial cells are called the medial edge epithelium (MEE). With the absence of periderm, the cells of the MEE send out filopodial projections (Cox 2004). As intercellular interactions increase, the tight junctions between the MEE cells break down, allowing the opposed cells to intercalate. This process is complete with the formation of a single layer of epithelial cells called the medial edge seam (MES). Once formed, the MES begins to dissolve. Three mechanisms appear to contribute to the dissolution of the MES: (1) terminal differentiation, (2) migration out of the medial edge to the nasal and oral surfaces of the palatal shelves, and (3) epithelium to mesenchymal transition (Nawshad 2008). These three mechanisms are hypothesized to also be involved in fusion of the lip, although this remains to be tested experimentally. Given the similarity in the cell types and the overall process of fusion, it would not be surprising to find a set of genes and pathways that are required for development of both the lip and the palate.

In sum, given the orchestration of many parts and the complexity of each of the parts, it is easy to imagine why cleft lip and palate is the most common craniofacial birth defect and one of the most common birth defects overall. Also, the inherent complexity in temporal and spatial requirement for cellular and molecular functions suggests a complex range of systems to regulate the function of pathways and genes.

6.3 Genetics of Orofacial Clefting Disorders

To overcome the complexity of the etiology of orofacial clefting disorders, multiple strategies have been used to identify the genetic factors involved. These include direct genetic analysis of human populations (syndromic and isolated), gene expression studies, characterization of mouse transgenics and knockouts, and palate culture assays (Schutte and Murray 1999). Based on these criteria, a list of 357 strong candidate genes was compiled (Jugessur et al. 2009). In addition, the Online Mendelian Inheritance in Man (OMIM; www.ncbi.nlm.nih.gov/omim) is an online

Table 6.1 Loci associated with cleft lip and palate by GWAS

Locus	Nearby gene	Associated SNP	Reference
1p22	*ABCA4*	rs560426	(Beaty et al. 2010)
1q32–q41	*IRF6*[a]	rs642961[b]	(Birnbaum et al. 2009; Beaty et al. 2010)
2p21	*THADA*	rs7590268	(Mangold et al. 2009)
8q24	Intergenic	rs987525	(Birnbaum et al. 2009; Grant et al. 2009; Mangold et al. 2009; Beaty et al. 2010)
10q25.3	*VAX1*	rs7078160	(Mangold et al. 2009; Beaty et al. 2010)
13q31.1	*SPRY2*	rs9574565	(Mangold et al. 2009)
15q13.3	*FMN1*	rs1258763	(Mangold et al. 2009)
17q22	*NOG*	rs17760296	(Mangold et al. 2009)
18q22.3	Intergenic	rs17085106	(Grant et al. 2009)
20q11.2	*MAFB*	rs13041247	(Beaty et al. 2010)

[a]*IRF6* is the only gene in this list that was previously associated with cleft lip and palate
[b]The SNP rs642961 is the only SNP in this list that is known to be the actual risk allele. All other SNPs are likely to be associated through linkage disequilibrium with the actual risk allele

resource for human genetic disorders. A search using the key terms "cleft lip or cleft palate" retrieved over 600 entries. Early genetic studies focused on syndromic forms of orofacial clefting because the tools for gene discovery at that time were best suited to find Mendelian disorders and chromosomal abnormalities. OMIM lists about 150 orofacial clefting disorders that have a Mendelian inheritance pattern, and of these, the gene involved has been identified for 54 (Dixon et al. 2011).

From this list, we will pay special attention to two genes: interferon regulatory factor 6 (*IRF6*) and tumor protein p63 (*TP63*/*p63*). These two genes are noteworthy for three reasons. (1) Both *IRF6* and *TP63* encode transcription factors. (2) Mutations in each gene can cause either cleft lip or cleft palate, even in the same family. This "mixed cleft" phenotype in the same family shows that *IRF6* and *TP63* are required for development of both the lip and the palate. Thus, these two distinct developmental processes share at least one common genetic pathway. (3) *IRF6* and *TP63* actually function in the same genetic pathway. In fact, they interact genetically (Thomason et al. 2010), and the mechanism of this interaction will be discussed in a later section.

With the advent of whole genome SNP arrays, family-based linkage studies for orofacial clefting disorders were supplemented by population-based association studies. Association studies can be more sensitive than linkage studies and are able to detect DNA variants that have weaker effects (Risch and Merikangas 1996). To date, four genome-wide association studies (GWAS) have been performed on populations with isolated cleft lip and palate (Birnbaum et al. 2009; Grant et al. 2009; Mangold et al. 2009; Beaty et al. 2010). From these studies, ten loci were identified (Table 6.1). Only one of these loci was previously associated with cleft lip and palate, *IRF6* (Zucchero et al. 2004). Interestingly, all ten associated loci were intergenic. Thus, isolated cleft lip and palate is like other common disorders in that most loci found by GWAS are located between genes or in introns (Hindorff et al. 2009; Gunther et al. 2011). Thus, there is a strong likelihood that the DNA variants that

Table 6.2 DNA elements and variants that regulate expression of genes that are required for development of the lip and palate (nd=not determined)

Gene	Element	Motifs	Reference
BMP4	Promoter	nd	(Suazo et al. 2009)
DLX5/6	Enhancer	MEF2C[a]	(Verzi et al. 2007)
FOXE1	Promoter	MYF-5[b]	Venza et al. 2009
HAND2	Enhancer	ET-1[a]	(Yanagisawa et al. 2003)
IRF6	Enhancer (MCS9.7)	TFAP2A[c], TP63, Ebox, MAFB	(Rahimov et al. 2008; Moretti et al. 2010; Thomason et al. 2010; Fakhouri et al. 2012)
IRF6	Enhancer (MCS2.4, 3.6)	CSL	(Restivo et al. 2011)
IRF6	Promoter	CpG island	(Botti et al. 2011)
PDGFC	Promoter	nd[d]	(Choi et al. 2009)
PITX2	Enhancer	NF1, TCF[e]	(Ai et al. 2007)
SATB2	Enhancer	nd	(FitzPatrick et al. 2003)
SOX9	Enhancer	MSX1[f]	(Benko et al. 2009)
TBX22	Promoter	nd[g]	(Pauws et al. 2009)
TBX22	Enhancer	MN1[a]	(Liu et al. 2008)
TF63	Enhancer	TFAP2, TP63[h]	(Antonini et al. 2006)

[a]Knockout of *trans* factor abolished enhancer activity
[b]SNP rs111846096 is associated with cleft lip and palate
[c]SNP rs642961 is associated with cleft lip and palate
[d]SNP rs28999109 is associated with cleft lip and palate
[e]DNA binding sites were mutated in a murine model
[f]Private mutation found in family with Pierre Robin sequence
[g]SNP rs41307258 is associated with cleft palate
[h]DNA binding sites were mutated in cell culture experiments

account for isolated cleft lip and palate may be found in elements that regulate gene expression. Table 6.2 includes a list of genes involved in cleft lip and palate where a regulatory element has been identified, and in some cases, the regulatory element contains a DNA variant that alters its function. In the next section, we will focus on one example of a DNA variant in a regulatory element that is associated with cleft lip and palate.

6.4 rs642961, a Common DNA Variant in a Regulatory Element for *IRF6*, Is Associated with Cleft Lip and Palate

IRF6 encodes a member of the interferon regulatory factor family of transcription factors. Mutations in the exons of *IRF6* cause two Mendelian orofacial clefting disorders: Van der Woude syndrome (MIM 119300) and popliteal pterygium syndrome (MIM 119500) (Kondo et al. 2002). Van der Woude syndrome is significant to the field of orofacial clefting because it is the most common syndromic form of cleft lip and palate, accounting for 2% of all orofacial clefts, and because it is an outstanding

clinical model for isolated cleft lip and palate. We define a clinical model as a rare disease that has a very similar phenotype for a common disease. For example, Van der Woude syndrome, like isolated cleft lip and palate, can manifest as a mixed cleft phenotype in the same family. Moreover, the only clinical difference is the presence of paramedian pits (or mounds) in the lower lip in 85% of patients with Van der Woude syndrome. Thus, since 15% of patients lack lip pits, they are a perfect phenocopy for isolated cleft lip and palate. Because of this remarkable phenotypic similarity, researchers hypothesized that DNA variation in the Van der Woude syndrome gene would also increase the risk for isolated cleft lip and palate.

To test this hypothesis, Drs. Jeff Murray and Mary Marazita led an international team that discovered that a common DNA variant in *IRF6* was highly associated with isolated cleft lip and palate throughout the world (Zucchero et al. 2004). Although the DNA variant was a non-synonymous SNP (V274I) at a conserved residue, this variant probably was not the allele that accounted for the disease association. In this chapter, we will call such an allele the disease risk allele. The main rationale for this hypothesis was that the associated allele was the ancestral allele, i.e., the allele that is found in other mammals. Since cleft lip and palate is a lethal event in other mammals, there would be a strong purifying selection against this allele. To explain their observations, the authors argued that the V274 allele was in linkage disequilibrium (see below) with the disease risk allele (Zucchero et al. 2004). Since no other common DNA variants were found in the exons of *IRF6*, the researchers hypothesized that the disease risk allele would be in a regulatory element.

To find the disease risk allele at the *IRF6* locus, a multidisciplinary group of investigators combined genomic, human, and murine genetics and molecular analyses. First, they hypothesized that the regulatory element would be highly conserved. By sequencing the *IRF6* locus in 17 species, they identified 41 multispecies conserved sequences (MCS) in the 140-kb haplotype block that contained *IRF6* (Rahimov et al. 2008). The human genome is divided into haplotype blocks, which are regions of variable size along each chromosome where alleles are in linkage disequilibrium. In other words, the alleles within the haplotype block are highly likely to co-segregate during meiosis. Next, the investigators sequenced these 41 conserved regions in cases of isolated cleft lip and palate and controls to find new DNA variants. One new DNA variant, rs642961, was significantly overrepresented in cases over controls and was highly associated with cleft lip in populations from multiple geographic origins (Rahimov et al. 2008).

rs642961 is located in *MCS9.7*, the conserved region located 9.7 kb upstream of the *IRF6* transcriptional start site (Fig. 6.2). This region contains multiple epigenetic signatures that are consistent with enhancer elements, including mono- and trimethylation of the lysine at position 4 of histone H3, a DNaseI-hypersensitive site, acetylated lysine at position 27 of histone H3, and a high level of conservation among mammals. *MCS9.7* also contains a number of binding sites for transcription factors that are important for craniofacial development, including TP63, transcription factor activator enhancer binding protein 2 alpha (TFAP2A, AP-2A), and v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (MAFB). To test whether *MCS9.7* is an enhancer, the authors performed a transient transgenic enhancer assay in mice. In this assay, the putative enhancer is cloned into a vector that contains a

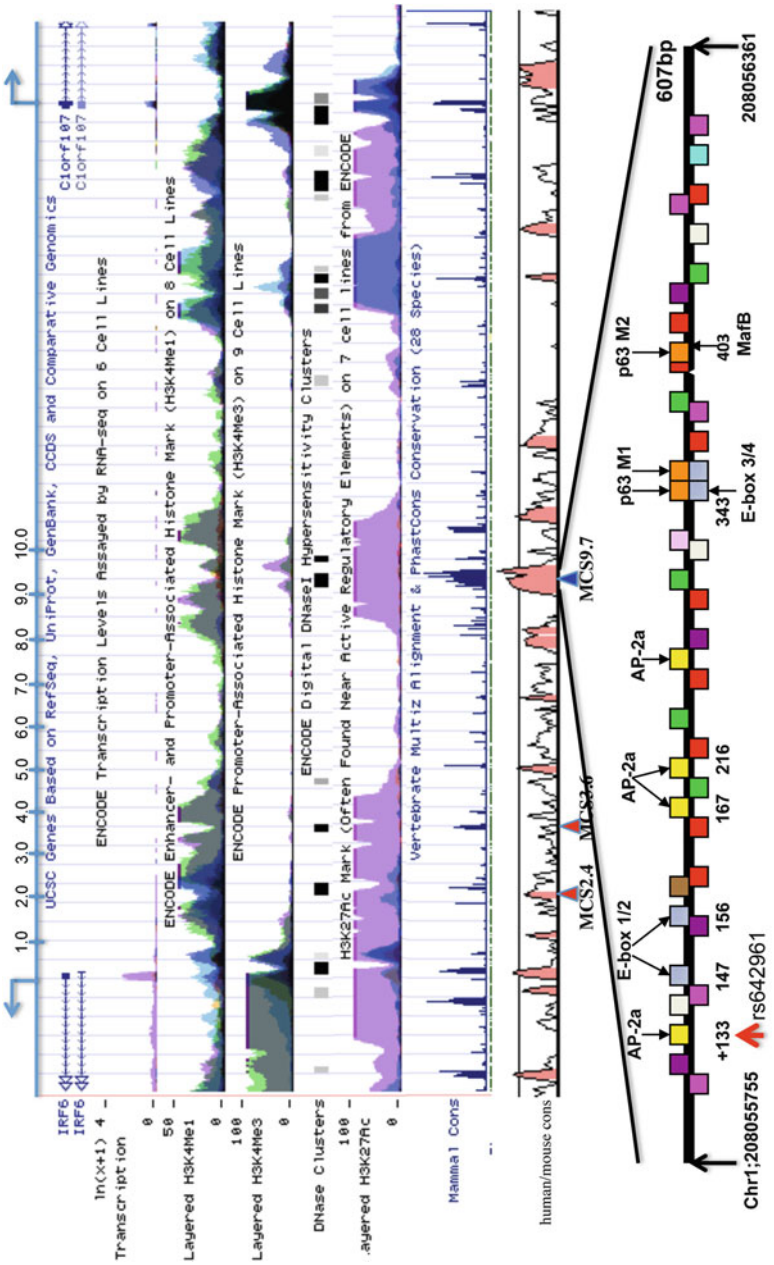
1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
IRF6
IRF6
C1orf107
C1orf107
ln(X+1) 4 -
Transcription
ENCODE Transcription Levels Assayed by RNA-seq on 6 Cell Lines
0 -
50 -
ENCODE Enhancer- and Promoter-Associated Histone Mark (H3K4Me1) on 8 Cell Lines
Layered H3K4Me1
0 -
100 -
ENCODE Promoter-Associated Histone Mark (H3K4Me3) on 9 Cell Lines
Layered H3K4Me3
0 -
ENCODE Digital DNaseI Hypersensitivity Clusters
DNase Clusters
100 -
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0 -
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Mammal Cons
human/mouse cons
MCS2.4
MCS3.6
MCS9.7
AP-2a
E-box 1/2
AP-2a
AP-2a
p63 M1
p63 M2
607bp
+133
147
156
167
216
343
403
E-box 3/4
MafB
Chr1;208055755
rs642961
208056361

Fig. 6.2 *Epigenetic signatures and regulatory landmarks upstream of IRF6.* Tracks from *top* to *bottom* represent the following. Scale bar in kb for a 25-kb region upstream of *IRF6* on chromosome 1q32–q41. *Arrows* show direction of transcription from the transcriptional start sites for *IRF6 and C1orf107*. The next track shows exons 1 and 2 for *IRF6* and *C1orf107*, and the transcriptional signal as obtained from RNA-Seq. A high signal is observed beneath exon 1 of *IRF6*. The next four tracks show the epigenetic markers of monomethylation and trimethylation of histone H3 at lysine 4 (H3K4me1 and H3K4me3, respectively), the DNaseI-hypersensitive sites, and the acetylation of lysine 27 of histone H3 (H3K27ac). These marks are enriched between a 9.0- and 10-kb region upstream of *IRF6*. The next two tracks show a peak of high conservation of multiple mammalian species (http://genome.ucsc.edu), and of mouse (http://genome.lbl.gov/vista), respectively. The highly conserved region was identified as the *MCS9.7* enhancer element. An enlarged view of this element shows many putative transcription factor binding sites. The TFAP2 and TP63 binding sites have been confirmed by ChIP assays. The common regulatory SNP rs642961 (red arrow) is located in the first binding site of TFAP2 within the MCS9.7 element. Two other predicted enhancer elements at multispecies conserved sequences (MCS) that are located 2.4 and 3.6 kb upstream of *IRF6* (red arrow heads) are also shown

basal promoter driving a reporter gene, in this case *LacZ*. Thus, if the cloned sequence is an enhancer, it will drive *LacZ* expression in specific cells or tissues. An example of this kind of staining is shown in Fig. 6.1a. For *MCS9.7*, nine transgenic embryos showed a consistent expression pattern that closely replicated the endogenous expression of *Irf6*. The authors concluded that *MCS9.7* was an enhancer element and was likely to be an important regulatory sequence for *IRF6* (Rahimov et al. 2008). Other genes with expression patterns that overlap with MCS9.7 activity (Fig. 6.3) maybe involved in common pathways and may have common regulatory sequences.

Several additional lines of evidence suggest that rs642961 is the DNA variant that leads to an increased disease risk. First, unlike V274I, the associated allele for rs642961 is the derived allele. That is, it is not the ancestral allele. It is only found as a DNA variant in human populations. Second, the authors observed that the disease-associated allele for rs642961 altered a highly conserved DNA binding site for the AP-2 family of transcription factors. Using an in vitro DNA binding assay, the authors observed that the DNA variant abrogated binding by recombinant TFAP2A protein to this mutated site. This result is significant because previous studies showed that *Tfap2a* in mouse is required for craniofacial development (Schorle et al. 1996) and that mutations in *TFAP2A* cause branchio-oculo-facial syndrome, a disorder that has phenotypic overlap with Van der Woude syndrome, including orofacial clefts and occasional lip pits (Milunsky et al. 2008). In sum, these data are consistent with the hypothesis that the derived allele for rs642961 contributes risk for cleft lip and palate by altering the activity of the *MCS9.7* enhancer element through the abrogation of binding by TFAP2A.

Before leaving this study, we highlight three other points that are relevant to this chapter. First, the authors measured the effect of the risk allele on expression of *IRF6* in a transactivation assay in cell culture. Contrary to expectation, they observed increased expression in cells transfected with the risk allele. This observation points out a potential limitation of using cell lines in testing enhancer elements and DNA variants in those elements. Second, the large effect size for this DNA variant (odds ratio ~ 1.8) and the high carrier frequency for this risk allele in many populations (~22%) combine to give an overall population attributable risk of 12–18%, depending on cleft type and population. The population attributable risk can be depicted as the fraction of disease cases that would not have occurred, if by some mechanism we could remove this risk allele from the world's population. This large value for worldwide attributable risk is certainly consistent with the "common variant, common disease" hypothesis. Thus, rs642961 and isolated cleft lip and palate can be included as one of the rare examples of validating this hypothesis in all of human genetics. Second, the population attributable risk for this allele was much larger for cleft lip than for cleft lip with or without palate, 18% versus 10%. This observation suggests that this variant has a stronger negative effect on development of the lip than the palate. Thus, the authors hypothesize that other variants in *IRF6* could be found that can account for its effect on cleft palate. Since no other DNA variant within *MCS9.7* was associated with orofacial clefts, these results suggest the presence of other DNA variants in other enhancer elements that alter *IRF6* expression and contribute risk for cleft palate. Finally, on average, 22% of the world's population

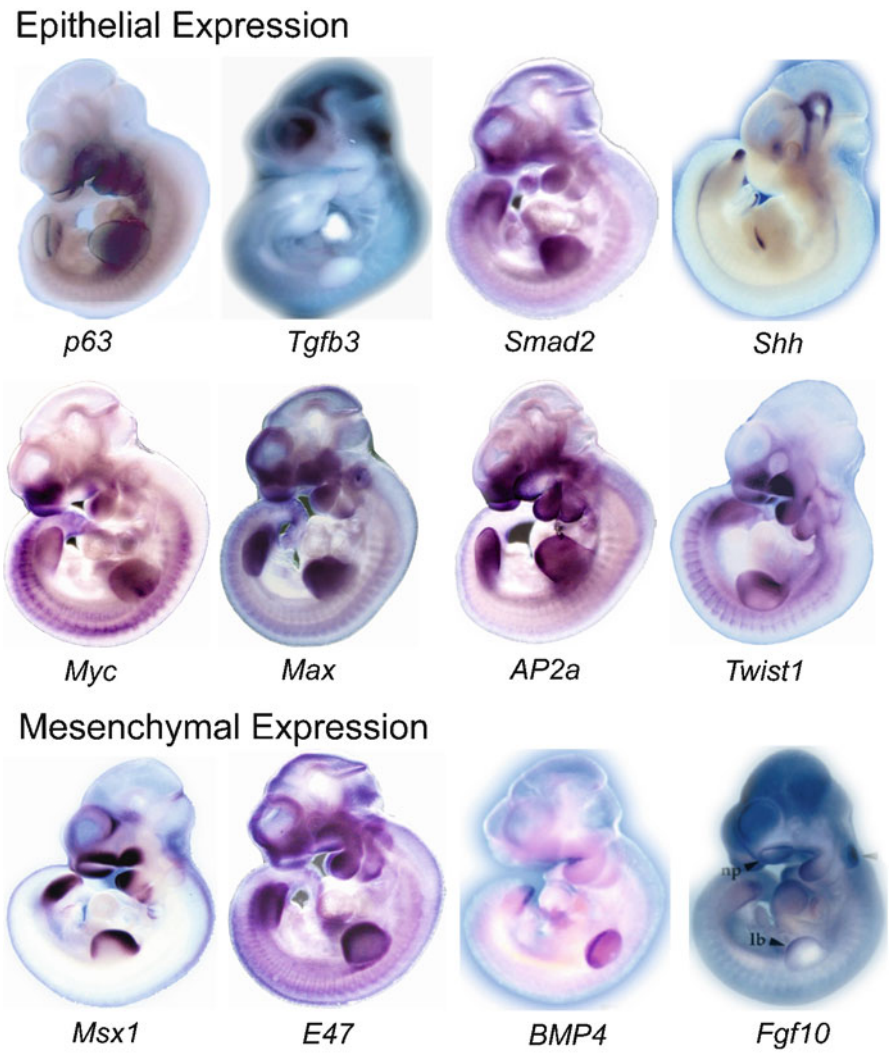

Fig. 6.3 *Whole mount* in situ *hybridization for murine genes that encode transcription and signaling factors at E10.5.* Images of in situ hybridization for the indicated genes were obtained from the EMAGE database (http://www.emouseatlas.org/emage) (Richardson et al. 2010). All of these *trans* factors share a common expression pattern that includes orofacial tissues and limbs. See legend of Fig. 6.1 for names of major embryonic structures. Grouping of epithelial and mesenchymal expression is based on immunostaining of coronal sections of embryonic heads at E13.5 (data not shown). Depending on time point and tissue, expression of *Tfap2a, Bmp4*, *Max*, *Myc*, *Smad2*, and *Twist1* can be observed in either or both epithelium and mesenchyme. *Twist1* is primarily mesenchymal, though placed under epithelium to fit in figure

are carriers for the rs642961 risk allele. Why? As discussed above, how can a risk allele for a disease that has negative evolutionary selection be common worldwide? In Sect. 6.6, we will address these last two points by carefully examining the activity of the *MCS9.7* enhancer.

6.5 *TP63* and *IRF6*, a Gene Regulatory Loop that Is Essential for Palatal Development

As data for *TP63* and *IRF6* have been gathered from human and mouse genetics, the suspicion that these two transcription factors function in a common genetic pathway has increased. For example, mutations in both genes cause multiple human syndromes that affect development of the skin, limbs, and face, including orofacial clefts (MIM 603273 and 607199). Similarly in the mouse, mutant strains for both genes show abnormal development of the skin, limbs, and face, including cleft palate. At the cellular level, both are required for the switch from proliferation to differentiation of keratinocytes during epidermal development (Koster and Roop 2004; Ingraham et al. 2006; Richardson et al. 2006). Finally, in two beautifully complementary studies, *TP63* and *IRF6* were shown to form a gene regulatory loop in both human and mouse keratinocytes (Moretti et al. 2010; Thomason et al. 2010). The study by Moretti and colleagues showed that *TP63* is required for IRF6 expression in three independent systems: (1) in murine keratinocytes that were treated with siRNA for the N-terminally deleted isoform of *TP63* (Δ*Np63*), (2) in epidermis from mice that lack *p63*, and (3) in skin obtained from a patient with ankyloblepharon ectodermal dysplasia clefting (MIM 106260), a syndrome that is caused by mutations in *TP63*. Moreover, they showed that p63 directly binds to two sites at the *IRF6* locus: one in the promoter region and one in intron 1, suggesting that ΔNp63 directly transactivates *IRF6*. This study also showed that the Irf6 protein downregulates the steady-state levels of ΔNp63, thus forming a negative feedback loop. The details of this feedback mechanism are not completely known at this time, but are thought to include the proteasome.

The Thomason study complemented the Moretti paper in several ways. First, the Thomason study also showed that ΔNp63 directly transactivates *IRF6*. However, in the Thomason paper, the binding sites for ΔNp63 were located approximately 10 kb upstream of the *IRF6* transcription start site. The differences in the ΔNp63 binding sites between these two studies are not mutually exclusive, but may reflect the different sources for the keratinocytes and different experimental conditions used. The binding site identified in the Thomason paper was interesting because it co-localized with the *MCS9.7* enhancer element. Computational and molecular studies identified two consensus binding sites for TP63 inside *MCS9.7*, and mutational analyses showed that both sites are required for full enhancer activity of *MCS9.7*. The Thomason study also used mutant strains of mice to show directly that *Tp63* and *Irf6* interact genetically. Whereas the individual heterozygous mice lack gross morphological abnormalities, nearly all embryos that were doubly heterozygous for *Tp63* and *Irf6* had a cleft palate. The full pathophysiological mechanism for this cleft has yet to be elucidated. However, at the histological level, the authors observed that the periderm failed to dissolve completely and that the medial edge epithelium from opposing palatal shelves failed to adhere. On the molecular level, whereas the level of Tp63 goes down in the medial edge epithelium during normal palatal fusion, Tp63 levels remained high in the medial edge epithelium in embryos that lacked *Irf6*.

This observation is consistent with the Moretti study that showed that Irf6 is required to destabilize ΔNp63 protein. Thus, in viewing these two studies in total, an elegant negative feedback loop emerges, whereby ΔNp63 directly transactivates *Irf6* expression, possibly through the *MCS9.7* enhancer element, and Irf6 protein then destabilizes ΔNp63 using a proteasome-dependent mechanism. In the next section, we will revisit the role of TP63 in regulating *IRF6* via the *MCS9.7* enhancer element.

6.6 Enhancer Activity of *MCS9.7* at the *IRF6* Locus

While the transient embryo experiments for *MCS9.7* enhancer activity were consistent between replicates and consistent with endogenous expression of *Irf6*, these data were only generated at one time point and were only viewed in whole mount (Rahimov et al. 2008). To test enhancer activity in other time points and to determine cell-specific expression patterns, it is necessary to generate a stable transgenic strain. Moreover, it is necessary to generate multiple transgenic lines because the standard protocol for generating transgenic mice relies on random insertion of the transgene into the genome. Thus, the pattern of enhancer activity may be altered by the structure of the chromatin at the integration site. This so-called position effect can be controlled by testing the enhancer activity in multiple independent transgenic lines. To characterize the enhancer activity of *MCS9.7*, two stable transgenic lines were created using the same vector that was used for the transient transgenic embryos in the earlier study (Fakhouri et al. 2012). Both stable lines showed an identical pattern of enhancer activity at all time points tested and were completely consistent with the nine transient transgenic embryos of the earlier study. Thus, there was no evidence for position effect in these *MCS9.7* transgenic lines.

The detailed characterization of the stable *MCS9.7-LacZ* transgenic strain exemplifies a number of biological questions that can be addressed with in vivo studies of transgenic enhancer lines. The primary question is whether the enhancer-reporter transgenic strain replicates the expression pattern of the endogenous gene. In the case of *MCS9.7* and *Irf6*, the answer to the primary question is yes, and surprisingly no. Since the answer is also no, secondary questions arise.

As expected from previous studies (Kondo et al. 2002; Ingraham et al. 2006; Knight et al. 2006; Richardson et al. 2006, 2009), enhancer activity of *MCS9.7* was observed along the edges of facial growth projections and branchial arches (Fig. 6.1a, b), the apical ridge of the limb buds (Fig. 6.1a), the palatal rugae (Fig. 6.1c, d), the medial edge of the secondary palatal shelves at mouse embryonic day 13.5 (E13.5) (Fig. 6.1c), the tooth germs (Fig. 6.1c, d), the periderm (Fig. 6.1e, f), the oral and nasal epithelia (Fig. 6.1g), the hair follicles (not shown here), and the epidermis of the skin (not shown here). Thus, *MCS9.7* is sufficient to recapitulate endogenous Irf6 expression in most tissues.

On the other hand, three observations were unexpected. First, the activity of the *MCS9.7* enhancer was more dynamic than anticipated. The original expression studies suggested that Irf6 was expressed in all palatal epithelium at all times.

Rather, the high sensitivity of the transgenic reporter system revealed that Irf6 expression was limited to the periderm prior to E13.5 (Fig. 6.1f), but at later time points, Irf6 was also expressed in the basal epithelium (Fig. 6.1h). This pattern at E13.5 is also unexpected because p63 is strongly expressed in the basal epithelium, but not in the periderm (Thomason et al. 2010; Fakhouri et al. 2012). Thus, despite the strong association between expression of Irf6 and p63 described above, these results show that p63 is neither necessary nor sufficient for Irf6 expression. Second, the *MCS9.7* enhancer was not active in the medial edge epithelium (MEE) at E14.5 (Fig. 6.1d, g). This was very surprising because endogenous Irf6 expression peaks in this tissue at this time point (Fig. 6.1h), strongly suggesting that some other regulatory element is required to drive Irf6 expression to the MEE. We will offer some speculations about the potential for another regulatory element in a later section. The third unexpected observation was that *MCS9.7* was active in regions where expression of *Irf6* had not been recognized, including the hindbrain (Fig. 6.1a), developing muscle in the tongue (Fig. 6.1e) and limb (not shown here). Subsequently, additional immunostaining revealed endogenous Irf6 expression in these regions. These observations have wider implications for orofacial clefting research. For instance, previous studies showed that individuals with Van der Woude syndrome are more likely to have cognitive dysfunction and abnormal brain development (Nopoulos et al. 2007a; Nopoulos et al. 2007b). Finding expression of Irf6 in early brain development may suggest a molecular rationale for these clinical observations.

6.7 DNA Variants in *FOXE1*, *PDGFC*, and *TBX22* Regulatory Elements Lead to an Increased Risk for Orofacial Clefting

Forkhead box E1 (*FOXE1*, also known as *TTF2* and *FKHL15*) is a single exon gene that encodes for a member of the forkhead family of transcription factors. *FOXE1* is required for normal craniofacial development in humans (Clifton-Bligh et al. 1998) and mice (De Felice et al. 1998). Like *IRF6*, DNA variants in *FOXE1* can cause orofacial clefts and contribute risk for orofacial clefts. Specifically, rare DNA variants in *FOXE1* cause Bamforth-Lazarus syndrome (MIM 241850), an autosomal recessive disorder that includes orofacial clefts as well as thyroid agenesis and choanal atresia. Also, relatively common DNA variants have been associated with cleft lip with and without cleft palate and also associated with cleft palate only (Moreno et al. 2009). In this study, DNA sequence analysis of the single exon did not detect any common DNA variants within the gene to account for the association. However, fine mapping identified DNA variants that were highly associated with orofacial clefts that were located 5′ of the gene and also 3′ of the gene. These data suggest the presence of multiple DNA variants that alter expression of *FOXE1* by affecting regulatory elements that might be on both sides of the gene.

In support of this hypothesis, a recent study identified a DNA variant (rs111846096) in the promoter region of *FOXE1* that was associated with cleft lip and palate (Venza et al. 2009). Although the sample size in this study was very small (N = 25), the

DNA variant is very interesting because it is located in a myogenic factor 5 (MYF5) DNA binding site. *MYF5* encodes a member of the myogenic transcription factor family. While this family is well known for its role in muscle development, the authors point out that murine embryos that lack *Myf5* and myogenic differentiation 1 (*MyoD*) have a cleft palate (Rot-Nikcevic et al. 2006). Moreover, the DNA variant abrogates MYF5 binding to this site, and the DNA variant is associated with a sharp decrease in expression of *FOXE1* from patient tissues. While these results need to be replicated in much larger and more diverse populations, the potential impact is high because the frequency of the associated allele is not rare (5%; http://www.ncbi.nlm.nih.gov/snp). Also, genetic studies suggest that DNA variation at *FOXE1* contributes significantly to orofacial clefts (Moreno et al. 2009).

Platelet-derived growth factor C (*PDGFC*) encodes one of the ligands for platelet-derived growth factor receptors. Human linkage and association studies suggest that *PDGFC*, or a nearby gene on chromosome 4q31–q32, is required for development of the lip and palate (Choi et al. 2009). In addition, mice that lack *Pdgfc* have a cleft palate (Ding et al. 2004). While sequence analysis of patient samples did not detect any DNA variants in the coding region of *PDGFC*, a novel DNA variant (rs28999109) was found 986 bp upstream of the transcriptional start site (Choi et al. 2009). The derived allele for rs288999109 was strongly associated with cleft lip with or without cleft palate in a cohort from China and other countries. In addition, this DNA variant significantly reduced the promoter activity for *PDGFC* in a transactivation assay in multiple cell lines. While it was predicted to alter the DNA binding site for six *trans* factors, the effect on any of these *trans* factors has not been tested in vitro or in vivo. Like the DNA variant in the *FOXE1* promoter, rs28999109 has the potential to have a high impact on cleft lip and palate susceptibility because the frequency of the associated allele is not rare (6.7%; http://www.ncbi.nlm.nih.gov/snp).

The final example of a relatively common DNA variant in a regulatory element that is associated with orofacial clefting is rs41307258. This DNA variant is located in the promoter region of T-box 22 (*TBX22*). *TBX22* encodes for a member of the T-box family of transcription factors and is located on the X chromosome. Like *IRF6* and *FOXE1*, DNA variation in *TBX22* can both cause and contribute risk for orofacial clefts. In this case, the cleft phenotype is cleft palate only (MIM 303400). Loss of function mutations in *TBX22* causes X-linked cleft palate and ankyloglossia in familial cases (Braybrook et al. 2001) and accounts for 4–8% of isolated cases of cleft palate (Marcano et al. 2004). Thus, it is important to appreciate that DNA variation in *TBX22* contributes to a broad spectrum of phenotypes and that an obvious genotype-phenotype relationship has not been detected. In a more recent study, the hypothesis that DNA variation in the promoter region of *TBX22* caused or contributed risk for cleft palate with or without ankyloglossia was tested (Pauws et al. 2009). While no novel DNA variants were identified, seven previously identified SNPs were analyzed. Two of these SNPs were highly associated with cleft palate, and when stratified for the presence of ankyloglossia, the association increased. Finally, a promoter activity assay in a single cell line was performed, and a significant decrease in promoter activity with the derived allele for rs41307258 was observed.

As the authors point out, the failure to detect an effect on promoter activity with the other DNA variant may simply reflect the difference between conditions in vivo and in a specific cell line in cell culture (Cirulli and Goldstein 2007).

6.8 Discovery of Long-Range Enhancers for *SOX9* and *SATB2* by Chromosomal Abnormalities in Patients with Orofacial Clefts

So far in this chapter, we have discussed DNA variants in regulatory elements that were located near the gene of interest. Certainly, a more daunting task is to identify regulatory elements, such as long-range enhancers, that are far away. The field of craniofacial genetics offers two good examples where long-range enhancers are involved in human disease. These examples share two common themes – the involved genes are located in gene deserts and the use of chromosomal abnormalities to help localize the regulatory element.

SRY (sex-determining region Y)-box 9 (*SOX9*) encodes a member of the sex-determining region Y (*SRY*)-related HMG box (*SOX*) family of transcription factors and is located on chromosome 17q24.3. Haploinsufficiency of *SOX9* causes campomelic dysplasia (Foster et al. 1994), an autosomal dominant disorder that includes abnormal skeletal and genital development and can include cleft palate (MIM 114290). Pierre Robin sequence is also an orofacial clefting disorder that was mapped to 17q24–q25 (Benko et al. 2009) and is described in detail in Chap. 7 of this book. A few families with Pierre Robin sequence were identified to have chromosomal abnormalities far away from *SOX9* that did not include the gene. These helped identify regulatory elements that regulate *SOX9* from a distance and when mutated could cause Pierre Robin sequence (described in detail in Chap. 7 of this book).

SATB homeobox 2 (*SATB2*) encodes a DNA-binding protein that regulates gene expression through chromatin modification and interaction (Dobreva et al. 2003; Britanova et al. 2005) and is required for embryogenesis, including development of the palate (Britanova et al. 2006; Dobreva et al. 2006). In humans, *SATB2* is located on chromosome 2q32–q33 and has been implicated in palatal development – one case of a de novo nonsense mutation in *SATB2* in an individual with multiple congenital anomalies, including cleft palate (Leoyklang et al. 2007), three cases of microdeletions that included part of the *SATB2* gene where one of these had cleft palate (Rosenfeld et al. 2009), and two cases of a balanced chromosomal translocation that were located within *SATB2* (FitzPatrick et al. 2003; Tegay et al. 2009). Significant for this chapter, a third balanced translocation was also found, but the location of the breakpoint was 3′ of *SATB2* (FitzPatrick et al. 2003) suggesting the presence of a distant enhancer that lies distal to *SATB2*. As exemplified by *SOX9*, there is a focused effort to screen patients with cleft palate for chromosomal abnormalities near the *SATB2* locus that can be used to refine the mapping of the predicted regulatory element.

6.9 Resources for Discovery of Risk Alleles in Regulatory Elements

The field of craniofacial genetics has followed a steady progression of gene discoveries that matches other human disorders. First came the lowest hanging fruit, the discovery of disease-causing alleles in genes involved in the rare Mendelian disorders using linkage analysis in families. Then came the alleles that contribute risk for the common but genetically complex disorders using GWAS in large population cohorts. Despite these advances, there are two obvious gaps in our knowledge of the genetics of cleft lip and palate (and other diseases) that are relevant to this chapter. First, in the Mendelian disorders, no disease-causing mutation has been found for a significant proportion of families. For example, in Van der Woude syndrome, no etiologic mutation has been found in about 25% of families (de Lima et al. 2009). Where are these missing mutations? Potential sources are mutations in other genes or mutations at the *IRF6* locus that are outside of the exons, such as in gene regulatory elements. For common diseases such as non-syndromic cleft lip and palate, ten loci were found by GWAS. However, the allele that actually contributes the risk is known for only one of these loci, *IRF6*. And even for *IRF6*, not all of the risks at this locus can be attributed to rs642961, the SNP whose derived allele alters the function of the enhancer element *MCS9.7*. Where are the other risk alleles at the *IRF6* locus and the other nine loci? One potential explanation is that these risk alleles are likely to be in regulatory elements. The rationale for this hypothesis is that most of the loci from the GWAS are located in regions between genes. Thus, the risk alleles are likely to be in regulatory elements or in genes that are difficult to detect such as noncoding RNAs.

A major challenge then is to find the regulatory elements and the DNA variants within them. There are two general approaches to achieve these two goals: (1) find the regulatory element and then sequence for the presence of DNA variants in case and control populations or (2) fine map the DNA variants that are associated with cleft lip and palate and then test the sequences surrounding the DNA variant for enhancer activity in vitro and in vivo. The two best cleft lip and palate-associated examples to find regulatory variants that contribute risk for cleft lip and palate, *IRF6* and *SOX9* (described in Chap. 7), both relied on the second approach. The reason that both studies used DNA sequence analysis first is indicative of the available experimental resources. Currently, it is much easier to perform high-throughput DNA sequence analysis than it is to perform high-throughput enhancer activity assays. Given the rapid pace of advances in DNA-sequencing technology, this pattern is likely to continue and strongly support the "sequence-first" paradigm. However, it is important to recognize that these two approaches are not mutually exclusive, and rapidly expanding resources exist for identifying sequences that are likely to contain regulatory elements. These include genome-wide analysis of chromatin structures using chromatin immunoprecipitation (ChIP), and genome-wide screening for enhancers using a transient transgenic embryo assay. Both sets of experiments are being performed by the Encyclopedia of DNA Elements (ENCODE), and their

Table 6.3 Transcription factors involved in craniofacial development for which genome-wide ChIP was performed

TF	Locus	ChIP[a]	Reference
ARNT	1q21.3	C	(Noordeen et al. 2009)
BARX2	11q25	C	(Stevens et al. 2004)
CDX4	Xq13.2	C	(Sturgeon et al. 2010)
HAND2	4q34.1	C	(Holler et al. 2010)
DLX1	2q31.1	C	(Zhou et al. 2004)
DLX2	2q31.1	C	(Zhou et al. 2004)
EGR3	8p23–p21	C	(Weigelt et al. 2011)
EVI1	3q24–q28	S/C	(Wang et al. 2011)
FOXH1	8q24.3	S/C	(Kim et al. 2011)
FOXP2	7q31	C	(Vernes et al. 2011)
GLI3	7p14.1	S/C	(Rodelsperger et al. 2010)
HIC1	17p13.3	C	(Van Rechem et al. 2009)
HIF1A	14q23.2	C	(Zhu et al. 2011)
IRF6	1q32.2	C	(Botti et al. 2011)
IRF9	14q11.2	C	(Kubosaki et al. 2010)
LEF1	4q25	C	(Yun and Im 2007)
NR3C1	5q31	S/C	(Pan et al. 2011)
PAX3	2q35-q37	C	(Lagha et al. 2008)
PITX2	4q25	S/C	(Gu et al. 2010)
RUNX2	6p21	C	(van der Deen et al. 2011)
SALL4	20q13.13–q13.2	C	(Yang et al. 2008)
SMAD1	4q31	S/C	(Morikawa et al. 2011)
SMAD2	18q21.1	S/C	(Liu et al. 2011)
SMAD3	15q22.3	S/C	(Liu et al. 2011)
SMAD4	18q21.1	S/C	(Kennedy et al. 2011)
SOX1	13q34	S/C	(Fang et al. 2011)
SOX9	17q24.3–q25.1	S/C	(Nishiyama et al. 2009)
STAT3	17q21.31	S/C	(Durant et al. 2010)
TBX21	17q21.3	S/C	(Lu et al. 2011)
TFAP2A	6p24	S/C	(Ramos et al. 2010)
TP63	3q27	S/C	(Kouwenhoven et al. 2010)
ZIC3	Xq26.2	C	(Lim et al. 2010)

[a]S = ChIP-Seq; C = ChIP-Chip

data is available at the following website: http://genome.ucsc.edu/ENCODE. Another rapidly expanding set of data is ChIP experiments performed with transcription factors that are known to be involved in cleft lip and palate (Table 6.3). These data sets contain DNA sequences that are likely to be regulatory elements. A list of transcription factors (TF) involved in cleft lip and palate was drawn from a previously published list of 357 candidate genes (Jugessur et al. 2009). Of the 89 transcription factors in that list, ChIP followed by sequencing (ChIP-Seq) or chip (ChIP-Chip) analyses were performed on 32 TFs from this list.

These two discovery approaches assume that a locus for cleft lip and palate has already been identified. However, even with multiple GWAS from diverse

Table 6.4 Criteria and resources for discovery of genes involved in cleft lip and palate

Criteria	URL
Known human gene or locus	www.ncbi.nlm.nih.gov/omim www.genome.gov/gwastudies/
Mutant murine strain with cleft	www.informatics.jax.org
Expression in human craniofacial tissues	http://humgen.wustl.edu/COGENE www.facebase.org
Expression in murine craniofacial tissues (e.g., see Fig. 6.3)	http://www.emouseatlas.org/emap/home.html www.facebase.org
Pathway analysis, i.e., gene in same pathway as known genes	http://david.abcc.ncifcrf.gov/home.jsp www.genome.jp/kegg/
SNP database	www.ncbi.nlm.nih.gov/projects/SNP/

populations, not all loci have been identified. Thus, resources are needed to identify strong candidates for genes involved in cleft lip and palate. Table 6.4 contains a list of criteria for candidate genes and the available resources to address each criterion.

6.10 Summary

In this chapter, we emphasized the discovery and functional characterization of rs642961, a DNA variant near *IRF6* that contributes significant risk for cleft lip and palate. This DNA variant is noteworthy because of its high impact in orofacial clefting, because its effect was well characterized in vitro, and because the activity of the enhancer in which it is located was extremely well characterized in vivo. However, there are at least three missing pieces to complete this puzzle. First, what is the effect of this DNA variant during palatal development? This is a challenging question in humans because the target tissues are early in embryonic development. Thus, there are both ethical and technical challenges to overcome. The technical challenges include collecting enough human fetal samples that have the appropriate genotype and then to perform quantitative gene expression measurements from highly specific cell types, e.g., periderm and basal epithelium from the oral cavity during development of the lip and palate. Mutant murine models would provide a reasonable alternative. For example, a transgenic strain could be created that contains the *MCS9.7-LacZ* transgene in which the *MCS9.7* contains the risk allele for rs642961. A more elaborate experiment would be to actually create a knockin strain that contains the risk allele in its native locus. Careful analysis of these murine strains would then allow more directed hypotheses to be tested in human fetuses, thereby reducing both ethical and technical challenges.

The second missing piece of the *IRF6* puzzle is the effect of the risk allele for rs642961 in non-craniofacial tissues. Recall the earlier discussion of the paradox whereby the risk allele for rs642961 is common, and yet its effect should have strong negative evolutionary pressure. Can this paradox be resolved by compensating

positive evolutionary pressure through an effect in another tissue? Again, the mutant murine models, especially the knockin strain, will provide for important resources to address this question.

The final missing pieces of the *IRF6* puzzle are the missing mutations. Specifically, where are the disease-causing mutations in the 25% of families with Van der Woude syndrome that lack mutations in the coding region? And, where are the additional DNA variants that contribute risk for non-syndromic cleft lip and palate? For the Van der Woude families, the enhancer element *MCS9.7* is an excellent candidate region to find disease-causing mutations. Also, for both the Van der Woude families and for the non-syndromic cleft lip and palate cases, mutations could be identified in other enhancer elements that drive *IRF6* expression. We hypothesize that at least one more exists. The rationale for this hypothesis is that the *MCS9.7* enhancer is not active in the medial edge epithelium when the palatal shelves are fusing, even though endogenous *IRF6* expression is high. Also, previous studies suggest that the enhancer activity in the medial edge epithelium is driven by transforming growth factor beta 3 (Tgfb3) signaling. Thus, the approaches and resources described in Sect. 6.9 are being applied to find this very important missing piece.

Abbreviations

Δ*Np63*	N-terminally deleted isoform of *TP63*
ChIP	Chromatin immunoprecipitation
ChIP-Seq	ChIP followed by sequencing
ChIP-Chip	ChIP followed by microarray analysis
ENCODE	Encyclopedia of DNA Elements
FOXE1	Forkhead box E1
GWAS	Genome-wide association studies
IRF6	Interferon regulatory factor 6
Kb	Kilobase
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B
MCS	Multispecies conserved sequences
MEE	Medial edge epithelium
MES	Medial edge seam
MyoD	Myogenic differentiation 1
OMIM	Online Mendelian inheritance of man
PDGFC	Platelet-derived growth factor C
SATB2	SATB homeobox 2
SNP	Single nucleotide polymorphism
SOX9	SRY (sex-determining region Y)-box 9
TBX22	T-box 22
TFAP2A	Transcription factor activator enhancer binding protein 2 alpha
TP63	Tumor protein p63

References

Ai D, Wang J, Amen M, Lu MF, Amendt BA, Martin JF (2007) Nuclear factor 1 and T-cell factor/LEF recognition elements regulate Pitx2 transcription in pituitary development. Mol Cell Biol 27:5765–5775

Antonini D, Rossi B, Han R, Minichiello A, Di Palma T, Corrado M, Banfi S, Zannini M, Brissette JL, Missero C (2006) An autoregulatory loop directs the tissue-specific expression of p63 through a long-range evolutionarily conserved enhancer. Mol Cell Biol 26:3308–3318

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF (2010) A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 42(6):525–529

Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT, McBride D, Golzio C, Fisher M, Perry P, Abadie V, Ayuso C, Holder-Espinasse M, Kilpatrick N, Lees MM, Picard A, Temple IK, Thomas P, Vazquez MP, Vekemans M, Roest Crollius H, Hastie ND, Munnich A, Etchevers HC, Pelet A, Farlie PG, Fitzpatrick DR, Lyonnet S (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet 41:359–364

Bille C, Winther JF, Bautz A, Murray JC, Olsen J, Christensen K (2005) Cancer risk in persons with oral cleft–a population-based study of 8,093 cases. Am J Epidemiol 161:1047–1055

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Berge SJ, Reich RH, Schiefke F, Hemprich A, Potzsch S, Steegers-Theunissen RP, Potzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nothen MM, Mangold E (2009) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41:473–477

Botti E, Spallone G, Moretti F, Marinari B, Pinetti V, Galanti S, De Meo PD, De Nicola F, Ganci F, Castrignano T, Pesole G, Chimenti S, Guerrini L, Fanciulli M, Blandino G, Karin M, Costanzo A (2011) Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. Proc Natl Acad Sci USA 108:13710–13715

Braybrook C, Doudney K, Marcano AC, Arnason A, Bjornsson A, Patton MA, Goodfellow PJ, Moore GE, Stanier P (2001) The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. Nat Genet 29:179–183

Britanova O, Akopov S, Lukyanov S, Gruss P, Tarabykin V (2005) Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. Eur J Neurosci 21:658–668

Britanova O, Depew MJ, Schwark M, Thomas BL, Miletich I, Sharpe P, Tarabykin V (2006) Satb2 haploinsufficiency phenocopies 2q32–q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. Am J Hum Genet 79:668–678

Byrne C, Tainsky M, Fuchs E (1994) Programming gene expression in developing epidermis. Development 120:2369–2383

Choi SJ, Marazita ML, Hart PS, Sulima PP, Field LL, McHenry TG, Govil M, Cooper ME, Letra A, Menezes R, Narayanan S, Mansilla MA, Granjeiro JM, Vieira AR, Lidral AC, Murray JC, Hart TC (2009) The PDGF-C regulatory region SNP rs28999109 decreases promoter transcriptional activity and is associated with CL/P. Eur J Hum Genet 17:774–784

Christensen K, Juel K, Herskind AM, Murray JC (2004) Long term follow up study of survival associated with cleft lip and palate at birth. BMJ 328:1405

Cirulli ET, Goldstein DB (2007) In vitro assays fail to predict in vivo effects of regulatory polymorphisms. Hum Mol Genet 16:1931–1939

Clifton-Bligh RJ, Wentworth JM, Heinz P, Crisp MS, John R, Lazarus JH, Ludgate M, Chatterjee VK (1998) Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nat Genet 19:399–401

Cox TC (2004) Taking it to the max: the genetic and developmental mechanisms coordinating midfacial morphogenesis and dysmorphology. Clin Genet 65:163–176

De Felice M, Ovitt C, Biffali E, Rodriguez-Mallon A, Arra C, Anastassiadis K, Macchia PE, Mattei MG, Mariano A, Scholer H, Macchia V, Di Lauro R (1998) A mouse model for hereditary thyroid dysgenesis and cleft palate. Nat Genet 19:395–398

de Lima RL, Hoper SA, Ghassibe M, Cooper ME, Rorick NK, Kondo S, Katz L, Marazita ML, Compton J, Bale S, Hehr U, Dixon MJ, Daack-Hirsch S, Boute O, Bayet B, Revencu N, Verellen-Dumoulin C, Vikkula M, Richieri-Costa A, Moretti-Ferreira D, Murray JC, Schutte BC (2009) Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. Genet Med 11:241–247

Ding H, Wu X, Bostrom H, Kim I, Wong N, Tsoi B, O'Rourke M, Koh GY, Soriano P, Betsholtz C, Hart TC, Marazita ML, Field LL, Tam PP, Nagy A (2004) A specific requirement for PDGF-C in palate formation and PDGFR-alpha signaling. Nat Genet 36:1111–1116

Dixon MJ, Marazita ML, Beaty TH, Murray JC (2011) Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet 12:167–178

Dobreva G, Chahrour M, Dautzenberg M, Chirivella L, Kanzler B, Farinas I, Karsenty G, Grosschedl R (2006) SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell 125:971–986

Dobreva G, Dambacher J, Grosschedl R (2003) SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression. Genes Dev 17:3048–3061

Durant L, Watford WT, Ramos HL, Laurence A, Vahedi G, Wei L, Takahashi H, Sun HW, Kanno Y, Powrie F, O'Shea JJ (2010) Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. Immunity 32:605–615

Fakhouri WD, Rhea L, Du T, Sweezer E, Morrison H, Fitzpatrick D, Yang B, Dunnwald M, Schutte BC (2012) MCS9.7 enhancer activity is highly, but not completely, associated with expression of Irf6 and p63. Dev Dyn 241(2):340–9

Fang X, Yoon JG, Li L, Yu W, Shao J, Hua D, Zheng S, Hood L, Goodlett DR, Foltz G, Lin B (2011) The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. BMC Genomics 12:11

FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H, Markham AF, Fantes JA, Bonthron DT (2003) Identification of SATB2 as the cleft palate gene on 2q32–q33. Hum Mol Genet 12:2491–2501

Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN et al (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 372:525–530

Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah HD, Kirschner RE, Hakonarson H (2009) A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 155(6):909–13

Grosen D, Bille C, Petersen I, Skytthe A, Hjelmborg JB, Pedersen JK, Murray JC, Christensen K (2011) Risk of oral clefts in twins. Epidemiology 22:313–319

Gu F, Hsu HK, Hsu PY, Wu J, Ma Y, Parvin J, Huang TH, Jin VX (2010) Inference of hierarchical regulatory network of estrogen-dependent breast cancer through ChIP-based data. BMC Syst Biol 4:170

Gunther T, Schmitt AO, Bortfeldt RH, Hinney A, Hebebrand J, Brockmann GA (2011) Where in the genome are significant single nucleotide polymorphisms from genome-wide association studies located? OMICS 15:507–512

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 106:9362–9367

Holler KL, Hendershot TJ, Troy SE, Vincentz JW, Firulli AB, Howard MJ (2010) Targeted deletion of Hand2 in cardiac neural crest-derived cells influences cardiac gene expression and outflow tract development. Dev Biol 341:291–304

Ingraham CR, Kinoshita A, Kondo S, Yang B, Sajan S, Trout KJ, Malik MI, Dunnwald M, Goudy SL, Lovett M, Murray JC, Schutte BC (2006) Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). Nat Genet 38:1335–1340

Jugessur A, Shi M, Gjessing HK, Lie RT, Wilcox AJ, Weinberg CR, Christensen K, Boyles AL, Daack-Hirsch S, Trung TN, Bille C, Lidral AC, Murray JC (2009) Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia. PLoS One 4:e5385

Kennedy BA, Deatherage DE, Gu F, Tang B, Chan MW, Nephew KP, Huang TH, Jin VX (2011) ChIP-seq defined genome-wide map of TGFbeta/SMAD4 targets: implications with clinical outcome of ovarian cancer. PLoS One 6:e22606

Kim SW, Yoon SJ, Chuong E, Oyolu C, Wills AE, Gupta R, Baker J (2011) Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Dev Biol 357: 492–504

Knight AS, Schutte BC, Jiang R, Dixon MJ (2006) Developmental expression analysis of the mouse and chick orthologues of IRF6: the gene mutated in Van der Woude syndrome. Dev Dyn 235:1441–1447

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289

Koster MI, Roop DR (2004) The role of p63 in development and differentiation of the epidermis. J Dermatol Sci 34:3–9

Kouwenhoven EN, van Heeringen SJ, Tena JJ, Oti M, Dutilh BE, Alonso ME, de la Calle-Mustienes E, Smeenk L, Rinne T, Parsaulian L, Bolat E, Jurgelenaite R, Huynen MA, Hoischen A, Veltman JA, Brunner HG, Roscioli T, Oates E, Wilson M, Manzanares M, Gomez-Skarmeta JL, Stunnenberg HG, Lohrum M, van Bokhoven H, Zhou H (2010) Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. PLoS Genet 6:e1001065

Kubosaki A, Lindgren G, Tagami M, Simon C, Tomaru Y, Miura H, Suzuki T, Arner E, Forrest AR, Irvine KM, Schroder K, Hasegawa Y, Kanamori-Katayama M, Rehli M, Hume DA, Kawai J, Suzuki M, Suzuki H, Hayashizaki Y (2010) The combination of gene perturbation assay and ChIP-chip reveals functional direct target genes for IRF8 in THP-1 cells. Mol Immunol 47:2295–2302

Lagha M, Kormish JD, Rocancourt D, Manceau M, Epstein JA, Zaret KS, Relaix F, Buckingham ME (2008) Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program. Genes Dev 22:1828–1837

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Leoyklang P, Suphapeetiporn K, Siriwan P, Desudchit T, Chaowanapanja P, Gahl WA, Shotelersuk V (2007) Heterozygous nonsense mutation SATB2 associated with cleft palate, osteoporosis, and cognitive defects. Hum Mutat 28:732–738

Lie RT, Wilcox AJ, Skjaerven R (1994) A population-based study of the risk of recurrence of birth defects. N Engl J Med 331:1–4

Lim LS, Hong FH, Kunarso G, Stanton LW (2010) The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. Stem Cells 28:1961–1969

Liu W, Lan Y, Pauws E, Meester-Smoor MA, Stanier P, Zwarthoff EC, Jiang R (2008) The Mn1 transcription factor acts upstream of Tbx22 and preferentially regulates posterior palate growth in mice. Development 135:3959–3968

Liu Z, Lin X, Cai Z, Zhang Z, Han C, Jia S, Meng A, Wang Q (2011) Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. J Biol Chem 286:28520–28532

Lu KT, Kanno Y, Cannons JL, Handon R, Bible P, Elkahloun AG, Anderson SM, Wei L, Sun H, O'Shea JJ, Schwartzberg PL (2011) Functional and epigenetic studies reveal multistep differentiation and plasticity of in vitro-generated and in vivo-derived follicular T helper cells. Immunity 35:622–632

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Potzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann HE, Steegers-Theunissen RP, Kramer FJ, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P, Nothen MM (2009) Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 42:24–26

Marcano AC, Doudney K, Braybrook C, Squires R, Patton MA, Lees MM, Richieri-Costa A, Lidral AC, Murray JC, Moore GE, Stanier P (2004) TBX22 mutations are a frequent cause of cleft palate. J Med Genet 41:68–74

Milunsky JM, Maher TA, Zhao G, Roberts AE, Stalker HJ, Zori RT, Burch MN, Clemens M, Mulliken JB, Smith R, Lin AE (2008) TFAP2A mutations result in branchio-oculo-facial syndrome. Am J Hum Genet 82:1171–1177

Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S, L'Heureux J, Valencia-Ramirez C, Rivera D, Lopez AM, Moreno MA, Hing A, Lammer EJ, Jones M, Christensen K, Lie RT, Jugessur A, Wilcox AJ, Chines P, Pugh E, Doheny K, Arcos-Burgos M, Marazita ML, Murray JC, Lidral AC (2009) FOXE1 association with both isolated cleft lip with or without cleft palate; and isolated cleft palate. Hum Mol Genet 18(24):4879–96

Moretti F, Marinari B, Lo Iacono N, Botti E, Giunta A, Spallone G, Garaffo G, Vernersson-Lindahl E, Merlo G, Mills AA, Ballaro C, Alema S, Chimenti S, Guerrini L, Costanzo A (2010) A regulatory feedback loop involving p63 and IRF6 links the pathogenesis of 2 genetically different human ectodermal dysplasias. J Clin Invest 120:1570–1577

Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH, Aburatani H, Miyazono K (2011) ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. Nucleic Acids Res 39:8712–8727

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC (2009) Cleft lip and palate. Lancet 374: 1773–1785

Nawshad A (2008) Palatal seam disintegration: to die or not to die? That is no longer the question. Dev Dyn 237:2643–2656

Nishiyama A, Xin L, Sharov AA, Thomas M, Mowrer G, Meyers E, Piao Y, Mehta S, Yee S, Nakatake Y, Stagg C, Sharova L, Correa-Cerro LS, Bassey U, Hoang H, Kim E, Tapnio R, Qian Y, Dudekula D, Zalzman M, Li M, Falco G, Yang HT, Lee SL, Monti M, Stanghellini I, Islam MN, Nagaraja R, Goldberg I, Wang W, Longo DL, Schlessinger D, Ko MS (2009) Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell 5:420–433

Noordeen NA, Khera TK, Sun G, Longbottom ER, Pullen TJ, da Silva XG, Rutter GA, Leclerc I (2009) Carbohydrate-responsive element-binding protein (ChREBP) is a negative regulator of ARNT/HIF-1beta gene expression in pancreatic islet beta-cells. Diabetes 59:153–160

Nopoulos P, Richman L, Andreasen N, Murray JC, Schutte B (2007a) Cognitive dysfunction in adults with Van der Woude syndrome. Genet Med 9:213–218

Nopoulos P, Richman L, Andreasen NC, Murray JC, Schutte B (2007b) Abnormal brain structure in adults with Van der Woude syndrome. Clin Genet 71:511–517

Pan D, Kocherginsky M, Conzen SD (2011) Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. Cancer Res 71:6360–6370

Pauws E, Moore GE, Stanier P (2009) A functional haplotype variant in the TBX22 promoter is associated with cleft palate and ankyloglossia. J Med Genet 46:555–561

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, Pennacchio LA, Schutte BC, Murray JC (2008) Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet 40:1341–1347

Ramos YF, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, van Galen M, van Dam H, van Ommen GJ, den Dunnen JT, Zantema A, t Hoen PA (2010) Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Res 38:5396–5408

Restivo G, Nguyen BC, Dziunycz P, Ristorcelli E, Ryan RJ, Ozuysal OY, Di Piazza M, Radtke F, Dixon MJ, Hofbauer GF, Lefort K, Dotto GP (2011) IRF6 is a mediator of Notch pro-differentiation and tumour suppressive function in keratinocytes. EMBO J 30(22):4571–85

Richardson L, Venkataraman S, Stevenson P, Yang Y, Burton N, Rao J, Fisher M, Baldock RA, Davidson DR, Christiansen JH (2010) EMAGEmouse embryo spatial gene expression database: 2010 update. Nucleic Acids Res 38:D703–709

Richardson RJ, Dixon J, Jiang R, Dixon MJ (2009) Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. Hum Mol Genet 18:2632–2642

Richardson RJ, Dixon J, Malhotra S, Hardman MJ, Knowles L, Boot-Handford RP, Shore P, Whitmarsh A, Dixon MJ (2006) Irf6 is a key determinant of the keratinocyte proliferation-differentiation switch. Nat Genet 38:1329–1334

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rodelsperger C, Guo G, Kolanczyk M, Pletschacher A, Kohler S, Bauer S, Schulz MH, Robinson PN (2010) Integrative analysis of genomic, functional and protein interaction data predicts long-range enhancer-target gene interactions. Nucleic Acids Res 39:2492–2502

Rosenfeld JA, Ballif BC, Lucas A, Spence EJ, Powell C, Aylsworth AS, Torchia BA, Shaffer LG (2009) Small deletions of SATB2 cause some of the clinical features of the 2q33.1 microdeletion syndrome. PLoS One 4:e6568

Rot-Nikcevic I, Reddy T, Downing KJ, Belliveau AC, Hallgrimsson B, Hall BK, Kablar B (2006) Myf5−/− :MyoD−/− amyogenic fetuses reveal the importance of early contraction and static loading by striated muscle in mouse skeletogenesis. Dev Genes Evol 216:1–9

Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell PJ (1996) Transcription factor AP-2 essential for cranial closure and craniofacial development. Nature 381:235–238

Schutte BC, Murray JC (1999) The many faces and factors of orofacial clefts. Hum Mol Genet 8: 1853–1859

Stevens TA, Iacovoni JS, Edelman DB, Meech R (2004) Identification of novel binding elements and gene targets for the homeodomain protein BARX2. J Biol Chem 279:14520–14530

Sturgeon K, Kaneko T, Biemann M, Gauthier A, Chawengsaksophak K, Cordes SP (2010) Cdx1 refines positional identity of the vertebrate hindbrain by directly repressing Mafb expression. Development 138:65–74

Suazo J, Santos JL, Jara L, Blanco R (2009) Association between bone morphogenetic protein 4 gene polymorphisms with nonsyndromic cleft lip with or without cleft palate in a chilean population. DNA Cell Biol 29:59–64

Tegay DH, Chan KK, Leung L, Wang C, Burkett S, Stone G, Stanyon R, Toriello HV, Hatchwell E (2009) Toriello-Carey syndrome in a patient with a de novo balanced translocation [46, XY, t(2;14)(q33;q22)] interrupting SATB2. Clin Genet 75:259–264

Thomason HA, Zhou H, Kouwenhoven EN, Dotto GP, Restivo G, Nguyen BC, Little H, Dixon MJ, van Bokhoven H, Dixon J (2010) Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. J Clin Invest 120(5):156–9

van der Deen M, Akech J, Lapointe D, Gupta S, Young DW, Montecino MA, Galindo M, Lian JB, Stein JL, Stein GS, van Wijnen AJ (2011) Genomic promoter occupancy of runt-related transcription factor RUNX2 in osteosarcoma cells identifies genes involved in cell adhesion and motility. J Biol Chem 287(7):4503–17

Van Rechem C, Boulay G, Leprince D (2009) HIC1 interacts with a specific subunit of SWI/SNF complexes, ARID1A/BAF250A. Biochem Biophys Res Commun 385:586–590

Venza M, Visalli M, Venza I, Torino C, Tripodo B, Melita R, Teti D (2009) Altered binding of MYF-5 to FOXE1 promoter in non-syndromic and CHARGE-associated cleft palate. J Oral Pathol Med 38:18–23

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145

Verzi MP, Agarwal P, Brown C, McCulley DJ, Schwarz JJ, Black BL (2007) The transcription factor MEF2C is required for craniofacial development. Dev Cell 12:645–652

Wang J, Lunyak VV, Jordan IK (2011) Genome-wide prediction and analysis of human chromatin boundary elements. Nucleic Acids Res 40(2):511–29

Weigelt K, Carvalho LA, Drexhage RC, Wijkhuijs A, de Wit H, van Beveren NJ, Birkenhager TK, Bergink V, Drexhage HA (2011) TREM-1 and DAP12 expression in monocytes of patients with severe psychiatric disorders. EGR3, ATF3 and PU.1 as important transcription factors. Brain Behav Immun 25:1162–1169

Yanagisawa H, Clouthier DE, Richardson JA, Charite J, Olson EN (2003) Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. Development 130:1069–1078

Yang J, Chai L, Fowles TC, Alipio Z, Xu D, Fink LM, Ward DC, Ma Y (2008) Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. Proc Natl Acad Sci USA 105:19756–19761

Yoon H, Chung IS, Seol EY, Park BY, Park HW (2000) Development of the lip and palate in staged human embryos and early fetuses. Yonsei Med J 41:477–484

Yun K, Im SH (2007) Transcriptional regulation of MMP13 by Lef1 in chondrocytes. Biochem Biophys Res Commun 364:1009–1014

Zhou QP, Le TN, Qiu X, Spencer V, de Melo J, Du G, Plews M, Fonseca M, Sun JM, Davie JR, Eisenstat DD (2004) Identification of a direct Dlx homeodomain target in the developing mouse forebrain and retina by optimization of chromatin immunoprecipitation. Nucleic Acids Res 32:884–892

Zhu JL, Basso O, Hasle H, Winther JF, Olsen JH, Olsen J (2002) Do parents of children with congenital malformations have a higher cancer risk? A nationwide study in Denmark. Br J Cancer 87:524–528

Zhu S, Zhou Y, Wang L, Zhang J, Wu H, Xiong J, Tian Y, Wang C (2011) Transcriptional upregulation of MT2-MMP in response to hypoxia is promoted by HIF-1alpha in cancer cells. Mol Carcinog 50:770–780

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, Natsume N, Yoshiura K, Vieira AR, Orioli IM, Castilla EE, Moreno L, Arcos-Burgos M, Lidral AC, Field LL, Liu YE, Ray A, Goldstein TH, Schultz RE, Shi M, Johnson MK, Kondo S, Schutte BC, Marazita ML, Murray JC (2004) Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med 351:769–780

Chapter 7
Cis-Regulatory Disruption at the *SOX9* Locus as a Cause of Pierre Robin Sequence

Christopher T. Gordon, Sabina Benko, Jeanne Amiel, and Stanislas Lyonnet

Abstract Mutations in the coding sequence of *SOX9* cause the severe congenital skeletal disorder campomelic dysplasia (CD). A range of genomic lesions in the region upstream of the *SOX9* coding sequence are also associated with CD, although often with milder phenotypic effects. Studies in humans and animal models suggest that these non-coding lesions disrupt *SOX9* expression in specific tissues during embryonic development. Several lesions at the *SOX9* locus, including translocations and microdeletions greater than 1 Mb upstream of the transcription start site, are associated with isolated Pierre Robin sequence (PRS), a craniofacial anomaly that is typically one part of the full-blown CD phenotype. In this chapter, we discuss how the lesions far upstream of *SOX9* suggest a requirement for craniofacial-specific regulatory elements during *SOX9* transcription in embryonic development and how the *cis*-ruption of these elements alone might result in isolated PRS, an endophenotype of CD.

Keywords Pierre Robin sequence • *SOX9* • Campomelic dysplasia • Craniofacial • Chondrogenesis • Enhancer • Conserved non-coding element • Cranial neural crest

7.1 Introduction

7.1.1 Pierre Robin Sequence

Pierre Robin sequence (PRS; OMIM 261800) is a craniofacial defect characterised by mandibular hypoplasia (micrognathia and retrognathia), U-shaped cleft secondary palate and glossoptosis (retropositioned tongue) (Fig. 7.1). These features result

C.T. Gordon (✉) • S. Benko • J. Amiel • S. Lyonnet (✉)
Department of Genetics, University Paris Descartes and INSERM U-781,
149 rue de Sèvres, Paris 75743, France
e-mail: chris.gordon@inserm.fr; stanislas.lyonnet@inserm.fr

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_7,

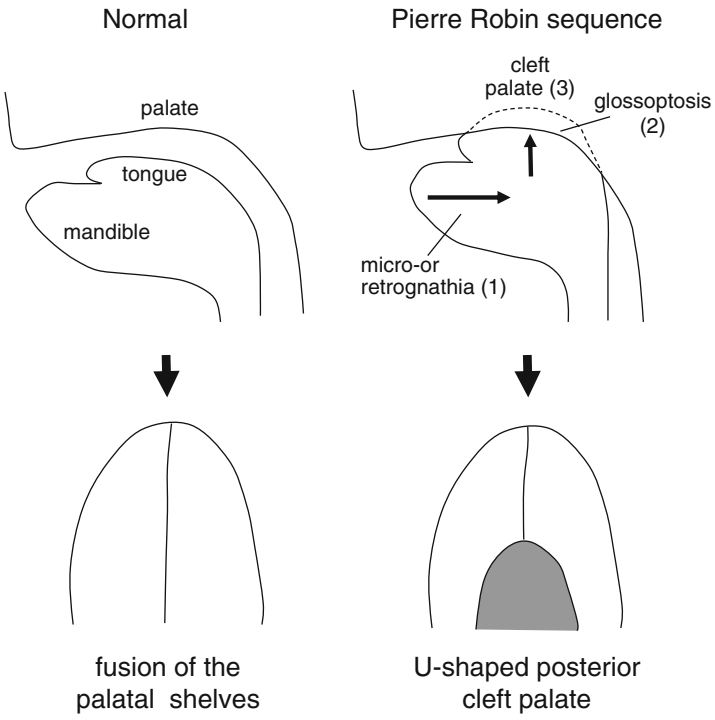

Fig. 7.1 *A schematic diagram of the developmental defects thought to underlie Pierre Robin sequence.* In the normal situation, outgrowth of the mandible allows descent of the tongue and fusion of the palatal shelves. In Pierre Robin sequence, reduced mandibular outgrowth (*1*) leads to retroposition of the tongue (*2*), preventing fusion of the palatal shelves and (*3*) resulting in a U-shaped posterior cleft palate

in respiratory and feeding difficulties in the postnatal period, typically requiring surgical intervention for cleft repair with or without tracheostomy and tube feeding. PRS is labelled as a sequence in reference to the theory that a cascade of abnormalities during foetal life would give rise to the phenotype: mandibular hypoplasia would lead to the tongue remaining posteriorly placed, resulting in physical obstruction of the paired palatal shelves and failure of palatal fusion. In this model, a defect in mandibular outgrowth would be the initiating pathogenic event. However, given that the mesenchymal and connective tissue components of both the palate and mandible are derived from cranial neural crest cells, a defect during early cranial neural crest production could also cause the PRS phenotype; therefore, in this case, PRS could be considered as a syndrome rather than a sequence (Cohen 1999). Features consistent with hindbrain dysfunction have also been reported in PRS patients, including sucking and swallowing disorders, oesophageal reflux and cardiac rhythm anomalies (Abadie et al. 2002). These models suggest heterogeneity in the events that initiate and influence the PRS phenotype.

PRS can occur as an isolated feature or may exist in the context of a syndrome (Cohen 1999; Holder-Espinasse et al. 2001; van den Elzen et al. 2001; Evans et al. 2006). Isolated PRS is typically sporadic, although a few familial cases have been reported. Genetically defined syndromes in which PRS is consistently a feature include Treacher Collins syndrome (OMIM 154500), typically caused by mutations in Treacher Collins-Franceschetti syndrome 1 (*TCOF1*), which is required for early development of cranial neural crest cells, and velocardiofacial syndrome (OMIM 192430), caused by microdeletions on 22q11.2 or rare mutations in T-box 1 (*TBX1*) (Yagi et al. 2003) (which is located within this chromosomal region), in which the development of multiple cell types within the pharyngeal arches is disrupted. The genetic bases of several rare disorders in which PRS is a component have recently been described. RNA-binding motif 10 (*RBM10*) mutations were shown to cause TARP syndrome (*t*alipes equinovarus, *a*trial septal defect, *R*obin sequence and *p*ersistent left superior vena cava; OMIM 311900), and expression of *Rbm10* in the early branchial arches of mouse embryos is consistent with the PRS component of TARP syndrome (Johnston et al. 2010). Mutations in component of oligomeric golgi complex 1 (*COG1*) at 17q25.1 were identified in two patients with a cerebrocostomandibular-like syndrome (OMIM 611209) – the phenotype included PRS in one patient and micrognathia plus high palate in the other (Zeevaert et al. 2009). A translocation disrupting the Fas-associated factor 1 (*FAF1*) gene at 1p32.3 was identified in a family displaying PRS plus some other mild craniofacial features, and experiments in zebrafish suggested that *FAF1* may function upstream of SRY (sex-determining region Y)-box 9 (*SOX9*) during development of craniofacial cartilages (Ghassibe-Sabbagh et al. 2011). PRS is also observed in a minor proportion of patients with lymphedema-distichiasis (OMIM 153400), which is caused by mutations in the transcription factor forkhead box C2 (*FOXC2*) (Tanpaiboon et al. 2010). Mutations in the collagen-encoding genes *COL2A1*, *COL11A1* and *COL11A2* are responsible for syndromic collagenopathies such as Stickler syndrome (OMIM 108300, 604841, 184840) and, less often, isolated PRS (Melkoniemi et al. 2003). PRS is also frequently a component of campomelic dysplasia (CD), caused by mutations in *SOX9* on chromosome 17q24.3 (OMIM 114290).

7.1.2 SOX9: Roles in Embryonic Development and Congenital Disease

SOX factors constitute a family of transcriptional regulators that bind DNA via the high-mobility group (HMG) domain and play key roles during many embryonic events. *SOX9* is expressed in several developing organs in human and mouse embryos (Wright et al. 1995; Ng et al. 1997; Zhao et al. 1997; Benko et al. 2009; Pritchett et al. 2010). Targeted deletion of *Sox9* in mice has revealed essential functions in a number of tissues including the heart, central nervous system, notochord, skeleton, testis, pancreas, gut and inner ear (Stolt et al. 2003; Akiyama et al. 2004a; Barrionuevo et al. 2006a; Barrionuevo et al. 2006b; Bastide et al. 2007; Seymour

et al. 2007; Barrionuevo et al. 2008). During chondrogenesis, *Sox9* is required for mesenchymal condensation as well as subsequently for cartilage differentiation (Bi et al. 1999, 2001; Akiyama et al. 2002; Barna and Niswander 2007). *Sox9* also plays an important role in neural crest production (Spokony et al. 2002; Cheung et al. 2005; McKeown et al. 2005; Sakai et al. 2006), and knockout mice have indicated that *Sox9* is essential for development of craniofacial structures (Bi et al. 2001; Kist et al. 2002; Mori-Akiyama et al. 2003).

7.1.3 Campomelic Dysplasia and Acampomelic Campomelic Dysplasia

The discovery of mutations within the coding sequence of *SOX9* in CD patients, as well as the identification of upstream translocation breakpoints, revealed a critical role for *SOX9* in human skeletal and testis development (Foster et al. 1994; Wagner et al. 1994). Features that have been described in CD patients are campomelia (bowing of the long bones, predominantly in the lower limbs), hypoplasia of the scapulae, abnormal development of the pelvic bones, congenital dislocation of the hips, hypomineralised thoracic pedicles, abnormal cervical vertebrae, a small chest, a missing pair of ribs, scoliosis and/or kyphosis, respiratory distress, talipes equinovarus (clubfeet), delayed ossification of epiphyses, short first metacarpals, XY sex reversal, relative macrocephaly, midface hypoplasia, flat nasal bridge, low-set ears, PRS, absence of the olfactory tract, congenital heart disease and renal abnormalities (Mansour et al. 1995). Death typically occurs in the postnatal period due to respiratory compromise. For patients harbouring genomic lesions upstream of the *SOX9* coding sequence, the severity of the phenotype is variable. For translocation breakpoints falling less than ~375 kb upstream (proximal translocation breakpoint cluster in Fig. 7.2), there is a tendency for campomelia and XY sex reversal to be present, while breakpoints ~932–789 kb upstream (distal translocation breakpoint cluster in Fig. 7.2) result in a phenotype without campomelia and a lower incidence of abnormal sex development (Foster et al. 1994; Wagner et al. 1994; Ninomiya et al. 1996; Wirth et al. 1996; Wunderle et al. 1998; Pfeifer et al. 1999; Hill-Harfe et al. 2005; Velagaleti et al. 2005; Leipoldt et al. 2007; Refai et al. 2010). These latter cases are referred to as acampomelic campomelic dysplasia (ACD). Hypomorphic point mutations within the *SOX9* coding sequence can also give rise to ACD (Staffler et al. 2010). Despite the absence of campomelia, features affecting the axial skeleton and face, such as scoliosis, scapular hypoplasia, pelvic abnormalities and PRS, are still frequently observed in ACD. The reduced severity of phenotype with more distant translocation breakpoints suggests that a greater proportion of the genomic domain controlling *SOX9* expression remains intact. Several large deletions upstream of *SOX9* associated with phenotypes milder than full-blown CD have also been described, and these cases provide support for the loss of specific regulatory sequences, as opposed to the possibility that the translocation cases induce non-specific position effects (Pop et al. 2004;

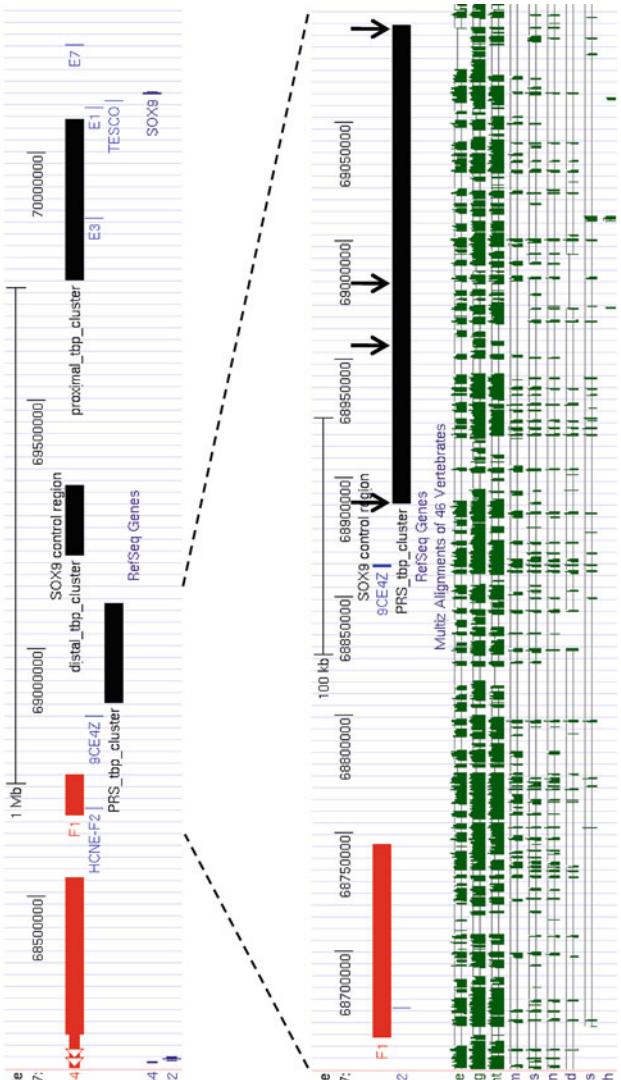

Fig. 7.2 *UCSC genome browser* (http://genome.ucsc.edu) *screenshot depicting the ~2 Mb genomic interval between KCNJ2 and SOX9. Black bars* demarcate translocation breakpoint (tbp) clusters. The proximal and distal clusters are based on data summarised by Leipoldt et al. (2007), and one case reported in Refai et al. (2010). The PRS tbp cluster is derived from Jakobsen et al. (2007) and Benko et al. (2009). *Red bars* represent microdeletions identified by Benko et al. (2009) in isolated PRS patients; note the centromeric limit of the Sp4 deletion is undefined, as indicated by the thin-ended bar. (Sp denotes sporadic and F denotes familial.) *Blue bars* represent enhancers that drive reporter expression in transgenic mice; HCNE-F2 and 9CE4Z were identified by Benko et al. (2009); E1, E3 and E7 were characterised by Bagheri-Fam et al. (2006) and *TESCO* is the testis enhancer identified by Sekido and Lovell-Badge (2008). The lower panel is an expanded view of the region containing the F1 deletion and the PRS tbp cluster. Note there are many conserved noncoding elements (CNEs) within this region, as indicated by the Multiz Alignment track. *Black vertical arrows* indicate the approximate positions of translocation breakpoints. Coordinates are for GRCh37/hg19

Lecointre et al. 2009; White et al. 2011). Interestingly, large duplications upstream of *SOX9* have recently been reported in patients with anonychia-brachydactyly (Kurth et al. 2009). Although nail and digit anomalies can occur in CD patients, it is currently unclear why these large duplications result in such a limited phenotype.

7.2 Tissue-Specific *Cis*-Regulatory Elements at the *SOX9* Locus

Enhancers situated at a large distance 5′ or 3′ from the proximal promoter of a target gene are thought to regulate tissue- and stage-specific transcription, and this may be achieved by a range of mechanisms (Bulger and Groudine 2011). At the *SOX9* locus, a number of tissue-specific regulatory elements have been identified via reporter assays in transgenic mice. In an early study, a large portion of the endogenous *Sox9* expression pattern was reproduced by *lacZ* expression driven by 350 kb of genomic sequence upstream of *SOX9* but not by 75 kb of upstream sequence, suggesting that enhancers were indeed spread over a large genomic range (Wunderle et al. 1998). The larger transgene tested by Wunderle et al. (1998) drove expression in developing skeletal tissues, consistent with the fact that several CD patients harbour translocation breakpoints that would remove from the *SOX9* locus at least part of the regulatory domain contained within the transgene. Comparison of non-coding genomic sequence between vertebrate species separated by large evolutionary distance has identified many conserved non-coding elements (CNEs) in the human genome, and several candidate *cis*-regulatory elements in the region surrounding *SOX9* have been identified and validated using enhancer assays in transgenic mice, following such analysis (Bagheri-Fam et al. 2001, 2006). For example, Bagheri-Fam et al. (2006) demonstrated that a CNE 251 kb upstream of *SOX9* drove *lacZ* expression specifically within cranial neural crest cells, branchial arch mesenchyme and the otic vesicle (E3 in Fig. 7.2). They also showed that another CNE, 95 kb downstream of *SOX9*, drove *lacZ* expression in the telencephalon and midbrain (E7 in Fig. 7.2) (Bagheri-Fam et al. 2006). An enhancer that drives reporter expression in the male gonad has also been identified (*TESCO*; testis-specific enhancer of *Sox9* core), ~10 kb upstream of the *Sox9* promoter (Fig. 7.2) (Sekido and Lovell-Badge 2008). The transcription factors sex-determining region Y (Sry; the male sex determination factor), steroidogenic factor 1 (Sf1) and Sox9 were shown to bind this testis enhancer in a dynamic fashion. Interestingly, the XY sex reversal cases associated with translocation breakpoints upstream of *SOX9* do not remove *TESCO* from the *SOX9* locus, suggesting either that other essential gonad enhancers exist further upstream, that *TESCO* function requires the presence of other general (i.e. not necessarily testis-specific) upstream enhancers or that the distant lesions alter chromatin structure at the *SOX9* locus in such a way as to have a negative, indirect influence on the function of *TESCO*.

7.3 Isolated PRS at the *SOX9* Locus

All reported translocation breakpoints less than 1 Mb upstream of *SOX9* result in a phenotype resembling either full-blown CD, or at least some anomalies of the axial skeleton and face, in ACD cases. Recently, a cluster of breakpoints further upstream than 1 Mb have been described in patients with isolated PRS (Jakobsen et al. 2007; Benko et al. 2009). The four translocation breakpoints fall 1.23–1.03 Mb upstream of *SOX9* (PRS translocation breakpoint cluster in Fig. 7.2), within the 1.9 Mb gene desert between *SOX9* and the nearest centromeric gene, potassium inwardly rectifying channel subfamily J member 2 (*KCNJ2*). Benko et al. (2009) employed high-density array comparative genomic hybridization (CGH) to screen other non-syndromic PRS patients for microdeletions in the genomic region surrounding *SOX9*. They identified two deletions situated further upstream than the PRS translocation breakpoint cluster, with 75 kb deleted in one familial case (F1 in Fig. 7.2) and >319 kb deleted in a sporadic case (Sp4 in Fig. 7.2), and one deletion of 36 kb at 1.52 Mb telomeric to *SOX9* in another sporadic case (Sp2) (not shown). Of these three deletions, F1 had the strongest association with isolated PRS, given that it segregates with several affected family members. A single base variant was also detected in another PRS family (family F2); this variant falls within a CNE (HCNE-F2 in Fig. 7.2) in the region deleted in the F1 case. In vitro, the variant sequence modified the binding of MSX1, which is required for orofacial growth and patterning in humans and mice (Satokata and Maas 1994; van den Boogaard et al. 2000). Although this variant was absent from a collection of control patients (Benko et al. 2009), it has recently been reported as a single nucleotide polymorphism (SNP; rs78542003), appearing in a sample of West Africans. This suggests that the variant may not be the sole factor contributing to the PRS phenotype in the F2 family and highlights the difficulties involved in determining the pathogenicity of any non-coding single base variant.

The clustering of translocation breakpoints and deletions greater than 1 Mb upstream of *SOX9* suggested that one or more enhancers, normally required for expression of *SOX9* during the development of tissues affected in PRS patients, may have been disrupted by these genomic lesions. To test the craniofacial activity of candidate enhancers from the region, transgenic mice were generated with two different CNEs driving expression of *lacZ* (HCNE-F2 and 9CE4Z in Fig. 7.2), and each element displayed activity within the branchial arches (Benko et al. 2009). It should be noted that within the region upstream of the PRS translocation breakpoint cluster, up to and including the F1-deleted region, there are a lot of other highly conserved elements (see the Multiz Alignment track in Fig. 7.2), whose regulatory activity is currently unknown. It is possible that the PRS phenotype in these patients may be the result of the loss of several craniofacial elements and not just the two characterised branchial arch enhancers.

It has previously been reported that the *SOX9* locus can physically associate with genomic regions greater than 1 Mb up- or downstream of the *SOX9* promoter (Velagaleti et al. 2005). Benko et al. (2009) utilised interphase fluorescence

in situ hybridization in mandibular cells expressing *Sox9* in vivo to additionally demonstrate long-range chromatin modifications surrounding the *Sox9* locus. These experiments support the possibility that the region containing distant upstream craniofacial enhancers makes contact with the *SOX9* promoter, via long-range looping, to effect appropriate *SOX9* transcription during development. A spontaneous mouse mutant, *Odd Sex*, in which a transgene insertion ~1 Mb upstream of *Sox9* induces ectopic upregulation of *Sox9* in the embryonic gonad of females and female-to-male sex reversal, provides further evidence that *Sox9* can be transcriptionally regulated by elements situated at a large distance from the promoter (Bishop et al. 2000; Qin et al. 2004).

Studies in mice with a targeted deletion of *Sox9* highlight its essential role during craniofacial development. Conditionally deleting *Sox9* in cranial neural crest cells results in cleft palate and the absence of cartilage elements that are normally derived from the neural crest (Mori-Akiyama et al. 2003). Also, heterozygous deletion of *Sox9* in all tissues, or in the cranial neural crest alone, produces the PRS-like phenotypes of cleft palate and micrognathia (Bi et al. 2001; Kist et al. 2002; Mori-Akiyama et al. 2003). This supports the argument that a reduction in *SOX9* dosage in human craniofacial tissue, as is predicted for the patients harbouring translocations and deletions at the *SOX9* locus, could cause PRS. Conversely, overexpression of *Sox9* in chick and mouse embryos also results in abnormal craniofacial development (Akiyama et al. 2004b; Eames et al. 2004), further arguing that precise regulation of *Sox9* expression is required during normal development. Finally, mutations in the collagen genes *COL2A1* and *COL11A2* are associated with isolated PRS or Stickler syndrome, and each is a direct transcriptional target of SOX9 (Bell et al. 1997; Lefebvre et al. 1997; Bridgewater et al. 1998; Liu et al. 2000; Suzuki et al. 2006). These findings collectively suggest that the loss of craniofacial enhancers far upstream of *SOX9* in patients with isolated PRS causes a reduction of *SOX9* levels within chondrogenic mesenchyme of the first branchial arch, resulting in mandibular hypoplasia and cleft palate.

The data described above are consistent with the hypothesis that PRS arises during fetal life as a sequence of events, with the initiating event being mandibular hypoplasia. However, one can also imagine that dysregulation of *SOX9* expression in other cell types could contribute to the PRS phenotype in patients with lesions far upstream of *SOX9*. *SOX9* plays a key role in early neural crest cell production in the dorsal neural tube in several animal models, and a deficit in expression at this stage may result in a failure to populate the branchial arches with adequate numbers of neural crest, potentially leading to a simultaneous reduction in growth of both the mandible and palate (both of which are first arch derivatives), as opposed to a phenotype solely originating with defective mandibular chondrogenesis. This proposed aetiology would be similar to that underlying Treacher Collins syndrome, where mutation of *TCOF1*, which is required for generation of cranial neural crest cells in mice (Dixon et al. 2006), results in craniofacial defects that include PRS. Also, *Sox9* is expressed in the mesenchyme of the palate during fusion of the palatal shelves (Yamashiro et al. 2004; Nie 2006); *SOX9* dysregulation at this discrete site could

plausibly result in the cleft palate component of PRS. Finally, given the theory that a brainstem anomaly may be involved in some PRS cases, and that *Sox9* is required for production of neural stem cells in the central nervous system (Scott et al. 2010), disruption of *SOX9* expression in neural cells cannot be excluded as a mechanism contributing to PRS.

Dysregulation of *SOX9* expression in craniofacial tissue appears to be the most likely cause of isolated PRS in the patients reported in Jakobsen et al. (2007) and Benko et al. (2009). However, it remains possible that altered expression of other genes in the 17q24.3 region may influence the development of the PRS phenotype. The gene desert upstream of *SOX9* is bordered by *KCNJ2* (see Fig. 7.2), which codes for an inward-rectifying potassium channel involved in the maintenance of resting membrane potential in muscle (Jongsma and Wilders 2001). *KCNJ2* coding sequence mutations, which are thought to function as dominant negatives (Preisig-Muller et al. 2002) cause Andersen syndrome (OMIM 170390), which is characterised by cardiac arrythmias, periodic paralysis and dysmorphic features that occasionally include micrognathia and cleft palate (Plaster et al. 2001; Tristani-Firouzi et al. 2002; Donaldson et al. 2003; Yoon et al. 2006). Mice homozygous null for *Kcnj2* display cleft secondary palate (Zaritsky et al. 2000), and the possibility that alteration of *KCNJ2* expression, due to disruption of its regulatory elements, contributes to isolated PRS cannot currently be excluded.

In the region downstream of *SOX9*, a deletion at 1.52 Mb from the *SOX9* promoter was reported in association with a sporadic case of isolated PRS (Sp2) (Benko et al. 2009). Also, a translocation breakpoint at ~1.3 Mb downstream (albeit in the context of cytogenetic anomalies on other chromosomes) was identified in a patient displaying ACD, XY sex reversal and PRS (Velagaleti et al. 2005). Although it could be speculated from these data that other craniofacial enhancers for *SOX9* may exist far downstream, this scenario is complicated by the fact that a number of genes fall within the 1.5-Mb region downstream of *SOX9*. Indeed, a splicing mutation in one of these genes, *COG1*, results in a skeletal dysplasia that has similarities to cerebrocostomandibular syndrome and includes PRS as part of the phenotype (Zeevaert et al. 2009). Also, sidekick homolog 2 (*SDK2*), which lies adjacent to the Sp2 deletion, is specifically expressed in cartilage (Day et al. 2009). Therefore, transcriptional dysregulation of *COG1* or *SDK2*, which are closer to the Sp2 deletion than *SOX9*, should also be considered as a pathogenic mechanism. It is also possible that a given enhancer may not just regulate one gene but could regulate multiple genes within a region. Given the ability of some enhancers to function over a large genomic range, it is a difficult task to definitively ascribe a target gene to any given enhancer. Perhaps, the ultimate test of the contribution of *SOX9* dysregulation to the PRS phenotype would involve analysis of mice harbouring a targeted deletion of the regulatory elements presumed to drive *Sox9* expression during craniofacial development. If disruption of *SOX9* expression really is the sole cause of the PRS phenotype in the patients reported by Jakobsen et al. (2007) and Benko et al. (2009), such mice should phenocopy human PRS, accompanied by alteration of *Sox9* expression in craniofacial tissue, without dysregulation of neighbouring genes in the region.

7.4 Perspectives

The number of highly conserved non-coding elements in the ~1.5-Mb region upstream of *SOX9* suggests that there may exist many more enhancers than that already discovered. Some of the major questions regarding *cis*-regulation of *SOX9* are as follows: how many enhancers are capable of driving expression in any one cell type, and of these, are they all essential or is there some redundancy amongst similar enhancers? There is evidence for enhancer redundancy for other *Sox* genes, where transgenic assays using elements from the *Sox10* or *Sox2* loci indicate spatially overlapping activity of separate enhancers (Uchikawa et al. 2003; Werner et al. 2007; Antonellis et al. 2008). Do multiple elements with activity in the same tissue communicate with each other and the *SOX9* promoter simultaneously, perhaps having an additive effect? It is also unclear how the different tissue-specific regulatory activities upstream of *SOX9* are co-ordinated in the 3D space of the nucleus. Enhancers with different activities may be randomly dispersed on linear genomic DNA, or there may be clusters of enhancers for particular tissues. Thus far, the craniofacial regulatory activities appear dispersed – they exist between the *SOX9* promoter and 350 kb upstream (Wunderle et al. 1998; Bagheri-Fam et al. 2006; Sekido and Lovell-Badge 2008), and also greater than 1 Mb upstream (Benko et al. 2009). The distant upstream lesions may result in isolated PRS if there is a higher density of craniofacial enhancers in the disrupted region than those for other tissues. It is also possible that mandibular outgrowth is more sensitive to slight reductions in *SOX9* expression levels than other tissues and that the distant upstream lesions in isolated PRS patients disrupt *SOX9* transcription in a mild, non-specific fashion, giving rise to an apparently tissue-specific defect.

PRS is likely to be genetically heterogeneous. Although many mice with targeted deletion of coding genes display PRS-related features, these are typically in the context of a phenotype affecting multiple organs. Similarly in humans, PRS usually occurs as part of a multi-system disorder when associated with lesions in coding genes. For the many cases of unexplained isolated PRS, perhaps disruption of regulatory non-coding DNA surrounding pleiotropic developmental genes is the underlying cause, as for the lesions at the *SOX9* locus.

Abbreviations

ACD	Acampomelic campomelic dysplasia
CD	Campomelic dysplasia
CGH	Comparative genomic hybridization
CNE	Conserved non-coding element
HMG	High-mobility group
Mb	Megabase
PRS	Pierre Robin sequence
SOX9	SRY (sex-determining region Y)-box 9

References

Abadie V, Morisseau-Durand MP et al (2002) Brainstem dysfunction: a possible neuroembryological pathogenesis of isolated Pierre Robin sequence. Eur J Pediatr 161(5):275–280

Akiyama H, Chaboissier MC et al (2004a) Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. Proc Natl Acad Sci USA 101(17):6502–6507

Akiyama H, Chaboissier MC et al (2002) The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev 16(21):2813–2828

Akiyama H, Lyons JP et al (2004b) Interactions between Sox9 and beta-catenin control chondrocyte differentiation. Genes Dev 18(9):1072–1087

Antonellis A, Huynh JL et al (2008) Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. PLoS Genet 4(9):e1000174

Bagheri-Fam S, Barrionuevo F et al (2006) Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. Dev Biol 291(2):382–397

Bagheri-Fam S, Ferraz C et al (2001) Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. Genomics 78(1–2):73–82

Barna M, Niswander L (2007) Visualization of cartilage formation: insight into cellular properties of skeletal progenitors and chondrodysplasia syndromes. Dev Cell 12(6):931–941

Barrionuevo F, Bagheri-Fam S et al (2006a) Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. Biol Reprod 74(1):195–201

Barrionuevo F, Naumann A et al (2008) Sox9 is required for invagination of the otic placode in mice. Dev Biol 317(1):213–224

Barrionuevo F, Taketo MM et al (2006b) Sox9 is required for notochord maintenance in mice. Dev Biol 295(1):128–140

Bastide P, Darido C et al (2007) Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol 178(4):635–648

Bell DM, Leung KK et al (1997) SOX9 directly regulates the type-II collagen gene. Nat Genet 16(2):174–178

Benko S, Fantes JA et al (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet 41(3):359–364

Bi W, Deng JM et al (1999) Sox9 is required for cartilage formation. Nat Genet 22(1):85–89

Bi W, Huang W et al (2001) Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci USA 98(12):6698–6703

Bishop CE, Whitworth DJ et al (2000) A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. Nat Genet 26(4):490–494

Bridgewater LC, Lefebvre V et al (1998) Chondrocyte-specific enhancer elements in the Col11a2 gene resemble the Col2a1 tissue-specific enhancer. J Biol Chem 273(24):14998–15006

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144(3):327–339

Cheung M, Chaboissier MC et al (2005) The transcriptional control of trunk neural crest induction, survival, and delamination. Dev Cell 8(2):179–192

Cohen MM Jr (1999) Robin sequences and complexes: causal heterogeneity and pathogenetic/phenotypic variability. Am J Med Genet 84(4):311–315

Day A, Dong J et al (2009) Disease gene characterization through large-scale co-expression analysis. PLoS One 4(12):e8491

Dixon J, Jones NC et al (2006) Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. Proc Natl Acad Sci USA 103(36):13403–13408

Donaldson MR, Jensen JL et al (2003) PIP2 binding residues of Kir2.1 are common targets of mutations causing Andersen syndrome. Neurology 60(11):1811–1816

Eames BF, Sharpe PT et al (2004) Hierarchy revealed in the specification of three skeletal fates by Sox9 and Runx2. Dev Biol 274(1):188–200
Evans AK, Rahbar R et al (2006) Robin sequence: a retrospective review of 115 patients. Int J Pediatr Otorhinolaryngol 70(6):973–980
Foster JW, Dominguez-Steglich MA et al (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 372(6506):525–530
Ghassibe-Sabbagh M, Desmyter L et al (2011) FAF1, a gene that is disrupted in cleft palate and has conserved function in zebrafish. Am J Hum Genet 88(2):150–161
Hill-Harfe KL, Kaplan L et al (2005) Fine mapping of chromosome 17 translocation breakpoints>or=900 kb upstream of SOX9 in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. Am J Hum Genet 76(4):663–671
Holder-Espinasse M, Abadie V et al (2001) Pierre Robin sequence: a series of 117 consecutive cases. J Pediatr 139(4):588–590
Jakobsen LP, Ullmann R et al (2007) Pierre Robin sequence may be caused by dysregulation of SOX9 and KCNJ2. J Med Genet 44(6):381–386
Johnston JJ, Teer JK et al (2010) Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate. Am J Hum Genet 86(5):743–748
Jongsma HJ, Wilders R (2001) Channelopathies: Kir2.1 mutations jeopardize many cell functions. Curr Biol 11(18):R747–R750
Kist R, Schrewe H et al (2002) Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. Genesis 32(2):121–123
Kurth I, Klopocki E et al (2009) Duplications of noncoding elements 5′ of SOX9 are associated with brachydactyly-anonychia. Nat Genet 41(8):862–863
Lecointre C, Pichon O et al (2009) Familial acampomelic form of campomelic dysplasia caused by a 960 kb deletion upstream of SOX9. Am J Med Genet A 149A(6):1183–1189
Lefebvre V, Huang W et al (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol 17(4):2336–2346
Leipoldt M, Erdel M et al (2007) Two novel translocation breakpoints upstream of SOX9 define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. Clin Genet 71(1):67–75
Liu Y, Li H et al (2000) Identification of an enhancer sequence within the first intron required for cartilage-specific transcription of the alpha2(XI) collagen gene. J Biol Chem 275(17): 12712–12718
Mansour S, Hall CM et al (1995) A clinical and genetic study of campomelic dysplasia. J Med Genet 32(6):415–420
McKeown SJ, Lee VM et al (2005) Sox10 overexpression induces neural crest-like cells from all dorsoventral levels of the neural tube but inhibits differentiation. Dev Dyn 233(2):430–444
Melkoniemi M, Koillinen H et al (2003) Collagen XI sequence variations in nonsyndromic cleft palate, Robin sequence and micrognathia. Eur J Hum Genet 11(3):265–270
Mori-Akiyama Y, Akiyama H et al (2003) Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. Proc Natl Acad Sci USA 100(16):9360–9365
Ng LJ, Wheatley S et al (1997) SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. Dev Biol 183(1):108–121
Nie X (2006) Sox9 mRNA expression in the developing palate and craniofacial muscles and skeletons. Acta Odontol Scand 64(2):97–103
Ninomiya S, Isomura M et al (1996) Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12; 17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. Hum Mol Genet 5(1):69–72
Pfeifer D, Kist R et al (1999) Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am J Hum Genet 65(1): 111–124
Plaster NM, Tawil R et al (2001) Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome. Cell 105(4):511–519

Pop R, Conz C et al (2004) Screening of the 1 Mb SOX9 5□ control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. J Med Genet 41(4):e47

Preisig-Muller R, Schlichthorl G et al (2002) Heteromerization of Kir2.x potassium channels contributes to the phenotype of Andersen's syndrome. Proc Natl Acad Sci USA 99(11):7774–7779

Pritchett J, Athwal V et al (2010) Understanding the role of SOX9 in acquired diseases: lessons from development. Trends Mol Med 17(3):166–174

Qin Y, Kong LK et al (2004) Long-range activation of Sox9 in Odd Sex (Ods) mice. Hum Mol Genet 13(12):1213–1218

Refai O, Friedman A et al (2010) De novo 12; 17 translocation upstream of SOX9 resulting in 46, XX testicular disorder of sex development. Am J Med Genet A 152A(2):422–426

Sakai D, Suzuki T et al (2006) Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development. Development 133(7):1323–1333

Satokata I, Maas R (1994) Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. Nat Genet 6(4):348–356

Scott CE, Wynn SL et al (2010) SOX9 induces and maintains neural stem cells. Nat Neurosci 13(10):1181–1189

Sekido R, Lovell-Badge R (2008) Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature 453(7197):930–934

Seymour PA, Freude KK et al (2007) SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc Natl Acad Sci USA 104(6):1865–1870

Spokony RF, Aoki Y et al (2002) The transcription factor Sox9 is required for cranial neural crest development in Xenopus. Development 129(2):421–432

Staffler A, Hammel M et al (2010) Heterozygous SOX9 mutations allowing for residual DNA-binding and transcriptional activation lead to the acampomelic variant of campomelic dysplasia. Hum Mutat 31(6):E1436–E1444

Stolt CC, Lommes P et al (2003) The Sox9 transcription factor determines glial fate choice in the developing spinal cord. Genes Dev 17(13):1677–1689

Suzuki T, Sakai D et al (2006) Sox genes regulate type 2 collagen expression in avian neural crest cells. Dev Growth Differ 48(8):477–486

Tanpaiboon P, Kantaputra P et al (2010) c. 595–596 insC of FOXC2 underlies lymphedema, distichiasis, ptosis, ankyloglossia, and Robin sequence in a Thai patient. Am J Med Genet A 152A(3):737–740

Tristani-Firouzi M, Jensen JL et al (2002) Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome). J Clin Invest 110(3):381–388

Uchikawa M, Ishida Y et al (2003) Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell 4(4):509–519

van den Boogaard MJ, Dorland M et al (2000) MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. Nat Genet 24(4):342–343

van den Elzen AP, Semmekrot BA et al (2001) Diagnosis and treatment of the Pierre Robin sequence: results of a retrospective clinical study and review of the literature. Eur J Pediatr 160(1):47–53

Velagaleti GV, Bien-Willner GA et al (2005) Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet 76(4):652–662

Wagner T, Wirth J et al (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79(6):1111–1120

Werner T, Hammer A et al (2007) Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. Nucleic Acids Res 35(19): 6526–6538

White S, Ohnesorg T et al (2011) Copy number variation in patients with disorders of sex development due to 46, XY gonadal dysgenesis. PLoS One 6(3):e17793

Wirth J, Wagner T et al (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. Hum Genet 97(2):186–193

Wright E, Hargrave MR et al (1995) The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. Nat Genet 9(1):15–20

Wunderle VM, Critcher R et al (1998) Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. Proc Natl Acad Sci USA 95(18):10649–10654

Yagi H, Furutani Y et al (2003) Role of TBX1 in human del22q11.2 syndrome. Lancet 362(9393):1366–1373

Yamashiro T, Wang XP et al (2004) Possible roles of Runx1 and Sox9 in incipient intramembranous ossification. J Bone Miner Res 19(10):1671–1677

Yoon G, Oberoi S et al (2006) Andersen-Tawil syndrome: prospective cohort analysis and expansion of the phenotype. Am J Med Genet A 140(4):312–321

Zaritsky JJ, Eckman DM et al (2000) Targeted disruption of Kir2.1 and Kir2.2 genes reveals the essential role of the inwardly rectifying K(+) current in K(+)-mediated vasodilation. Circ Res 87(2):160–166

Zeevaert R, Foulquier F et al (2009) Cerebrocostomandibular-like syndrome and a mutation in the conserved oligomeric Golgi complex, subunit 1. Hum Mol Genet 18(3):517–524

Zhao Q, Eberspaecher H et al (1997) Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. Dev Dyn 209(4):377–386

Chapter 8
Regulatory Mutations in Human Hereditary Deafness

Jonathan E. Bird and Thomas B. Friedman

Abstract Moderate to profound deafness is a common sensory deficit that is estimated by the World Health Organization to affect more than 275 million people worldwide (WHO 2010). The etiology of hearing loss is varied and can include environmental noise, physical trauma to the head, infections, ototoxic compounds, and the natural aging process. Heritable hearing loss segregating as a Mendelian trait is thought to constitute but a fraction of cases; nonetheless, its study has yielded rich information about the biology of hearing and its pathophysiology. This chapter is a critical review of gene regulation in the auditory system and draws upon the dissection of human hereditary nonsyndromic hearing loss and relevant animal models. This body of work encompasses mutant alleles of transcription factors, promoters, long-range enhancers, and microRNAs that have been associated with hearing loss including genes such as *ESSRB*, *EYA4*, *GRHL2*, *HGF*, *MIR96, POU3F4*, and *POU4F3*. At the conclusion of this chapter, we speculate how future studies can capitalize on new sequencing technologies to broaden our knowledge of gene regulation in both normal hearing and deafness.

Keywords Deafness • Cochlea • DFNA • DFNB • DFNX • *POU4F3* • *POU3F4* • *EYA4* • *MIR96* • *GRHL2* • *HGF* • *ESRRB*

8.1 Introduction

Hearing is a complex sensory phenomenon that couples the initial detection of sound with extensive neural processing in the brainstem and auditory cortex. The overall performance of this system is truly remarkable, both in terms of sound

J.E. Bird • T.B. Friedman (✉)
Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, 5 Research Court, Rockville, MD 20850, USA
e-mail: birdjo@nidcd.nih.gov; friedman@nidcd.nih.gov

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_8,

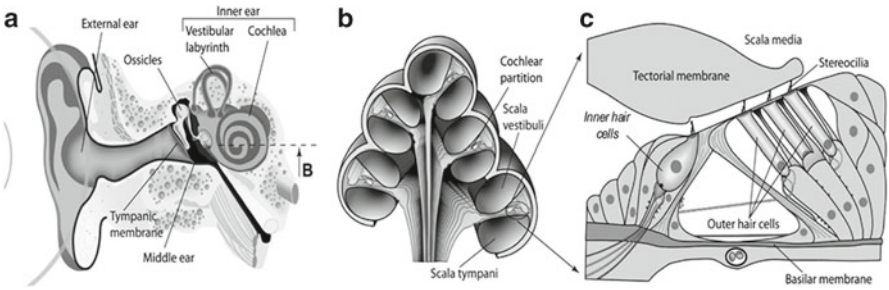

Fig. 8.1 *Structure of the human peripheral auditory system.* (**a**) Sound waves enter through the external ear and cause the tympanic membrane to vibrate. Ossicles in the middle ear amplify and transmit displacements from the tympanic membrane to the cochlea. (**b**) A cross section of the cochlea reveals three fluid-filled compartments, the scala media, tympani, and vestibuli. (**c**) A zoomed-in view of a single cochlear turn. The organ of Corti contains the inner and outer sensory hair cells. Stereocilia protrude from the apex of hair cells and are bathed in a potassium-rich endolymph contained within the scala media. Displacement of the organ of Corti relative to the tectorial membrane deflects stereocilia and opens mechanically gated ion channels that depolarize the hair cell. Inner hair cells receive primarily afferent innervation from the spiral ganglion neurons that ultimately synapse in brainstem cochlear nucleus. Outer hair cells exhibit somatic motility that forms part of the cochlear amplifier. This figure was modified with permission from Gregory Frolenkov (Frolenkov et al. 2004)

selectivity and sensitivity. The human auditory system can detect sound over a wide range of frequencies, 20 Hz–20 kHz, and yet still discriminate subtle perturbations of a few hertz. The system is also sensitive enough to perceive sound pressure changes as small as 20 micropascals, a fluctuation five billion times smaller than atmospheric pressure. Much of this performance is due to the exquisite operation of the cochlea, the sensory end organ that transduces sound into neural impulses.

Sound waves initially enter the ear via the external auditory canal where they displace the tympanic membrane (Fig. 8.1a). Oscillations of the tympanic membrane are coupled through the middle ear via the ossicular chain, a triad of tiny bones that create an impedance transformer to maximize energy transfer. The final bone in the ossicular chain, the stapes, contacts the oval window and conducts sound energy directly into the cochlea, a coiled, snail-like structure that is completely encased within the temporal bone (Fig. 8.1b). The cochlea is partitioned into three fluid-filled compartments, the scala media, vestibuli, and tympani. A flexible, collagenous basilar membrane separates the scala tympani from the scala media, such that incident sound energy entering through the oval window displaces the scala vestibuli and sets up oscillations of the basilar membrane. The mechanics of the basilar membrane vary along the length of the cochlea, the so-called tonotopic axis, such that high and low frequencies are resonant at the base and apex, respectively. By stimulating unique portions of the basilar membrane in a frequency-dependent manner, the cochlea effectively behaves like a spectral analyzer, separating sounds into their fundamental frequency components. An intricate mechanosensor, the organ of Corti, sits atop the basilar membrane and detects these frequency-coded displacements (Fig. 8.1c).

The organ of Corti is composed of hair cells, the primary sensory receptors cells of the inner ear, and a variety of nonsensory supporting cells. Hair cells are further classified as either inner or outer hair cells, each displaying distinct specializations that contribute toward the overall functioning of the cochlea (reviewed by Fettiplace and Hackney 2006). One of the most widely studied are the mechanosensitive "hairs," more correctly called stereocilia, that emerge from the cell surface and convert mechanical displacements into electrical currents. Displacements of a few nanometers are sufficient to gate ion channels at the tips of individual stereocilia and modulate cation flux into the hair cell. The resulting membrane depolarization drives neurotransmitter release at the basal pole of the hair cell and stimulates afferent neurotransmission to the brainstem cochlear nucleus via the eighth cranial nerve. Another remarkable hair cell specialization is somatic motility, a property where outer hair cells change length cyclically as their receptor potential oscillates. Somatic motility allows outer hair cells to expend energy and do work to compensate for mechanical losses within the basilar membrane, thus allowing smaller signals to be detected. These are just two of the many specializations that endow the cochlea with its unique properties.

Genetic analyses of human hereditary deafness have contributed many of the seminal insights into hair cell biology and cochlear physiology (reviewed by Richardson et al. 2011). At the core of this effort has been the ascertainment of large human pedigrees segregating an abnormal hearing phenotype inherited as a monogenic disorder. Subsequent genetic mapping using STR (short tandem repeats) or SNP (single nucleotide polymorphism) markers attempts to identify a specific chromosomal interval (locus) that associates with the hearing loss phenotype. Once defined, DNA sequencing is then used to identify potential pathological variants (causative mutations) within that locus. In every complex hearing organism amenable to genetic analysis, inherited hearing loss has been found to be highly genetically heterogeneous. To date, there are 115 human chromosomal loci that have been genetically mapped and published for hearing loss inherited as the exclusive trait (phenotype), which is referred to as nonsyndromic deafness. This is distinct from syndromic deafness, where hearing loss forms part of a more complex phenotype affecting multiple organ systems (Toriello et al. 2004). Approximately 55 nonsyndromic deafness genes with causative mutations have been identified to date (see the Hereditary Hearing Loss Homepage: http://hereditaryhearingloss.org), and thus, much work remains to uncover those underlying the remaining 60 nonsyndromic loci. A summary of mutations associated with a wide variety of human disorders, including hereditary deafness, is available from the Human Gene Mutation Database (http://www.biobase-international.com/product/hgmd). The wild-type functions of nonsyndromic and syndromic deafness genes identified to date are diverse. The products of these genes include molecular motors, cell-cell adhesion molecules, ion channels, and cytoskeleton-associated proteins as well as transcription factors/coactivators. Without exception, the identification of each of these genes has lent new understanding or in some cases unveiled a completely new aspect of inner ear biology.

Experimental animal models not only reinforce the finding that a bona fide human deafness gene or its noncoding regulatory variant has been correctly identified; they also provide a biological system in which to study the function of the wild-type gene

and the pathophysiology of deafness. In particular, mice and zebrafish have become popular models with which to study the biology of hearing. In mice, mutations causing hearing loss often arise spontaneously, and mutagenesis screens have been overwhelmingly successful in providing additional variants of known and novel deafness genes. In zebrafish, large-scale mutagenesis screens have identified genes involved in hearing, balance, and altered sensitivity to ototoxic drugs (Nicolson et al. 1998; Owens et al. 2008). By comparison, human geneticists map and identify deafness genes by taking advantage of the billions of families around the world, a small percentage of which are segregating hearing loss as a Mendelian disorder.

As gene identification progresses, it is likely we will approach an upper limit on the number involved in hereditary deafness. Cataloging the totality of genes involved in deafness is an important goal, yet ignoring for one moment the gargantuan task of discovering the cell biology specific to each of these gene products, we also need to understand how their expression is regulated throughout development and adult life. In this regard, seminal contributions have been made regarding the mechanisms of gene regulation during auditory system development, and we provide references to a few of several notable reviews (Chatterjee et al. 2010; Cotanche and Kaiser 2010; Friedman and Avraham 2009a; Kelley and Wu 2005; Soukup 2009). Of particular clinical relevance is the idea that gene regulation might be therapeutically manipulated to prevent or ameliorate deafness and potentially even restore hearing to deaf individuals.

Hair cells within the cochlea are fragile and prone to damage by excessive noise, ototoxic drugs, and aging (reviewed by Henderson and Bielefeld 2008; Ohlemiller and Frisna 2008; Rybak et al. 2008). The capacity of hair cells to tolerate insult and undergo subsequent repair is not well understood, though in cases of severe trauma, hair cells die and are removed from the organ of Corti. The adult mammalian cochlea is unable to replace hair cells, and the loss of significant numbers of hair cells results in permanent deafness. This is in striking contrast to birds, fish, and reptiles, where hair cells are readily replaced through a combination of supporting cell division and direct trans-differentiation (reviewed by Warchol 2011). Hair cell regeneration is an exciting example where manipulating existing gene regulatory pathways may have clinical utility for reversing hearing loss. The process of hair cell regeneration presumably recapitulates some of the earlier gene regulatory events that normally occur during development, and the identification of these is an active area of research (Alvarado et al. 2011; Hawkins et al. 2007, 2006). Another example is the phenomenon of cochlear preconditioning, where a moderate exposure to sound itself can provide an extended, protective effect against subsequent noise traumas (Niu and Canlon 2002). Might the gene regulatory pathways induced under these conditions be harnessed to protect hair cells?

This chapter focuses on what is known about gene regulation, marshaled from studies of mutations causing human nonsyndromic deafness inherited as a dominant or a recessive trait; syndromic deafness genes are only briefly touched upon in this chapter. We have taken a broad definition of gene regulation to include transcription factors and their transcriptional *cis*-acting elements and target genes, in addition to mutations that disrupt promoters, enhancers, repressors, microRNAs, and the gene regulatory networks in which they operate.

8.2 Nonsyndromic Deafness DFNA15 Caused by Mutations of *POU4F3*

Twelve members of a large five-generation Jewish family of Libyan descent were diagnosed with progressive hearing loss that segregated as an autosomal dominant trait. The age of onset was especially variable and first reported at 18–30 years of age (Avraham 2000; Vahava et al. 1998). Intrafamilial phenotypic variation such as this can be difficult to explain given that all of the affected individuals have the same major gene mutation. Phenotypic variability (severity, age of onset, degree of pleiotropy) of affected members within a family is common, and the causes are of general interest. Subjectivity explains some of the variation inherent to determining the age of hearing loss onset; a small initial hearing loss in adolescence is unlikely to be noticed or may be misremembered. Other explanations for inter- and intrafamiliar phenotypic variation are differences in genetic background, epigenetic changes, and environmental factors that modify the rate of hearing loss and/or severity of the fully evolved disorder.

A genome-wide screen using STRs revealed a novel deafness locus, designated *DFNA15* (MIM #602459) that was mapped to chromosome 5q31–q33 in this family. Many genes are included in this 25 cM (centimorgan) interval, and a number were ranked as likely candidates based upon their expression in the auditory system, or not. Notably, there was already a mouse deafness gene [POU class 4 homeobox 3 (*Pou4f3*)] for which the human ortholog mapped to chromosome 5q31 in the *DFNA15* locus (Vahava et al. 1998). *Pou4f3* is expressed in postmitotic auditory and vestibular hair cell nuclei, and a homozygous targeted deletion of *Pou4f3* resulted in a completely deaf mouse (Erkman et al. 1996; Hertzano et al. 2004; Xiang et al. 1997). The *POU4F3* (MIM #602460) gene is comprised of two exons and encodes a transcription factor that has dual DNA-binding sites, a POU domain and a POU_H (POU homeodomain) (Wegner et al. 1993). Depending on the downstream target gene, the cell type, and physiological context, POU transcription factors can act either as transcriptional activators or repressors (Budhram-Mahadeo et al. 1996; Dawson et al. 1996; Phillips and Luisi 2000).

The two exons of *POU4F3* were sequenced and all affected members of this family were found to be carriers of an eight base pair (bp) deletion located in exon 2. This mutation caused a translational frameshift that results in the inclusion of four missense amino acids followed by a premature stop codon. The predicted mutant protein, if synthesized and stable, would terminate in the POU_H domain. A truncating mutation is unlikely to be a benign polymorphism, although there are such examples in the literature for other genes. To examine whether the 8 bp deletion was a common polymorphism, exon 2 of *POU4F3* was sequenced in ethnically matched hearing individuals, but no carriers of this allele were detected in over 200 chromosomes (Vahava et al. 1998).

DFNA15 deafness appears to be a rare form of progressive deafness, but it is not restricted solely to the initial Israeli family. Collin and coauthors reported a dominant missense mutant allele (p.L289F) in the POU_H domain of *POU4F3* that segregated in a large Dutch family with 32 affected individuals that was not present

in 200 control chromosomes. Substitution of phenylalanine for leucine alters the DNA binding properties of the POU_H domain (Collin et al. 2008a). All of the mutant *POU4F3* alleles reported to date are dominant and associated with progressive hearing loss (Lee et al. 2010). Without additional data, the pathogenicity of these alleles could be due either to haploinsufficiency, a gain of function, or a dominant-negative process that interferes with the maintenance of hearing. Addressing this question, Weiss and coauthors examined the ability of in vitro synthesized wild-type and truncated mutant POU4F3 to bind the DNA sequence ATAATTAAT in an electrophoretic mobility shift assay (EMSA), an assay used to detect protein-nucleic acid interactions. The ATAATTAAT oligonucleotide is a POU DNA-binding target (Weiss et al. 2003). In this assay, truncated POU4F3 failed to bind and shift the mobility of the ATAATTAAT oligonucleotide. Also reported in this study were differences in stability, ability to activate transcription, and altered intracellular distribution of the mutant protein, indicating that several molecular defects caused by one copy of truncated POU4F3 may contribute to the DFNA15 phenotype.

There are no reported dominant mutant alleles of mouse *Pou4f3* that recapitulate the progressive human DFNA15 hearing loss phenotype. Unlike the dominant mutant alleles of *POU4F3* in humans, a deletion of mouse *Pou4f3* is recessive. Homozygous mutant (*Pou4f3*$^{-/-}$) hair cells develop, although never fully mature, and they undergo apoptosis at embryonic day 17 (E17). This indicates that POU4F3 is not necessary for hair cell fate specification or early differentiation events but rather appears to be required for terminal differentiation and survival of hair cells. As a result, the sensory epithelium of the neonatal organ of Corti appears to be devoid of hair cells when examined for stereocilia bundles by scanning electron microscopy (Erkman et al. 1996; Xiang et al. 1998). A similar phenotype was observed for mice homozygous for the *Pou4f3*^{ddl} allele that has a two nucleotide frameshifting deletion in exon 2 (Hertzano et al. 2004).

Several studies have reported transcriptional targets of POU4F3. Inner ears from *Pou4f3*$^{ddl/ddl}$ mice were used in a global gene expression profiling study that identified *Gfi1*, a nuclear zinc finger transcription factor, also known as growth factor independence 1, as a direct or indirect downstream target of *Pou4f3* regulation (Hertzano et al. 2004). In the absence of POU4F3, *Gfi1* expression is downregulated, and *Gfi1* mutant mice are deaf. Only outer hair cells die in *Gfi1*-deficient mice, recapitulating part of the phenotype of *Pou3f4*-deficient mice (Wallis et al. 2003). Interestingly, inner hair cells and vestibular hair cells survive longer in *Gfi1* mutant mice compared to *Pou3f4* mutant mice, indicating that additional POU4F3-regulated genes influence survival in these specific cell types.

A LIM domain transcription factor expressed in hair cells, LIM homeobox 3 (*Lhx3*), was also identified as a gene regulated by POU4F3 (Hertzano et al. 2007). The hearing phenotype of mature *Lhx3*$^{-/-}$ mice was not reported, as homozygosity for this allele results in death around birth (postnatal day 0). However, since the cochlea is already significantly developed during embryogenesis, albeit not yet fully functional, the authors examined the sensory epithelium and hair cells from E15.5 embryos cultured in vitro. Though immature, cultured cochleae from E15.5 *Lhx3*$^{-/-}$ embryos appeared grossly normal, indicating that other members of the *Lhx* family

can compensate for the loss of LHX3 or that the phenotype is subtle enough not to be evident by anatomical observations alone. It is possible that LHX3 is irrelevant for inner ear function but nevertheless regulated by *Pou4f3*, yet this appears unlikely given human genetic evidence from small pedigrees and singletons. Recessive loss-of-function nonsense and missense mutations of *LHX3* are associated with a syndrome characterized by bilateral sensorineural hearing loss, hypopituitarism, and cervical abnormalities (MIM 600577) (Bonfig et al. 2011; Rajab et al. 2008). A mouse model is now needed to confirm this association and provide insight to the cochlear pathophysiology. Embryonic lethality in *Lhx3*$^{-/-}$ embryos can hopefully be circumvented using a conditional *Lhx* allele in combination with cochlear-specific expression of Cre recombinase to obtain a restricted deletion of *Lhx*.

Another target of POU4F3 is *Caprin1* (cell cycle-associated protein 1; cytoplasmic activation and proliferation-associated protein 1), which was identified by subtractive RNA hybridization using OC-2 cells, a cell line derived from the embryonic sensory epithelia of the immortomouse (Rivolta et al. 1998; Towers et al. 2011). The immortomouse carries a transgene encoding a temperature-sensitive mutation of the SV40 large T antigen that binds p53 at the permissive temperature of 33°C permitting immortalization of cells that are otherwise difficult to grow in vitro (Whitehead and Robinson 2009). Using immortomouse-derived OC-2 cells, *Caprin1* mRNA was found to be reduced when *Pou3f4* was overexpressed and conversely was enhanced when *Pou4f3* was targeted for degradation using antisense RNA, suggesting that it is repressed by *Pou4f3* (Towers et al. 2011). The function of CAPRIN1 in the cochlea remains unknown, although it was shown to associate with hair cell RNA stress granules in response to treatment with neomycin, an ototoxic aminoglycoside antibiotic. These data indicate that CAPRIN1 may be involved in a cellular stress response pathway, though the functional relevance of this association remains to be demonstrated in an animal model.

These combined studies support an exciting hypothesis that *Pou4f3* regulates a hair cell survival pathway. The future challenge will be to understand how POU4F3, and its known target genes *Gfi1*, *Lhx3*, and *Caprin1*, function to promote hair cell survival. Do they regulate hair cell death pathways directly or perhaps work indirectly by modulating metabolic or antioxidant gene expression? Determining this will require the identification of additional POU4F3 effector genes as well as understanding how heterodimerization influences transcriptional activity. The study of this larger network promises to reveal how POU4F3 exerts its protective effects and is an excellent example where gene regulation might be manipulated clinically to preserve hair cells during normal aging or following environmental or pharmacological insult.

8.3 Mutations of *POU3F4* Cause Sex-Linked Nonsyndromic Deafness DFNX2

Five loci for nonsyndromic deafness have been mapped to the X chromosome (*DFNX1-DFNX5*). Currently, only mutations in three genes [phosphoribosyl pyrophosphate synthetase 1 (*PRPS1*, *DFNX1*); POU class 3 homeobox 4 (*POU3F4*;

DFNX2); small muscle protein, X-linked (*SMPX*, *DFNX4*)] have been reported. Recessive mutations at the *DFNX2* locus (previously designated *DFN3*) are the most common cause of sex-linked deafness although *DFNX2* alleles overall are responsible for only a small percent of human nonsyndromic deafness. DFNX2 is characterized by rapidly progressing hearing loss and reduced penetrance for a conductive component due to fixation of the stapes footplate (Bitner-Glindzicz et al. 1995). Individuals with conductive DFNX2 hearing loss may choose a stapedectomy procedure to correct the conductive hearing loss. However, mobilization of the fixed stapes footplate can result in the rapid outflow of perilymphatic fluid with associated hearing loss, referred to as a perilymphatic gusher.

DFNX2 hearing loss was mapped to Xq21.1 by linkage analyses and refined cytogenetically by overlapping deletions and microdeletions in subjects with X-linked hearing loss (Brunner et al. 1988; Wallis et al. 1988). Mouse *Pou3f4* (POU domain, class 3, transcription factor 4) was already known as a sex-linked gene associated with hearing loss in this species, and therefore, human *POU3F4* (also referred to as *BRN4*) was a good positional candidate for DFNX2 hearing loss (de Kok et al. 1995b). Several mutant alleles including small deletions and point mutations were reported in the single protein-coding exon of *POU3F4* (Bitner-Glindzicz et al. 1995; Cremers et al. 2000; de Kok et al. 1995b). A more fascinating class of mutations associated with DFNX2 hearing loss are chromosomal anomalies located approximately 1 Mb upstream of the *POU3F4* coding region. Chromosomal inversions and microdeletions have been reported in this region that potentially disrupt distant *cis*-acting regulatory elements of *POU3F4* (de Kok et al. 1995a, 1996; Naranjo et al. 2010). Alternatively, structural changes can alter the chromosomal neighborhood of an otherwise wild-type *POU3F4* resulting in anomalous gene expression, a phenomenon referred to as a position effect (reviewed by Kleinjan and van Heyningen 1998), first described nearly a century ago in the fruit fly by A. H. Sturtevant.

Guided by the position of microdeletions and inversions in human-affected subjects, several groups have now attempted to identify specific regulatory elements upstream of the *POU3F4* coding region. Ahn and coauthors focused on one region of high sequence conservation shared in all species from frogs to humans, located approximately 920 kilobase (kb) upstream of *POU3F4*. A 3.4-kb fragment flanking this element was fused to a minimal promoter driving Cre recombinase and used to generate transgenic mice. When crossed against a ROSA reporter mouse, β-galactosidase activity was detected in several structures within the inner ear, confirming that this fragment included a *cis*-acting regulatory element conferring otic tissue specificity (Ahn et al. 2009). A related approach, using GFP rather than Cre recombinase as a reporter, was used to examine the activity of three separate, highly conserved noncoding regions (HNCR) upstream of *POU3F4* (Naranjo et al. 2010; Robert-Moreno et al. 2010). Individually, each of these regions could drive expression of GFP in the developing zebrafish inner ear, though each element conferred a slightly different timing and domain of expression. In addition, two transcription factors that have been well documented to participate in inner ear development, *Pax2* and *Sox2*, were shown to interact with one of these regions, HNCR 81675, by chromatin immunoprecipitation (ChIP) (Robert-Moreno et al. 2010).

Different mutant alleles of mouse *Pou3f4* (synonyms include *Brn4* and *Otf9*) have shed light on the function of this transcription factor in the inner ear. Two of these mutations are null alleles, while *Pou3f4*^{Slf} (sex-linked fidget) is a radiation-induced chromosomal inversion with one breakpoint near *Pou3f4*. This rearrangement does not alter the protein-coding region of *Pou3f4*. *Pou3f4*^{Slf} represses *Pou4f3* expression in the otic capsule, yet there is no altered expression of *Pou4f3* evident in the neural tube (Phippard et al. 2000). The location of the *slf* inversion breakpoint further supports the presence of an evolutionarily conserved but distant upstream regulatory region for *POU3F4*/*Pou3f4* that is cochlear specific.

The two reported engineered null alleles of *Pou3f4* have different phenotypes that are difficult to reconcile (Minowa et al. 1999; Phippard et al. 1999). Minowa and coauthors replaced the entire coding region of *Pou3f4* with a PGK neomycin resistance cassette (*Pou3f4*^{tm1Tno}) and found that 11-week-old mice were profoundly deaf as measured by ABR (auditory brainstem response) analysis. In contrast, the knockout allele of *Pou3f4* in which a lacZ cassette was substituted for the single protein coding exon (*Pou3f4*^{tm1Cren}) produced mice that were reported to have cochlear bone dysplasia and only a mild hearing loss assessed using Preyer's reflex (Phippard et al. 1999). The disparity in hearing loss between these two null alleles of *Pou3f4* could be due to the differing methodologies used to measure hearing. While ABR analysis is an objective measure of hearing, Preyer's reflex is manifest as a flicking of the mouse external ear (pinna) in response to a sudden onset sound stimulus (Jero et al. 2001). The use of Preyer's reflex is highly subjective and insensitive to anything other than profound deafness. Other possible explanations for the phenotypic difference between these two *Pou3f4* null alleles could be environmental or due to variations in the genetic background.

Modifier variants are difficult to identify. Examples of successful experiments that pinpointed modifier alleles with significant impact on hearing ability include mouse *mdfw* that modifies the phenotype of mice heterozygous for the *dfw* (deaf-waddler) allele (Noben-Trauth et al. 2003, 1997). In humans, recessive nonsyndromic deafness DFNB26, which was mapped to chromosome 4, is completely suppressed by one copy of a rare dominant allele at the *DFNM1* locus on chromosome 1 (Riazuddin et al. 2000). A second modifier of auditory function in humans is a variant of ATPase, Ca^{2+}-transporting plasma membrane 2 (*ATP2B2* or *PMCA2*) that moderates the severity of sensorineural hearing loss due to a mutation in cadherin-related 23 (*CDH23*) (Friedman et al. 2000; Schultz et al. 2005). A third example of an auditory genetic modifier is a promoter variant of an otherwise wild-type *MYO7A* that modifies the extent of DFNA11 progressive low-frequency hearing loss in a large family (HL2) of English decent (Street et al. 2011). Mutations of *MYO7A* can cause either Usher syndrome type 1B (MIM #276900), nonsyndromic deafness DFNB2 (MIM #600060), or dominant DFNA11 (MIM # 601317) progressive hearing loss (Friedman et al. 2011; Riazuddin et al. 2008). All the affected individuals in family HL2 carried a *MYO7A* glycine-to-arginine amino acid substitution (p.G772R). However, the degree of hearing loss in family HL2 varied considerably, and this was associated with a single nucleotide SNP (T/C) in the promoter of the wild-type *MYO7A* allele. Additional data suggested that the T^{-4128} allele

reduces the level of transcription of the wild-type *MYO7A* allele *in trans* to the p.G772R mutant. This combination exacerbates DFNA11 hearing loss in comparison to the less severe hearing deficit of DFNA11 individuals in family HL2 carrying the C^{-4128} allele (Street et al. 2011). It will be interesting to determine if homozygotes for the *MYO7A* T^{-4128} modifier allele have a hearing loss phenotype.

8.4 Mutations of the Transcriptional Coactivator *EYA4* Are Associated with Nonsyndromic Progressive Hearing Loss DFNA10 and Loss of Hearing Coupled with Dilated Cardiomyopathy

DFNA10 is an autosomal dominant progressive hearing loss locus that was genetically mapped to chromosome 6q22.3-q23.2 in a four-generation family from the United States (O'Neill et al. 1996). The onset of hearing loss began at ages ranging from 20 to 50 years and progressed to severe/profound deafness (De Leenheer et al. 2001, 2002; Verstreken et al. 2000). An additional two families refined the *DFNA10* locus to a 3.7 cM region (Verhoeven et al. 2000). Among the genes in the *DFNA10* interval is *EYA4*, the ortholog of "eyes absent Drosophila homolog 4," a transcriptional coactivator required, as the name implies, for fruit fly eye development. Affected members of the three unrelated DFNA10 families were each found to be heterozygous for one of three different mutant alleles of *EYA4*, each predicted to cause premature termination of translation within the EYA domain (Wayne et al. 2001).

EYA4 is one of four paralogs (*EYA1-EYA4*) of the EYA family of transcription factors. Human *EYA4* has 21 exons and encodes a highly conserved C-terminus of approximately 270 residues, referred to as the EYA domain. EYA4 has no known or predicted DNA binding domain and is not thought to bind DNA directly. Instead, the EYA domain provides an interaction interface for the SIX homeodomain-containing transcription factors and for two DACH paralogs (Bonini et al. 1998; Borsani et al. 1999; Hanson 2001). For example, SIX homeobox 3 (SIX3) interacts with the EYA domain of EYA4 (Abe et al. 2009). What function does EYA4 bring to this complex? The EYA4 N-terminus of approximately 360 residues appears to function as a *trans*-activator of the downstream target genes of the complex (Ohto et al. 1999). Once assembled, the SIX transcription factor, EYA4, and a DACH family member form a tripartite transcription factor complex that shuttles into the nucleus and together acts to activate or repress downstream target genes as part of a network that regulates the development and maintenance of a wide variety of organ systems (reviewed by Christensen et al. 2008).

EYA proteins of plants, flies, mice, and humans also possess intrinsic phosphatase activity toward transcriptional cofactors, in addition to the EYA protein itself (Jemc and Rebay 2007; Li et al. 2003; Rayapureddi et al. 2003; Tootle et al. 2003). A carboxy-terminus haloacid dehalogenase domain in EYA4 targets phosphotyrosine residues, while the amino-terminal domain targets phosphothreonine (Okabe et al. 2009). The target substrates of the EYA4 phosphatase in vivo and their

biological relevance to inner ear function remain unknown. Is the phosphatase activity important for EYA4 transactivation in the auditory system? In a more general sense, the presence of a transcriptional cofactor with enzymatic activity raises the possibility that the converse may also be true. Are there enzymes that have unrecognized functions as transcription factors or coactivators?

Aside from mutations that truncate the C-terminus EYA domain of EYA4 and are associated with nonsyndromic deafness, other mutations of *EYA4* have been associated with a syndromic cardio-auditory phenotype (MIM # 605362). In a single large family, a 4,846-bp deletion of *EYA4,* including exons encoding part of the N-terminus region and the entire EYA domain, was associated with dilated cardiomyopathy (DCM) with reduced penetrance and coinheritance of juvenile onset progressive hearing loss. This deletion was not found in 300 control chromosomes (Schonberger et al. 2000, 2005). These data indicate that deafness and DCM result when both the N-terminus and EYA domain are disrupted. Since DCM is often a late-onset disorder, was DCM overlooked in the affected individuals of the original DFNA10 families? In this regard, Makishima and coauthors (2007) evaluated nine DFNA10-affected individuals from yet another North American Caucasian family of European ancestry for pleiotropic effects of a truncating mutation of *EYA4*. Members carrying a dominant *EYA4* frameshift mutation that leaves the N-terminus variable region intact, but deletes the EYA domain, were examined for a potential heart phenotype. In this report, electrocardiograph, echocardiograph, and magnetic resonance imaging studies revealed no evidence for DCM (Makishima et al. 2007). On the basis of these data, a genotype-phenotype relationship has been proposed for *EYA4*, where heterozygous truncations that include the N-terminal transactivation domain result in deafness and DCM, while heterozygous downstream truncations of the Eya domain alone are associated with hearing loss only. The human phenotype for homozygous *EYA4* mutations has not been reported. To date, only a single heterozygous mutant allele associated with hearing loss and DCM has been reported (Schonberger et al. 2005). One concern is that coinheritance of deafness and DCM by chance alone might mask. two independent genetic etiologies. In support of a mutated *EYA4*-mediated cardiomyopathy/deafness phenotype, expression of EYA4 is found in both the heart and inner ear (Schonberger et al. 2005). Moreover, four different antisense morpholinos designed to downregulate zebrafish *eya4* (68% identical in amino acid sequence to human EYA4) showed cardiovascular abnormalities and compromised ventricular function (Schonberger et al. 2005). These data indicate that EYA4 has an evolutionarily conserved, crucial function in the heart. This is not surprising if the Mikado's Pooh-Bah can trace his ancestry back in time to a "protoplasmal primordial atomic globule."

The direct effects on gene expression in the auditory system caused by mutations of *EYA4* are largely unknown, although there is an engineered *Eya4* mutant mouse (*Eya4*$^{-/-}$) that could be used to investigate this (Depreux et al. 2008). The *Eya4*$^{-/-}$ mouse was constructed by deleting exons 8 through 10 and replacing this deleted sequence with a PGK neo and zeocin cassette. On a 129S6/SvEv background, homozygous mutant mice die just after birth, while homozygous mice on a mostly CBA/J background are viable, although males are sterile. The hearing of these

Eya4$^{-/-}$ mice was evaluated by ABR and distortion product otoacoustic emissions (DPOAE) and found to be profoundly deaf while wild-type controls had normal hearing. Homozygous *Eya4*$^{-/-}$ mice had otitis media (effusion and inflammation) regardless of the four different genetic backgrounds and a variety of inner ear abnormalities including maldevelopment of the tympanic membrane and Eustachian tube. Since *Eya1*, SIX homeobox 1 (*Six1*) and F-box protein 11 (*Fbxo11*) mutant mice also display increased otitis media susceptibility, expression of these genes were examined in *Eya4*$^{-/-}$ and wild-type littermates. Expression levels of *Eya1*, *Six1* (at E12.5), and *Fbxo11* (at P1) were found to be comparable to wild-type littermates and thus unlikely to be directly regulated by EYA4 (Depreux et al. 2008).

In order for the *Eya4* null allele to be an accurate model of DFNA10 deafness, it is important to determine that a heterozygous *Eya4* mutant mouse (*Eya4*$^{+/-}$) can recapitulate the human phenotype. The hearing phenotype of a heterozygote was not originally reported (Depreux et al. 2008), but unpublished data were generously provided by M. Charles Liberman. Heterozygote *Eya4*$^{+/-}$mice appear to have wild-type hearing, and thus, the null allele of mouse *Eya4* is not an accurate model for human progressive deafness DFNA10. It is important to keep in mind that in mice, an allele comparable to a pathogenic mutation in humans may not recapitulate the human disorder. It appears that haploinsufficiency of EYA4 is an unlikely explanation for DFNA10 hearing loss and that either a dominant negative or gain of function is the more likely pathological mechanism. It remains to be seen what gene networks are regulated by EYA4 in the cochlea and how these are perturbed in DFNA10 deafness.

8.5 Mutations of *MIR96* Cause Human DFNA50 Progressive Deafness and the Diminuendo Phenotype in Mouse

Dominantly inherited, postlingual, progressive deafness segregating in a large multigenerational Spanish family was genetically mapped with a LOD score of 10.7 to markers defining a novel 3.8 cM interval on chromosome 7q32, designated *DFNA50* (Modamio-Hoybjor et al. 2004). The earliest perceived hearing loss was at 12 years of age and affected frequencies from 250 to 8,000 Hz. Since hearing was near normal for the first decade of life, the inner ears of affected individuals presumably developed normally but were unable to function correctly. All of the affected members of this Spanish family were found to be carriers of a transition mutation (+13G>A) within the 7-nucleotide seed region of *MIR96* encoding microRNA 96 (synonyms, *MIRN96*, *has-mir-96*; *miR-96*) (Mencia et al. 2009). In a second small Spanish family, also segregating progressive deafness as a dominant trait, a transversion (+14C>A) mutation (adjacent nucleotide to +13G>A) was identified, also in the seed region of *MIR96*. Ophthalmologic evaluation revealed no obvious retinal phenotype in carriers of either of the two *MIR96* mutations, despite expression in photoreceptor cells (Mencia et al. 2009). At the time of publication in 2009, this was the first report of a microRNA mutation responsible for a monogenic human disorder.

A second example of a Mendelian disorder due to mutations of a microRNA gene cluster was reported recently (de Pontual et al. 2011). Microdeletions of *MiR-17-92* (*MIR17HG*) are associated with Feingold syndrome (MIM #164280), which is characterized by microcephaly, short stature, digital anomalies, and a variety of other features with reduced penetrance including hearing loss (Feingold et al. 1997).

MicroRNAs are an ancient class of noncoding single-stranded RNAs, approximately 20–24 nucleotides long found in animals and plants that regulate post-transcriptional gene expression (reviewed by Brodersen and Voinnet 2009). MicroRNAs have highly conserved, cell-specific patterns of expression among divergent species, and each is predicted to interact with hundreds of potential mRNA targets (Christodoulou et al. 2010). Newly discovered microRNAs continue to be reported, with at least 1,000 identified in mammals. At least 150 of these microRNAs are expressed in the inner ear (Elkan-Miller et al. 2011; Friedman et al. 2009b; Weston et al. 2011), where they have now been implicated in diverse processes such as cell specification, development, and hair cell homeostasis (Kuhn et al. 2011; Li and Fekete 2010a). The microRNA-mediated regulatory process occurs when a mature microRNA anneals to a target mRNA, frequently in its 3′ untranslated region (3′ UTR), delivering with it the RNA-induced silencing complex (RISC). Once associated with the target mRNA, microRNAs generally repress gene translation by either blocking translation or promoting mRNA degradation via RISC endonuclease activity (Guo et al. 2010), although there are now examples of microRNAs that can enhance target mRNA translation (Lin et al. 2011a; Orom et al. 2008). The specificity of a microRNA annealing to its target mRNA is directed primarily by a 7-nucleotide core "seed" sequence. The computational identification of mRNA targets with the complementary sequence would then seem trivial; however, the effects of wobble base pairing in RNA-RNA duplexes complicate this effort. Furthermore, efficient mRNA degradation can occur in the presence of significant mismatches to the microRNA. Thus, experimentally establishing and validating the specific targets of a microRNA and how these are altered by a mutation presents a daunting challenge.

The human *MIR183* family of *MIR96*, *MIR182*, and *MIR183* on chromosome 7q32.2 is transcribed as a polycistronic RNA, which in principle provides stoichiometric amounts of each microRNA. In zebrafish and mice, *Mir96*, *Mir182*, and *Mir183* are expressed in sensory cells of the olfactory epithelium, retina and inner ear hair cells (Friedman et al. 2009b; Weston et al. 2006, 2011; Wienholds et al. 2005; Xu et al. 2007). What mRNA transcripts in the inner ear are regulated by *MIR96*? How do mutations in the seed region of *MIR96* result in hearing loss? Do *MIR96* mutations increase or decrease the half-life of target mRNAs, or do they shift the specificity (off-target effects) of *MIR96* to anneal to mRNAs not normally targeted in the wild type? Using the programs miRanda, TargetScan, and PITA, Mencia and coauthors computationally identified 700 potential targets of *MIR96* (Mencia et al. 2009). As a proof of principle, five of them [aquaporin 5 (*AQP5*), cadherin, EGF LAG seven-pass G-type receptor 2 (*CELSR2*), outer dense fiber of sperm tails 2 (*ODF2*), myosin VIIA Rab interacting protein (*MYRIP*), and RYK receptor-like tyrosine kinase (*RYK*)] were examined further using a luciferase reporter coupled to the 3′ UTRs of these genes. Using luciferase activity as a proxy for luciferase mRNA

levels, wild-type *MIR96* was shown to downregulate these transcripts. Critically, this downregulation was impaired when *DFNA50* mutant *MIR96s* were tested. In addition to examining the regulation of *bona fide* target mRNAs, it is equally important to consider the possibility that there may be off-target effects induced by mutant *MIR96*. Is DFNA50 hearing loss caused by off-target effects of mutant *MIR96* and/or altered regulation of genuine *MIR96* targets? One way to test this would be to find and examine deletion heterozygotes of *MIR96*. If these individuals were to have normal hearing at an older age, this would exclude haploinsufficiency as the mechanism responsible for hearing loss and indicate that the action of the *DFNA50* mutation is likely aberrant regulation of "off-target" genes.

Questions about the molecular pathogenicity of human *MIR96* mutations are beginning to be answered using the diminuendo deaf-circling mouse that arose from an ENU-induced mutagenesis screen. Genetic mapping of the diminuendo hearing loss locus and positional cloning of the responsible mutant gene identified an A>T transversion of *Mir96* (synonyms *mir 96*, *Mirn96*, *mmu-mir-96*) that altered the seed region (wt, TTGGCACT>diminuendo, TGGCTCT) of the mature microRNA. The mouse *Mir96* and human *MIR96* ortholog have identical sequence in this region. The diminuendo allele is referred to as *Mir96*^{Dmdo} and was shown to be a semidominant allele since there is a hearing loss in the heterozygote and more severe abnormalities in the homozygote (Lewis et al. 2009). Homozygous mutant mice show early-onset hair cell loss that progresses quickly to profound deafness, while heterozygotes have an intermediate phenotype of progressive hearing loss beginning at 15 days of age (P15) that recapitulates the human DFNA50 phenotype (Kuhn et al. 2011; Lewis et al. 2009). The inner ear phenotype of the homozygous *Mir96*^{Dmdo} mouse is unusual. Hair cells in *Mir96*^{Dmdo} homozygotes appear to never fully mature, instead retaining the electrophysiological signature and morphological architecture of late embryonic hair cells, before eventually degenerating (Kuhn et al. 2011; Lewis et al. 2009).

How might *Mir96* regulate the maturation of hair cells? Are subtle perturbations in hundreds of mRNAs collectively responsible for the arrest of hair cell development in the diminuendo mouse? Alternatively, among the transcriptional "noise," are there a small number of crucial gene expression changes that account for the phenotype? Microarray analyses have revealed several potentially relevant findings. Lewis et al. (2009) compared gene expression in wild-type P4 (postnatal day 4) and *Mir96*^{Dmdo} organ of Corti and found many differences; for example, solute carrier family 26, member 5 (*Slc26a5*) encoding prestin, the outer hair cell somatic motor; *Gfi1*, a target of POU4F3 regulation; and protein tyrosine phosphatase, receptor type, Q (*Ptprq*), a phosphatase required for stereocilia shaft connector formation (Goodyear and Richardson 2003) were all downregulated. However, *Mir96*^{Dmdo} probably indirectly regulates these changes since the mRNAs encoded by these genes do not have sequence complimentary to *Mir96*. Nevertheless, mutations of these three genes in mouse are individually known to cause deafness, and their dysregulation may cumulatively contribute to hearing loss (Richardson et al. 2011). In addition, microphthalmia-associated transcription factor (*MITF*), a direct target of wild-type *MIR96*, encodes the MITF transcription factor that

is mutated in a human auditory-pigmentary syndrome (WS2A, Waardenburg syndrome type 2A, MIM #193510) and in the micropthalmia phenotype in mouse (reviewed by Schultz 2006). These data integrate *MIR96* into a mammalian neurosensory regulatory network for both syndromic and nonsyndromic deafness (Li and Fekete 2010a; Xu et al. 2007).

It remains to be seen whether perturbations in a few key pathways or a more general catastrophic gene dysregulation underlies the pathology of DFNA50 deafness. Since microRNAs can potentially block translation, in addition to regulating transcript stability, measuring mRNA transcript levels reveals only part of the story. In order to fully understand how MIR96 functions in the inner ear, an exhaustive catalog of transcriptional changes needs to be compared alongside proteomic datasets from mutant hair cells. As discussed toward the end of this chapter, these types of datasets are not simple to assemble and are a challenge for the future.

8.6 A Mutation of *GRHL2* Is Associated with Progressive Deafness DFNA28

Dominant mutations of genes encoding transcription factors *POU3F4*, *POU4F3*, and *EYA4* are associated with progressive hearing loss, also referred to as adult-onset hearing loss. The underlying pathology of this disorder can be particularly difficult to dissect; specifically, how to exclude the possibility of a subtle developmental defect that renders the adult auditory system less resilient to environmental stressors? Distinguishing between defective manufacture during development, as opposed to defective maintenance during adult life, is experimentally challenging, and this question remains unanswered for all inherited late-onset deafness in both mouse and man. DFNA28 progressive deafness is no exception.

A single large family was observed to cosegregate dominant, postlingual hearing loss that progressed with age in the high-frequency regions, similar to presbycusis; the slow, age-related neurosensory hearing loss that is a common disorder of the elderly. The phenotype in this family was genetically mapped to a 1.4 cM interval of chromosome 8q22 (Peters et al. 2002). The locus was designated *DFNA28* (MIM #608641) and encompassed seven annotated genes. The exons and adjacent intronic sequence of six of these genes were sequenced, and a 1-bp insertion (c.1609-1610insC) in exon 13 of grainyhead-like 2 (*GRHL2*; previous nomenclature *TFCP2L3*, transcription factor cellular promoter 2-related to TFCP2) was identified. This alteration resulted in a predicted premature translation stop codon in exon 14. The c.1609-1610insC frameshift insertion cosegregated with hearing loss in the four-generation family (nine affected members) and was not found in genomic DNA from 150 Caucasian and pan-ethnic control individuals. No additional DFNA28 pedigrees have been reported, though variants of *GRHL2* may be associated with presbycusis. A study by Van Laer and coauthors examined 768 single nucleotide polymorphisms (SNPs) tagging 70 known deafness genes in 2,418 DNA samples from subjects with age-related hearing loss (Van Laer et al. 2008).

Although no SNPs in this limited association study had p-values exceeding the Bonferroni-corrected threshold for multiple testing, several SNPs in intron 1 of *GRHL2* came close to doing so. As of yet, the actual causative variants of *GRHL2* that may contribute to presbycusis have not been identified. A separate study in the Han Chinese population sought to replicate the association of *GRHL2* polymorphisms and age-related hearing loss but was unable to do so (Lin et al. 2011b). This may represent different susceptibility loci to presbycusis in different ethnic groups.

Human *GRHL2* is similar in amino acid sequence to the *Drosophila melanogaster* grainyhead gene (*grh*, *Elf-1*) that is predominantly expressed in surface ectoderm (Biggin and Tjian 1988; Bray and Kafatos 1991; Ostrowski et al. 2002) and is required for wound repair (Mace et al. 2005). In vertebrates, there are three members (*GRHL1*, *GRHL2*, and *GRHL3*) of the grainyhead-like family of transcription factors that can form homo- or heterodimers with one another (Ting et al. 2003; Wilanowski et al. 2002). Mammalian *GRHL2* encodes transactivation, DNA binding and dimerization domains. Unless there is an isoform of *GRHL2* that can splice around exon 13, the c.1609-1610insC frameshift mutation is predicted to introduce 10 novel amino acids before a translation stop codon truncates *GRHL2* and removes the majority of the dimerization domain (Peters et al. 2002). If this mRNA and/or translated peptide is stable, the c.1609-1610insC mutation may create a dominant negative or gain-of-function variant. Alternatively, the introduction of a premature translation stop codon by the c.1609-1610insC mutation might target the *Grhl2* transcript for nonsense-mediated mRNA decay (NMD). In this case, the progressive hearing loss phenotype could be due to haploinsufficiency of GRHL2 in the auditory system. Since the phenotype is dominant and nonsyndromic, one dose of wild-type *GRHL2* would be sufficient for other organ systems to function normally. Although NMD is commonly invoked to explain the pathogenicity of a nonsense mutation, solid experimental evidence is required to substantiate its involvement. A striking example can be found in a study of the *Myo7a polka* allele, where a nonsense mutation triggers NMD in the cochlea, but not in the retina, where instead a stable protein hypomorph is produced (Schwander et al. 2009). This highlights the need for a mouse model carrying the humanized c.1609-1610insC mutation to properly understand the molecular pathology of DFNA28 deafness.

In a variety of organ systems, GRHL2 is essential for epithelial cell differentiation, neural tube closure and wound healing. These processes involve *trans*-activation of genes encoding E-cadherin, claudin-4, and RhoGEF19, all of which are components of the apical junctional complex (Boglev et al. 2011; Pyrgaki et al. 2011; Werth et al. 2010). In all likelihood, the majority of direct transcriptional targets of GRHL2 are yet to be identified. What might the functions of GRHL2 be in the auditory system? In inner ears of the wild-type mouse, *Grhl2* is expressed in all the epithelial cells, including sensory hair cells, which line the developing cochlear duct (Peters et al. 2002). Studying the mature auditory phenotype of a *Grhl2* null mouse is not currently possible as embryos die at E11.5 from neural tube closure failure (Werth et al. 2010). A conditional knockout of mouse *Grhl2* has not been reported. Experimentally manipulating the expression of *Grhl2* in the various

cell types of the auditory system will be critical in defining the roles of this transcription factor in bringing about and maintaining normal hearing.

A recent study of a *grhl2b* mutant has shed light on the function of this transcription factor in the developing zebrafish auditory system (Han et al. 2011). Zebrafish Grhl2b is 71% identical to human GRHL2 and broadly expressed throughout the embryo. Despite this broad expression, homozygous *grhl2b*$^{T086/T086}$ mutant zebrafish carrying a transposon-based *Tol2* gene trap in intron 1 has defects largely limited to the inner ear. In this gene-trap model, the mutant transcript would consist of the first 6 amino acids of Grhl2b fused in frame with the coding sequence for EGFP. The phenotype of developing mutant fish at 36 h postfertilization included enlarged otocysts, reduced or absent otoliths, and aberrantly formed semicircular canals. At 5 days postfertilization, mature fish exhibited hearing and balance deficits, although anatomically, hair cells appeared grossly normal (Han et al. 2011). Injection of wild-type mRNA transcribed from human *GRHL2* into mutant embryos rescued the mutant phenotype of homozygous *grhl2b*$^{T086/T086}$ fish. When the same experiment was repeated using human DFNA28 mutant *GRHL2*$^{(1609\text{-}1610insC)}$ mRNA, there was no rescue of the mutant phenotype (Han et al. 2011). These data suggest that the human *GRHL2*$^{1609\text{-}1610insC}$ is a loss-of-function allele. Given the involvement of *grh* and *Grhl3* in wound repair (Caddy et al. 2010; Mace et al. 2005), it is tempting to speculate that *Grhl2* might similarly contribute to epithelial repair and homeostasis in the cochlea.

8.7 Noncoding Mutations of *HGF* Cause Nonsyndromic Deafness DFNB39

Autosomal recessive, nonsyndromic hearing loss DFNB39 was initially mapped to a large interval on chromosome 7q11.22–q21.12 (Wajid et al. 2003). Additional families segregating nonsyndromic hearing loss refined the interval to 1.2 Mb (Schultz et al. 2009). All of the annotated and predicted genes in the smallest *DFNB39* interval were sequenced, but no missense, nonsense, or frameshift mutations were found. However, in a conserved region of hepatocyte growth factor (*HGF*) intron 4, two different overlapping microdeletions (3 and 10 bp) were found that cosegregated with deafness in these families. *HGF* encodes hepatocyte growth factor. These deletions of *HGF* turned out to be located not just in an intron but were part of the 3′ UTR of a novel short isoform of *HGF*. The functions of other reported shorter isoforms of *HGF* are poorly understood, and despite the wealth of literature available for HGF, no comprehensive study of their temporal or spatial regulation has been published.

HGF is a secreted protein that functions in a variety of organ systems as a potent mitogen, morphogen, and motogen (Birchmeier et al. 2003; Nakamura et al. 2011; Schmidt et al. 1995). HGF is also important for wound healing and regeneration, and somatic mutations affecting *HGF* expression have been implicated in carcinogenesis (Ma et al. 2009). The active form of HGF is produced through proteolytic cleavage

of the pro-HGF polypeptide to form an alpha chain (N-terminal and four kringle domains) and beta chain (serine protease-like domain) which then heterodimerize through disulfide bonds. The mature HGF protein binds the cell surface MET receptor in a 2:2 receptor-ligand complex (Kemp et al. 2006).

How mutations in the short isoform of *HGF* cause deafness is not yet known, but it is clear that the normal development of the mouse auditory system is sensitive to HGF expression levels. Transgenic mice that ubiquitously over-express full-length *Hgf* are viable but deaf (Schultz et al. 2009). Conversely, a cochlear-specific conditional knockout of *Hgf* is also deaf (Phaneuf et al. 2004; Schultz et al. 2009).Taken together, these data indicate that dysregulation (either too much or too little) of one or more isoforms of HGF can cause hearing loss (Schultz et al. 2009). How the *Hgf* transcript is normally regulated to prevent under or over production of the HGF isoforms remains unclear. One possibility is that the 3-bp and 10-bp deletions within the *Hgf* 3′ UTR associated with DFNB39 deafness remove a sequence that is complementary to a microRNA. Aside from regulation of the *Hgf* gene itself, the function of HGF protein in the cochlea is also unknown. A full evaluation of inner ear HGF isoforms and cellular signaling downstream of the MET receptor awaits investigation.

8.8 Mutations in *ESRRB* Cause DFNB35

Genetic mapping of recessively inherited hearing loss often utilizes families with consanguineous marriages or from more broadly endogamous populations. The majority of efforts to map and identify nonsyndromic deafness genes have utilized such families and taken advantage of genome-wide screens for marker homozygosity (Friedman et al. 1995). Using this method, severe to profound nonsyndromic deafness segregating as an autosomal recessive disorder in a single large Pakistani family was linked (multi-point LOD of 7.6) using homozygosity mapping to markers on chromosome 14q and the locus designated *DFNB35* (Ansar et al. 2003). Based on only one family, a linkage interval of 10 Mb of DNA (11.8 cM) on chromosome 14q was reported. The meiotic boundaries of other deafness loci on chromosome 14 (*DFNA9*, *DFNA23* and *DFNB5*) did not overlap with *DFNB35*. As additional consanguineous families with deafness were linked to genetic markers on chromosome 14q, the *DFNB35* interval was further refined to about 1 Mb of genomic DNA. This interval encompassed seven genes, one of which was the orphan estrogen-related receptor (*ESRRB*), a member of the nuclear hormone receptor (NHR) family of transcription factors (reviewed by Blumberg and Evans 1998). Subsequently, missense mutations and a frameshift allele of *ESRRB* were reported in six unrelated families, including the original family used to map *DFNB35* (Ben Said et al. 2011; Collin et al. 2008b). The hearing loss phenotype segregating in all the DFNB35 families was very similar. The deaf individuals in a family of Turkish origin did not have any obvious visual or renal problems, ruling out Usher and Alport syndromes, respectively, and one affected male in this family was fertile (Collin et al. 2008b). No clinically relevant features cosegregated with deafness in any of the other

affected members within the pedigree (Collin et al. 2008b). However, it is worth noting that in general, it is always difficult to exclude more subtle phenotypes, especially if the affected individuals do not receive a thorough evaluation by physicians. This concern applies equally to the many other families segregating presumptive nonsyndromic deafness that have been reported over the past 15 years.

ESRRB has at least three alternative splice isoforms, two of which are widely expressed throughout the embryo. The mutations in *ESRRB* that associate with deafness are all missense mutations and predicted to affect all three isoforms (Ben Said et al. 2011; Collin et al. 2008b). The longest isoform of *ESRRB* is expressed in testes and in cells of mesothelial origin, as well as the supporting cells and stria vascularis of the developing and adult inner ear (Collin et al. 2008b; Zhou et al. 2006). Expression was not detected in the sensory hair cells within the organ of Corti. Previous animal studies had demonstrated that mouse embryos homozygous for a null *Esrrb*$^{-/-}$ allele die at E10.5, necessitating the use of a conditional allele to study cochlear function (Luo et al. 1997; Mitsunaga et al. 2004). A conditional knockout of *Esrrb* recapitulated DFNB35 deafness and revealed that *Esrrb* is required for correct development of marginal cells in the stria vascularis (Chen and Nathans 2007).

Some of the transcriptional targets of ESRRB within the inner ear have already been identified. Chen and Nathans used microarray hybridization to compare gene expression in stria vascularis isolated from either wild-type or conditionally null *Esrrb*$^{-/-}$ cochleae (Chen and Nathans 2007). Of the changes confirmed by northern analyses, potassium voltage-gated channel, Isk-related family, member 1 (*Kcne1*), aldehyde dehydrogenase 1 family, member A2 (*Aldh1a2*), R-spondin 3 (*Rspo3*), prostaglandin D2 synthase 21 kDa (*Ptgds*), ATPase, Na+/K+ transporting, beta 2 polypeptide (*Atp1b2*), solute carrier family 12 member 2 (*Slc12a2*), potassium voltage-gated channel, KQT-like subfamily, member 1 (*Kcnq1*), and WNK lysine-deficient protein kinase 4 (*Wnk4*) were all significantly repressed in the absence of ESRRB (Chen and Nathans 2007). Another study has hypothesized that ESRRB might modulate the effects of the thyroid hormone pathway within the cochlea (Collin et al. 2008b). Thyroid hormone and the two thyroid receptors THRA and THRB are essential for hearing and regulate the expression of potassium voltage-gated channel, KQT-like subfamily, member 4 (*Kcnq4*) in outer hair cells and potassium inwardly rectifying channel, subfamily J, member 10 (*Kcnj10*) in stria vascularis (Forrest et al. 2002; Mustapha et al. 2009; Rusch et al. 2001; Winter et al. 2007). Future experiments using chromatin immunoprecipitation (ChIP) will be important to determine the direct transcriptional targets of ESRRB and dissect the molecular mechanisms underlying DFNB35 deafness.

Outside the cochlea, an unexpected function for *ESRRB* was recently reported by stem cell biologists attempting to differentiate fibroblasts into pluripotent stem cells (iPS). Mouse embryonic fibroblasts (MEFs) transfected with cDNAs encoding *Oct4*, *Sox2*, *c-Myc*, and *Klf4* are converted to iPS cells. Feng and coauthors reported that *Esrrb* in combination with *Oct4*, c-Myc, and *Sox2* could also reprogram MEFs to iPS cells (Feng et al. 2009; Heng et al. 2010). The significance of this in DFNB35 deafness is yet to be explored.

8.9 Concluding Remarks

Our understanding of the gene regulatory events underlying hearing loss is far from complete, providing investigators with exciting opportunities for discovery. Though many protein-coding genes have now been implicated in the biology of hearing, the identification of regulatory sequences has lagged behind. At least one reason for this disparity originates from how candidate loci have been evaluated. Since chromosomal intervals identified from linkage studies are generally large (>5 Mb), the subsequent search for pathogenic variants has typically targeted exonic coding regions of genes. This approach has clearly been effective, but is by its very nature biased against the discovery of causative variants in promoters, long-range enhancers, and repressors that can be a considerable distance from their respective targets, in addition to a myriad of other regulatory elements. It is interesting to ponder how many laboratories around the world have ascertained pedigrees segregating deafness but have hitherto been unable to find convincing exonic mutations. Massively parallel sequencing of the entire genome or of linked regions selected by targeted enrichment promises to address this bias, as well as to expedite the identification of pathogenic mutations in general (Rehman et al. 2010). Caution must be exercised with global exome capture methodologies as these would still be predicted to miss regulatory mutations in the vicinity of an exon.

The mouse Twirler (*Tw*) allele is a good example where a noncoding variant with a major effect would have been overlooked if only the exome has been sequenced for mutations. The dominant Twirler phenotype is characterized by obesity and malformation of the vestibular semicircular canals that result in circling behavior. The only DNA variant in the Twirler chromosome 18 linkage interval was a nucleotide change (c.58+181G>A) in a predicted MYB consensus binding site located in first intron of *Zeb1*, a transcription factor involved in mesenchymal cell fate (Hertzano et al. 2011; Kurima et al. 2011). Demonstrating causality between a complex phenotype and pathogenicity of a noncoding single base change is challenging in any organism since any rare variant may be in linkage disequilibrium with the genuine pathogenic allele. In an otherwise wild-type mouse, Kurima and colleagues experimentally introduced the c.58+181G>A into intron 1 of a wild-type *Zeb1* gene. Heterozygous mice for this engineered point mutation recapitulated the Twirler phenotype providing definitive evidence of causality of a noncoding nucleotide change, which was also shown by EMSA analysis to disrupt MYB binding (Kurima et al. 2011).

Analysis of massively parallel datasets is certain to bring new challenges, not least in the correct ascertainment of pathogenicity among a large number of nonpathogenic variants. Recent whole-genome studies have identified nonsense mutations in healthy individuals (Li et al. 2010b; MacArthur and Tyler-Smith 2010), highlighting the need for candidate mutant alleles to be confirmed with both rigorous functional analyses and large pedigrees demonstrating statistical linkage to the phenotype (Rehman et al. 2010). Since such human pedigrees are not always available, analysis of several unrelated sporadic cases with a variety of different pathogenic

alleles also raises confidence that the variant has been correctly identified (Lindhurst et al. 2011). The bar for validating novel human pathogenic alleles is set deliberately high to avoid false positives polluting the literature and also potentially misinforming genetic counselors as well as affected subjects.

Animal models will continue to be critical for dissecting gene regulation in the auditory system. The NIH Knockout Mouse Project (KOMP) and the European Conditional Mouse Mutagenesis Project (EUCOMM) alone will bolster our knowledge of the transcription factors involved in hearing and deafness (Skarnes et al. 2011). These consortia were tasked with generating either targeted null or gene-trapped alleles for every annotated gene in the mouse genome, ultimately making them freely available in public repositories. As more of these mice are generated and subjected to rigorous phenotyping, including auditory and vestibular testing (Hardisty-Hughes et al. 2010), new genes and transcription factors will undoubtedly be linked to sensory function. A significant limitation of these high-throughput efforts is that null alleles will not always be viable and may not accurately model gain-of-function alleles; a case in point is *Grhl2*, where neither a heterozygous nor homozygous null allele correctly recapitulates DFNA28 deafness. These mutant resources are also unlikely to uncover regulatory elements that are not in close proximity to the protein-coding exons. Thus, there is still a need to generate mouse models with humanized mutations in order to fully understand gene regulation in the auditory system.

In the absence of clear evidence from human genetics, how can we go about identifying regulatory elements that are important for auditory function? This is a daunting task, especially considering that some elements can be megabases away from the regulated gene itself. One approach to mapping these types of sequences is the introduction of mobile elements, such as *Sleeping Beauty* or *PiggyBac*, that can transpose multiple times to disrupt existing elements or introduce new ones (Ding et al. 2005; Kokubu et al. 2009; Rad et al. 2010). Transposon-based mutagenesis can target large loci (~1 kb–100 kb) and allows the effects of any specific integration to be subsequently examined in a live animal. Another powerful approach for identifying potential regulatory sequences is chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-Seq; Robertson et al. 2007). By immunoprecipitating known enhancer-/repressor-associated proteins or transcription factors cross-linked to chromatin, their binding sites across the genome can be extensively mapped. For example, ChIP-Seq has been successfully used to map a gamut of tissue-specific enhancers that bind the enhancer-associated protein p300 throughout the body (Visel et al. 2009). While many of these enhancers are strongly evolutionarily conserved, some are less so, highlighting the unbiased approach of ChIP-Seq to identify enhancers versus comparative genomics (Blow et al. 2010). With few exceptions, the activity of enhancers and repressors responsible for cochlear and hair cell-specific expression are largely unstudied and constitutes an important area for future research.

Ultimately, understanding cochlear gene regulation in its entirety will require multimodal genomic, transcriptomic, and proteomic datasets to be assembled. Though these types of datasets are technically challenging to generate, they promise

to reveal subtle modes of gene regulation. In support of this, the first study of this type in the auditory system was recently published. Elkan-Miller and coauthors combined expression arrays and quantitative proteomics to compare gene regulation by microRNAs between auditory and vestibular organs. By comparing transcriptional and translational changes in parallel, not only can the noise inherent to microRNA target prediction be overcome, but varying modes of microRNA regulation can also be potentially detected (Elkan-Miller et al. 2011). It will be important to expand upon these types of studies in the cochlea. Multimodal cochlear datasets will also likely reveal rare changes between the genomic DNA template and the transcribed RNA molecules, so-called RNA-DNA differences. In addition to increasing the potential for regulatory diversity, these changes are of potential significance for identifying novel deafness alleles. Is it possible that a mutant allele could be manifested solely in mRNA, such that its genomic DNA sequence was essentially wild type? The involvement of a regulatory mechanism like this in deafness would be unprecedented but is certainly possible.

The structural complexity of the cochlea presents one final confound to understanding gene regulation in vivo. The cochlea typically contains only a few tens of thousands of sensory hair cells, and ancillary cell types outnumber these by at least an order of magnitude. Investigators are left to decipher which signals are specific to hair cells among a high level of contamination. Fluorescence-activated cell sorting (FACS) is one method that has been used to purify hair cell and supporting cell populations from the mouse cochlea, using transgenic mice expressing EGFP under the atonal homolog 1 (*Atoh1* or *Math1*) or $p27^{kip1}$ promoter, respectively (Doetzlhofer et al. 2006; White et al. 2006). The identification of cell surface markers for the different cell types in the inner ear will also allow for antibody-based FACS sorting of this cell population (Hertzano et al. 2010, 2011). A completely different way to isolate potentially pure populations of sensory hair cells was recently published in a seminal study by Oshima and colleagues (2010). They demonstrated that with the use of appropriate transcription factors, mouse embryonic stem cells could be made to differentiate into immature yet functional hair cells in vitro. This exciting advance allows for a potentially unlimited source of hair cells to be grown in vitro, a goal that has doggedly eluded investigators for many years.

In conclusion, the study of human hereditary deafness has contributed enormously to our understanding of the cochlea and the genes involved in its function. As investigators delve deeper into the transcriptome and regulatory landscape of hair cells, it promises to open a new window on how they function normally, how they respond to noise and drug trauma, and how they change during the natural aging process.

Acknowledgements We thank Lisa Cunningham, Dennis Drayna, Ronna Hertzano, Robert Morell and Julie Schultz for helpful discussions. JEB and TBF were supported by the National Institute on Deafness and Other Communication Disorders intramural research funds DC000039-15 at the National Institutes of Health.

Abbreviations

3′UTR	3′ untranslated region
ABR	Auditory brainstem response
ALDH1A2	Aldehyde dehydrogenase 1 family member A2
ATP1B2	ATPase Na+/K + transporting, beta 2 polypeptide
bp	Base pair
CDH23	Cadherin-related 23
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation coupled with massively parallel sequencing
DFNA	Nonsyndromic deafness autosomal dominant
DFNB	Nonsyndromic deafness autosomal recessive
DFNX	Nonsyndromic deafness X-linked
DPOAE	Distortion product otoacoustic emissions
E17	Embryonic day 17
ESRRB	Estrogen-related receptor
EUCOMM	European conditional mouse mutagenesis project
EYA4	Eyes absent Drosophila homolog 4
FACS	Fluorescence-activated cell sorting
FBXO11	F-box protein 11
GFP	Green fluorescent protein
GRHL2	Grainyhead-like 2
HGF	Hepatocyte growth factor
HNCR	Highly conserved noncoding regions
Hz	Hertz
iPS	Induced pluripotent stems cells
Kb	Kilobase
KCNE1	Potassium voltage-gated channel, Isk-related family, member 1
KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1
KOMP	NIH knockout mouse project
Mb	Megabase
MEFs	Mouse embryonic fibroblasts
NMD	Nonsense-mediated mRNA decay
POU3F4	POU class 3 homeobox 4
POU4F3	POU class 4 homeobox 3
POU_H	POU homeodomain
PTGDS	Prostaglandin D2 synthase 21 kDa
PTPRQ	Protein tyrosine phosphatase receptor type, Q
RISC	RNA-induced silencing complex
RSPO3	R-spondin 3
SIX1	SIX homeobox 1
SLC12A2	Solute carrier family 12 member 2
SLC26A5	Solute carrier family 26 member 5

SNP	Single nucleotide polymorphism
STR	Short tandem repeats
WHO	World Health Organization
WNK4	Lysine-deficient protein kinase 4
WS2A	Waardenburg syndrome type 2A

References

Abe Y, Oka A, Mizuguchi M, Igarashi T, Ishikawa S, Aburatani H, Yokoyama S, Asahara H, Nagao K, Yamada M, Miyashita T (2009) EYA4, deleted in a case with middle interhemispheric variant of holoprosencephaly, interacts with SIX3 both physically and functionally. Hum Mutat 30(10):E946–E955

Ahn KJ, Passero F Jr, Crenshaw EB III (2009) Otic mesenchyme expression of Cre recombinase directed by the inner ear enhancer of the *Brn4/Pou3f4* gene. Genesis 47(3):137–141

Alvarado DM, Hawkins RD, Bashiardes S, Veile RA, Ku YC, Powder KE, Spriggs MK, Speck JD, Warchol ME, Lovett M (2011) An RNA interference-based screen of transcription factor genes identifies pathways necessary for sensory regeneration in the avian inner ear. J Neurosci 31(12):4535–4543

Ansar M, Din MA, Arshad M, Sohail M, Faiyaz-Ul-Haque M, Haque S, Ahmad W, Leal SM (2003) A novel autosomal recessive non-syndromic deafness locus (*DFNB35*) maps to 14q24.1–14q24.3 in large consanguineous kindred from Pakistan. Eur J Hum Genet 11(1):77–80

Avraham KB (2000) *DFNA15*. Adv Otorhinolaryngol 56:107–115

Ben Said M, Ayedi L, Mnejja M, Hakim B, Khalfallah A, Charfeddine I, Khifagi C, Turki K, Ayadi H, Benzina Z, Ghorbel A, Castillo ID, Masmoudi S, Aifa MH (2011) A novel missense mutation in the *ESRRB* gene causes *DFNB35* hearing loss in a Tunisian family. Eur J Med Genet 54(6):e535–e541

Biggin MD, Tjian R (1988) Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. Cell 53(5):699–711

Birchmeier C, Birchmeier W, Gherardi E, Vande Woude GF (2003) Met, metastasis, motility and more. Nat Rev Mol Cell Biol 4(12):915–925

Bitner-Glindzicz M, Turnpenny P, Hoglund P, Kaariainen H, Sankila EM, van der Maarel SM, de Kok YJ, Ropers HH, Cremers FP, Pembrey M et al (1995) Further mutations in Brain 4 (POU3F4) clarify the phenotype in the X-linked deafness, *DFN3*. Hum Mol Genet 4(8):1467–1469

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Bristow J, Ren B, Black BL, Rubin EM, Visel A, Pennacchio LA (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42(9):806–810

Blumberg B, Evans RM (1998) Orphan nuclear receptors – new ligands and new possibilities. Genes Dev 12(20):3149–3155

Boglev Y, Wilanowski T, Caddy J, Parekh V, Auden A, Darido C, Hislop NR, Cangkrama M, Ting SB, Jane SM (2011) The unique and cooperative roles of the grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. Dev Biol 349(2):512–522

Bonfig W, Krude H, Schmidt H (2011) A novel mutation of LHX3 is associated with combined pituitary hormone deficiency including ACTH deficiency, sensorineural hearing loss, and short neck-a case report and review of the literature. Eur J Pediatr 170(8):1017–1021

Bonini NM, Leiserson WM, Benzer S (1998) Multiple roles of the eyes absent gene in Drosophila. Dev Biol 196(1):42–57

Borsani G, DeGrandi A, Ballabio A, Bulfone A, Bernard L, Banfi S, Gattuso C, Mariani M, Dixon M, Donnai D, Metcalfe K, Winter R, Robertson M, Axton R, Brown A, van Heyningen V,

Hanson I (1999) *EYA4*, a novel vertebrate gene related to Drosophila eyes absent. Hum Mol Genet 8(1):11–23

Bray SJ, Kafatos FC (1991) Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila. Genes Dev 5(9):1672–1683

Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10(2):141–148

Brunner HG, van Bennekom A, Lambermon EM, Oei TL, Cremers WR, Wieringa B, Ropers HH (1988) The gene for X-linked progressive mixed deafness with perilymphatic gusher during stapes surgery (DFN3) is linked to PGK. Hum Genet 80(4):337–340

Budhram-Mahadeo V, Morris PJ, Lakin ND, Dawson SJ, Latchman DS (1996) The different activities of the two activation domains of the Brn-3a transcription factor are dependent on the context of the binding site. J Biol Chem 271(15):9108–9113

Caddy J, Wilanowski T, Darido C, Dworkin S, Ting SB, Zhao Q, Rank G, Auden A, Srivastava S, Papenfuss TA, Murdoch JN, Humbert PO, Parekh V, Boulos N, Weber T, Zuo J, Cunningham JM, Jane SM (2010) Epidermal wound repair is regulated by the planar cell polarity signaling pathway. Dev Cell 19(1):138–147

Chatterjee S, Kraus P, Lufkin T (2010) A symphony of inner ear developmental control genes. BMC Genet 11:68

Chen J, Nathans J (2007) Estrogen-related receptor beta/NR3B2 controls epithelial cell fate and endolymph production by the stria vascularis. Dev Cell 13(3):325–337

Christensen KL, Patrick AN, McCoy EL, Ford HL (2008) The six family of homeobox genes in development and cancer. Adv Cancer Res 101:93–126

Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D (2010) Ancient animal microRNAs and the evolution of tissue identity. Nature 463(7284):1084–1088

Collin RW, Chellappa R, Pauw RJ, Vriend G, Oostrik J, van Drunen W, Huygen PL, Admiraal R, Hoefsloot LH, Cremers FP, Xiang M, Cremers CW, Kremer H (2008a) Missense mutations in POU4F3 cause autosomal dominant hearing impairment DFNA15 and affect subcellular localization and DNA binding. Hum Mutat 29(4):545–554

Collin RW, Kalay E, Tariq M, Peters T, van der Zwaag B, Venselaar H, Oostrik J, Lee K, Ahmed ZM, Caylan R, Li Y, Spierenburg HA, Eyupoglu E, Heister A, Riazuddin S, Bahat E, Ansar M, Arslan S, Wollnik B, Brunner HG, Cremers CW, Karaguzel A, Ahmad W, Cremers FP, Vriend G, Friedman TB, Leal SM, Kremer H (2008b) Mutations of *ESRRB* encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. Am J Hum Genet 82(1):125–138

Cotanche DA, Kaiser CL (2010) Hair cell fate decisions in cochlear development and regeneration. Hear Res 266(1–2):18–25

Cremers FP, Cremers CW, Ropers HH (2000) The ins and outs of X-linked deafness type 3. Adv Otorhinolaryngol 56:184–195

Dawson SJ, Morris PJ, Latchman DS (1996) A single amino acid change converts an inhibitory transcription factor into an activator. J Biol Chem 271(20):11631–11633

de Kok YJ, Merkx GF, van der Maarel SM, Huber I, Malcolm S, Ropers HH, Cremers FP (1995a) A duplication/paracentric inversion associated with familial X-linked deafness (*DFN3*) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene. Hum Mol Genet 4(11):2145–2150

de Kok YJ, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers HH, Cremers FP (1995b) Association between X-linked mixed deafness and mutations in the POU domain gene *POU3F4*. Science 267(5198):685–688

de Kok YJ, Vossenaar ER, Cremers CW, Dahl N, Laporte J, Hu LJ, Lacombe D, Fischel-Ghodsian N, Friedman RA, Parnes LS, Thorpe P, Bitner-Glindzicz M, Pander HJ, Heilbronner H, Graveline J, den Dunnen JT, Brunner HG, Ropers HH, Cremers FP (1996) Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (*DFN3*) 900 kb proximal to the *DFN3* gene *POU3F4*. Hum Mol Genet 5(9):1229–1235

De Leenheer EM, Huygen PL, Wayne S, Smith RJ, Cremers CW (2001) The DFNA10 phenotype. Ann Otol Rhinol Laryngol 110(9):861–866

De Leenheer EM, Huygen PL, Wayne S, Verstreken M, Declau F, Van Camp G, Van de Heyning PH, Smith RJ, Cremers CW (2002) DFNA10/EYA4–the clinical picture. Adv Otorhinolaryngol 61:73–78

de Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, Van Haeringen A, Genevieve D, Goldenberg A, Oufadem M, Manouvrier S, Munnich A, Vidigal JA, Vekemans M, Lyonnet S, Henrion-Caude A, Ventura A, Amiel J (2011) Germline deletion of the mIR-17~92 cluster causes skeletal and growth defects in humans. Nat Genet 43(10):1026–1030

Depreux FF, Darrow K, Conner DA, Eavey RD, Liberman MC, Seidman CE, Seidman JG (2008) Eya4-deficient mice are a model for heritable otitis media. J Clin Invest 118(2):651–658

Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122(3):473–483

Doetzlhofer A, White P, Lee YS, Groves A, Segil N (2006) Prospective identification and purification of hair cell and supporting cell progenitors from the embryonic cochlea. Brain Res 1091(1):282–288

Elkan-Miller T, Ulitsky I, Hertzano R, Rudnicki A, Dror AA, Lenz DR, Elkon R, Irmler M, Beckers J, Shamir R, Avraham KB (2011) Integration of transcriptomics, proteomics, and microRNA analyses reveals novel microRNA regulation of targets in the mammalian inner ear. PLoS One 6(4):e18195

Erkman L, McEvilly RJ, Luo L, Ryan AK, Hooshmand F, O'Connell SM, Keithley EM, Rapaport DH, Ryan AF, Rosenfeld MG (1996) Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development. Nature 381(6583):603–606

Feingold M, Hall BD, Lacassie Y, Martinez-Frias ML (1997) Syndrome of microcephaly, facial and hand abnormalities, tracheoesophageal fistula, duodenal atresia, and developmental delay. Am J Med Genet 69(3):245–249

Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng HH (2009) Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol 11(2):197–203

Fettiplace R, Hackney CM (2006) The sensory and motor roles of auditory hair cells. Nat Rev Neurosci 7(1):19–29

Forrest D, Reh TA, Rusch A (2002) Neurodevelopmental control by thyroid hormone receptors. Curr Opin Neurobiol 12(1):49–56

Friedman LM, Avraham KB (2009) MicroRNAs and epigenetic regulation in the mammalian inner ear: implications for deafness. Mamm Genome 20(9–10):581–603

Friedman LM, Dror AA, Mor E, Tenne T, Toren G, Satoh T, Biesemeier DJ, Shomron N, Fekete DM, Hornstein E, Avraham KB (2009) MicroRNAs are essential for development and function of inner ear hair cells in vertebrates. Proc Natl Acad Sci USA 106(19):7915–7920

Friedman T, Battey J, Kachar B, Riazuddin S, Noben-Trauth K, Griffith A, Wilcox E (2000) Modifier genes of hereditary hearing loss. Curr Opin Neurobiol 10(4):487–493

Friedman TB, Liang Y, Weber JL, Hinnant JT, Barber TD, Winata S, Arhya IN, Asher JH Jr (1995) A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. Nat Genet 9(1):86–91

Friedman TB, Schultz JM, Ahmed ZM, Tsilou ET, Brewer CC (2011) Usher syndrome: hearing loss with vision loss. Adv Otorhinolaryngol 70:56–65

Frolenkov GI, Belyantseva IA, Friedman TB, Griffith AJ (2004) Genetic insights into the morphogenesis of inner ear hair cells. Nat Rev Genet 5(7):489–498

Goodyear RJ, Richardson GP (2003) A novel antigen sensitive to calcium chelation that is associated with the tip links and kinocilial links of sensory hair bundles. J Neurosci 23(12):4878–4887

Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466(7308):835–840

Han Y, Mu Y, Li X, Xu P, Tong J, Liu Z, Ma T, Zeng G, Yang S, Du J, Meng A (2011) Grhl2 deficiency impairs otic development and hearing ability in a zebrafish model of the progressive dominant hearing loss DFNA28. Hum Mol Genet 20(16):3213–3226

Hanson IM (2001) Mammalian homologues of the Drosophila eye specification genes. Semin Cell Dev Biol 12(6):475–484

Hardisty-Hughes RE, Parker A, Brown SD (2010) A hearing and vestibular phenotyping pipeline to identify mouse mutants with hearing impairment. Nat Protoc 5(1):177–190

Hawkins RD, Bashiardes S, Powder KE, Sajan SA, Bhonagiri V, Alvarado DM, Speck J, Warchol ME, Lovett M (2007) Large scale gene expression profiles of regenerating inner ear sensory epithelia. PLoS One 2(6):e525

Hawkins RD, Helms CA, Winston JB, Warchol ME, Lovett M (2006) Applying genomics to the avian inner ear: development of subtractive cDNA resources for exploring sensory function and hair cell regeneration. Genomics 87(6):801–808

Henderson DH, Bielefeld E (2008) Patterns and mechanisms of noise-induced cochlear pathology. In: Schacht J, Fay R (eds) Auditory trauma, protection and repair, vol 31. Springer, Dordrecht

Heng JC, Orlov YL, Ng HH (2010) Transcription factors for the modulation of pluripotency and reprogramming. Cold Spring Harb Symp Quant Biol 75:37–244

Hertzano R, Dror AA, Montcouquiol M, Ahmed ZM, Ellsworth B, Camper S, Friedman TB, Kelley MW, Avraham KB (2007) Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system. Eur J Neurosci 25(4):999–1005

Hertzano R, Elkon R, Kurima K, Morrisson A, Chan SL, Sallin M, Biedlingmaier A, Darling DS, Griffith AJ, Eisenman DJ, Strome SE (2011) Cell type-specific transcriptome analysis reveals a major role for Zeb1 and miR-200b in mouse inner ear morphogenesis. PLoS Genet 7(9):e1002309

Hertzano R, Montcouquiol M, Rashi-Elkeles S, Elkon R, Yucel R, Frankel WN, Rechavi G, Moroy T, Friedman TB, Kelley MW, Avraham KB (2004) Transcription profiling of inner ears from Pou4f3(ddl/ddl) identifies Gfi1 as a target of the *Pou4f3* deafness gene. Hum Mol Genet 13(18):2143–2153

Hertzano R, Puligilla C, Chan SL, Timothy C, Depireux DA, Ahmed Z, Wolf J, Eisenman DJ, Friedman TB, Riazuddin S, Kelley MW, Strome SE (2010) CD44 is a marker for the outer pillar cells in the early postnatal mouse inner ear. J Assoc Res Otolaryngol 11(3):407–418

Jemc J, Rebay I (2007) The eyes absent family of phosphotyrosine phosphatases: properties and roles in developmental regulation of transcription. Annu Rev Biochem 76:513–538

Jero J, Coling DE, Lalwani AK (2001) The use of Preyer's reflex in evaluation of hearing in mice. Acta Otolaryngol 121(5):585–589

Kelley MW, Wu DK (2005) Development of the inner ear. Springer, New York

Kemp LE, Mulloy B, Gherardi E (2006) Signalling by HGF/SF and Met: the role of heparan sulphate co-receptors. Biochem Soc Trans 34(Pt 3):414–417

Kleinjan DJ, van Heyningen V (1998) Position effect in human genetic disease. Hum Mol Genet 7(10):1611–1618

Kokubu C, Horie K, Abe K, Ikeda R, Mizuno S, Uno Y, Ogiwara S, Ohtsuka M, Isotani A, Okabe M, Imai K, Takeda J (2009) A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. Nat Genet 41(8):946–952

Kuhn S, Johnson SL, Furness DN, Chen J, Ingham N, Hilton JM, Steffes G, Lewis MA, Zampini V, Hackney CM, Masetto S, Holley MC, Steel KP, Marcotti W (2011) miR-96 regulates the progression of differentiation in mammalian cochlear inner and outer hair cells. Proc Natl Acad Sci USA 108(6):2355–2360

Kurima K, Hertzano R, Gavrilova O, Monahan K, Shpargel KB, Nadaraja G, Kawashima Y, Lee KY, Ito T, Higashi Y, Eisenman DJ, Strome SE, Griffith AJ (2011) A noncoding point mutation of zeb1 causes multiple developmental malformations and obesity in twirler mice. PLoS Genet 7(9):e1002307

Lee HK, Park HJ, Lee KY, Park R, Kim UK (2010) A novel frameshift mutation of *POU4F3* gene associated with autosomal dominant non-syndromic hearing loss. Biochem Biophys Res Commun 396(3):626–630

Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, Langford C, van Dongen S, Abreu-Goodger C, Piipari M, Redshaw N, Dalmay T, Moreno-Pelayo MA, Enright AJ, Steel KP (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. Nat Genet 41(5):614–618

Li H, Fekete DM (2010) MicroRNAs in hair cell development and deafness. Curr Opin Otolaryngol Head Neck Surg 18(5):459–465

Li X, Oghi KA, Zhang J, Krones A, Bush KT, Glass CK, Nigam SK, Aggarwal AK, Maas R, Rose DW, Rosenfeld MG (2003) Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. Nature 426(6964):247–254

Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparso T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jorgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang X, Li R, Yang H, Wang J, Hansen T, Pedersen O, Nielsen R (2010) Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet 42(11):969–972

Lin CC, Liu LZ, Addison JB, Wonderlin WF, Ivanov AV, Ruppert JM (2011a) A KLF4-miRNA-206 autoregulatory feedback loop can promote or inhibit protein translation depending upon cell context. Mol Cell Biol 31(12):2513–2527

Lin YH, Wu CC, Hsu CJ, Hwang JH, Liu TC (2011b) The grainyhead-like 2 gene (*GRHL2*) single nucleotide polymorphism is not associated with age-related hearing impairment in Han Chinese. Laryngoscope 121(6):1303–1307

Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, Turner J, Cannons JL, Bick D, Blakemore L, Blumhorst C, Brockmann K, Calder P, Cherman N, Deardorff MA, Everman DB, Golas G, Greenstein RM, Kato BM, Keppler-Noreuil KM, Kuznetsov SA, Miyamoto RT, Newman K, Ng D, O'Brien K, Rothenberg S, Schwartzentruber DJ, Singhal V, Tirabosco R, Upton J, Wientroub S, Zackai EH, Hoag K, Whitewood-Neal T, Robey PG, Schwartzberg PL, Darling TN, Tosi LL, Mullikin JC, Biesecker LG (2011) A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med 365(7):611–619

Luo J, Sladek R, Bader JA, Matthyssen A, Rossant J, Giguere V (1997) Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. Nature 388(6644):778–782

Ma J, DeFrances MC, Zou C, Johnson C, Ferrell R, Zarnegar R (2009) Somatic mutation and functional polymorphism of a novel regulatory element in the HGF gene promoter causes its aberrant expression in human breast cancer. J Clin Invest 119(3):478–491

MacArthur DG, Tyler-Smith C (2010) Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet 19(R2):R125–R130

Mace KA, Pearson JC, McGinnis W (2005) An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head. Science 308(5720):381–385

Makishima T, Madeo AC, Brewer CC, Zalewski CK, Butman JA, Sachdev V, Arai AE, Holbrook BM, Rosing DR, Griffith AJ (2007) Nonsyndromic hearing loss DFNA10 and a novel mutation of EYA4: evidence for correlation of normal cardiac phenotype with truncating mutations of the Eya domain. Am J Med Genet A 143A(14):1592–1598

Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, Moreno F, Moreno-Pelayo MA (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 41(5):609–613

Minowa O, Ikeda K, Sugitani Y, Oshima T, Nakai S, Katori Y, Suzuki M, Furukawa M, Kawase T, Zheng Y, Ogura M, Asada Y, Watanabe K, Yamanaka H, Gotoh S, Nishi-Takeshima M, Sugimoto T, Kikuchi T, Takasaka T, Noda T (1999) Altered cochlear fibrocytes in a mouse model of DFN3 nonsyndromic deafness. Science 285(5432):1408–1411

Mitsunaga K, Araki K, Mizusaki H, Morohashi K, Haruna K, Nakagata N, Giguere V, Yamamura K, Abe K (2004) Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos. Mech Dev 121(3):237–246

Modamio-Hoybjor S, Moreno-Pelayo MA, Mencia A, del Castillo I, Chardenoux S, Morais D, Lathrop M, Petit C, Moreno F (2004) A novel locus for autosomal dominant nonsyndromic hearing loss, DFNA50, maps to chromosome 7q32 between the *DFNB17* and *DFNB13* deafness loci. J Med Genet 41(2):e14

Mustapha M, Fang Q, Gong TW, Dolan DF, Raphael Y, Camper SA, Duncan RK (2009) Deafness and permanently reduced potassium channel gene expression and function in hypothyroid Pit1dw mutants. J Neurosci 29(4):1212–1223

Nakamura T, Sakai K, Matsumoto K (2011) Hepatocyte growth factor twenty years on: much more than a growth factor. J Gastroenterol Hepatol 26(Suppl 1):188–202

Naranjo S, Voesenek K, de la Calle-Mustienes E, Robert-Moreno A, Kokotas H, Grigoriadou M, Economides J, Van Camp G, Hilgert N, Moreno F, Alsina B, Petersen MB, Kremer H, Gomez-Skarmeta JL (2010) Multiple enhancers located in a 1-Mb region upstream of POU3F4 promote expression during inner ear development and may be required for hearing. Hum Genet 128(4):411–419

Nicolson T, Rusch A, Friedrich RW, Granato M, Ruppersberg JP, Nusslein-Volhard C (1998) Genetic analysis of vertebrate sensory hair cell mechanosensation: the zebrafish circler mutants. Neuron 20(2):271–283

Niu X, Canlon B (2002) Protective mechanisms of sound conditioning. Adv Otorhinolaryngol 59:96–105

Noben-Trauth K, Zheng QY, Johnson KR (2003) Association of cadherin 23 with polygenic inheritance and genetic modification of sensorineural hearing loss. Nat Genet 35(1):21–23

Noben-Trauth K, Zheng QY, Johnson KR, Nishina PM (1997) mdfw: a deafness susceptibility locus that interacts with deaf waddler (dfw). Genomics 44(3):266–272

O'Neill ME, Marietta J, Nishimura D, Wayne S, Van Camp G, Van Laer L, Negrini C, Wilcox ER, Chen A, Fukushima K, Ni L, Sheffield VC, Smith RJ (1996) A gene for autosomal dominant late-onset progressive non-syndromic hearing loss, DFNA10, maps to chromosome 6. Hum Mol Genet 5(6):853–856

Ohlemiller KK, Frisna RD (2008) Age-related hearing loss and its cellular and molecular bases. In: Schacht J, Faty R (eds) Auditory trauma, protection, and repair, vol 31. Springer, Dordrecht

Ohto H, Kamada S, Tago K, Tominaga SI, Ozaki H, Sato S, Kawakami K (1999) Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya. Mol Cell Biol 19(10):6815–6824

Okabe Y, Sano T, Nagata S (2009) Regulation of the innate immune response by threonine-phosphatase of Eyes absent. Nature 460(7254):520–524

Orom UA, Nielsen FC, Lund AH (2008) MicroRNA-10a binds the 5′UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell 30(4):460–471

Oshima K, Shin K, Diensthuber M, Peng AW, Ricci AJ, Heller S (2010) Mechanosensitive hair cell-like cells from embryonic and induced pluripotent stem cells. Cell 141(4):704–716

Ostrowski S, Dierick HA, Bejsovec A (2002) Genetic control of cuticle formation during embryonic development of Drosophila melanogaster. Genetics 161(1):171–182

Owens KN, Santos F, Roberts B, Linbo T, Coffin AB, Knisely AJ, Simon JA, Rubel EW, Raible DW (2008) Identification of genetic and chemical modulators of zebrafish mechanosensory hair cell death. PLoS Genet 4(2):e1000020

Peters LM, Anderson DW, Griffith AJ, Grundfast KM, San Agustin TB, Madeo AC, Friedman TB, Morell RJ (2002) Mutation of a transcription factor, TFCP2L3, causes progressive autosomal dominant hearing loss, DFNA28. Hum Mol Genet 11(23):2877–2885

Phaneuf D, Moscioni AD, LeClair C, Raper SE, Wilson JM (2004) Generation of a mouse expressing a conditional knockout of the hepatocyte growth factor gene: demonstration of impaired liver regeneration. DNA Cell Biol 23(9):592–603

Phillips K, Luisi B (2000) The virtuoso of versatility: POU proteins that flex to fit. J Mol Biol 302(5):1023–1039

Phippard D, Boyd Y, Reed V, Fisher G, Masson WK, Evans EP, Saunders JC, Crenshaw EB 3rd (2000) The sex-linked fidget mutation abolishes *Brn4/Pou3f4* gene expression in the embryonic inner ear. Hum Mol Genet 9(1):79–85

Phippard D, Lu L, Lee D, Saunders JC, Crenshaw EB 3rd (1999) Targeted mutagenesis of the POU-domain gene *Brn4/Pou3f4* causes developmental defects in the inner ear. J Neurosci 19(14):5980–5989

Pyrgaki C, Liu A, Niswander L (2011) Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. Dev Biol 353(1):38–49

Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA, de la Rosa J, Strong A, Lu D, Ellis P, Conte N, Yang FT, Liu P, Bradley A (2010) PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science 330(6007):1104–1107

Rajab A, Kelberman D, de Castro SC, Biebermann H, Shaikh H, Pearce K, Hall CM, Shaikh G, Gerrelli D, Grueters A, Krude H, Dattani MT (2008) Novel mutations in LHX3 are associated with hypopituitarism and sensorineural hearing loss. Hum Mol Genet 17(14):2150–2159

Rayapureddi JP, Kattamuri C, Steinmetz BD, Frankfort BJ, Ostrin EJ, Mardon G, Hegde RS (2003) Eyes absent represents a class of protein tyrosine phosphatases. Nature 426(6964):295–298

Rehman AU, Morell RJ, Belyantseva IA, Khan SY, Boger ET, Shahzad M, Ahmed ZM, Riazuddin S, Khan SN, Friedman TB (2010) Targeted capture and next-generation sequencing identifies *C9orf75*, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. Am J Hum Genet 86(3):378–388

Riazuddin S, Castelein CM, Ahmed ZM, Lalwani AK, Mastroianni MA, Naz S, Smith TN, Liburd NA, Friedman TB, Griffith AJ, Wilcox ER (2000) Dominant modifier DFNM1 suppresses recessive deafness DFNB26. Nat Genet 26(4):431–434

Riazuddin S, Nazli S, Ahmed ZM, Yang Y, Zulfiqar F, Shaikh RS, Zafar AU, Khan SN, Sabar F, Javid FT, Wilcox ER, Tsilou E, Boger ET, Sellers JR, Belyantseva IA, Friedman TB (2008) Mutation spectrum of MYO7A and evaluation of a novel nonsyndromic deafness DFNB2 allele with residual function. Hum Mutat 29(4):502–511

Richardson GP, de Monvel JB, Petit C (2011) How the genetics of deafness illuminates auditory physiology. Annu Rev Physiol 73:311–334

Rivolta MN, Grix N, Lawlor P, Ashmore JF, Jagger DJ, Holley MC (1998) Auditory hair cell precursors immortalized from the mammalian inner ear. Proc Biol Sci 265(1406):1595–1603

Robert-Moreno A, Naranjo S, de la Calle-Mustienes E, Gomez-Skarmeta JL, Alsina B (2010) Characterization of new otic enhancers of the pou3f4 gene reveal distinct signaling pathway regulation and spatio-temporal patterns. PLoS One 5(12):e15907

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4(8):651–657

Rusch A, Ng L, Goodyear R, Oliver D, Lisoukov I, Vennstrom B, Richardson G, Kelley MW, Forrest D (2001) Retardation of cochlear maturation and impaired hair cell function caused by deletion of all known thyroid hormone receptors. J Neurosci 21(24):9792–9800

Rybak LP, Talaska AE, Schacht J (2008) Drug-induced hearing loss. In: Schacht J, Fay R (eds) Auditory trauma, protection, and repair, vol 31. Springer, Dordrecht

Schmidt C, Bladt F, Goedecke S, Brinkmann V, Zschiesche W, Sharpe M, Gherardi E, Birchmeier C (1995) Scatter factor/hepatocyte growth factor is essential for liver development. Nature 373(6516):699–702

Schonberger J, Levy H, Grunig E, Sangwatanaroj S, Fatkin D, MacRae C, Stacker H, Halpin C, Eavey R, Philbin EF, Katus H, Seidman JG, Seidman CE (2000) Dilated cardiomyopathy and sensorineural hearing loss: a heritable syndrome that maps to 6q23-24. Circulation 101(15):1812–1818

Schonberger J, Wang L, Shin JT, Kim SD, Depreux FF, Zhu H, Zon L, Pizard A, Kim JB, Macrae CA, Mungall AJ, Seidman JG, Seidman CE (2005) Mutation in the transcriptional coactivator EYA4 causes dilated cardiomyopathy and sensorineural hearing loss. Nat Genet 37(4):418–422

Schultz JM (2006) Waardenburg syndrome. Semin Hear 27(3):171–181

Schultz JM, Khan SN, Ahmed ZM, Riazuddin S, Waryah AM, Chhatre D, Starost MF, Ploplis B, Buckley S, Velasquez D, Kabra M, Lee K, Hassan MJ, Ali G, Ansar M, Ghosh M, Wilcox ER, Ahmad W, Merlino G, Leal SM, Friedman TB, Morell RJ (2009) Noncoding mutations of *HGF* are associated with nonsyndromic hearing loss, DFNB39. Am J Hum Genet 85(1):25–39

Schultz JM, Yang Y, Caride AJ, Filoteo AG, Penheiter AR, Lagziel A, Morell RJ, Mohiddin SA, Fananapazir L, Madeo AC, Penniston JT, Griffith AJ (2005) Modification of human hearing loss by plasma-membrane calcium pump PMCA2. N Engl J Med 352(15):1557–1564

Schwander M, Lopes V, Sczaniecka A, Gibbs D, Lillo C, Delano D, Tarantino LM, Wiltshire T, Williams DS, Muller U (2009) A novel allele of myosin VIIa reveals a critical function for the C-terminal FERM domain for melanosome transport in retinal pigment epithelial cells. J Neurosci 29(50):15810–15818

Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A (2011) A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474(7351):337–342

Soukup GA (2009) Little but loud: small RNAs have a resounding affect on ear development. Brain Res 1277:104–114

Street VA, Li J, Robbins CA, Kallman JC (2011) A DNA variant within the MYO7A promoter regulates YY1 transcription factor binding and gene expression serving as a potential dominant DFNA11 auditory genetic modifier. J Biol Chem 286(17):15278–15286

Ting SB, Wilanowski T, Cerruti L, Zhao LL, Cunningham JM, Jane SM (2003) The identification and characterization of human sister-of-mammalian grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. Biochem J 370(Pt 3):953–962

Tootle TL, Silver SJ, Davies EL, Newman V, Latek RR, Mills IA, Selengut JD, Parlikar BE, Rebay I (2003) The transcription factor Eyes absent is a protein tyrosine phosphatase. Nature 426(6964):299–302

Toriello HV, Reardon W, Gorlin RJ (2004) Hereditary hearing loss and its syndromes, vol 2. Oxford University Press, New York

Towers ER, Kelly JJ, Sud R, Gale JE, Dawson SJ (2011) Caprin-1 is a target of the deafness gene Pou4f3 and is recruited to stress granules in cochlear hair cells in response to ototoxic damage. J Cell Sci 124(Pt 7):1145–1155

Vahava O, Morell R, Lynch ED, Weiss S, Kagan ME, Ahituv N, Morrow JE, Lee MK, Skvorak AB, Morton CC, Blumenfeld A, Frydman M, Friedman TB, King MC, Avraham KB (1998) Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. Science 279(5358):1950–1954

Van Laer L, Van Eyken E, Fransen E, Huyghe JR, Topsakal V, Hendrickx JJ, Hannula S, Maki-Torkko E, Jensen M, Demeester K, Baur M, Bonaconsa A, Mazzoli M, Espeso A, Verbruggen K, Huyghe J, Huygen P, Kunst S, Manninen M, Konings A, Diaz-Lacava AN, Steffens M, Wienker TF, Pyykko I, Cremers CW, Kremer H, Dhooge I, Stephens D, Orzan E, Pfister M, Bille M, Parving A, Sorri M, Van de Heyning PH, Van Camp G (2008) The grainyhead like 2 gene (*GRHL2*), alias *TFCP2L3*, is associated with age-related hearing impairment. Hum Mol Genet 17(2):159–169

Verhoeven K, Fagerheim T, Prasad S, Wayne S, De Clau F, Balemans W, Verstreken M, Schatteman I, Solem B, Van de Heyning P, Tranebjarg L, Smith RJ, Van Camp G (2000) Refined localization and two additional linked families for the DFNA10 locus for nonsyndromic hearing impairment. Hum Genet 107(1):7–11

Verstreken M, Declau F, Schatteman I, Van Velzen D, Verhoeven K, Van Camp G, Willems PJ, Kuhweide EW, Verhaert E, D'Haese P, Wuyts FL, Van de Heyning PH (2000) Audiometric analysis of a Belgian family linked to the DFNA10 locus. Am J Otol 21(5):675–681

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–858

Wajid M, Abbasi AA, Ansar M, Pham TL, Yan K, Haque S, Ahmad W, Leal SM (2003) DFNB39, a recessive form of sensorineural hearing impairment, maps to chromosome 7q11.22–q21.12. Eur J Hum Genet 11(10):812–815

Wallis C, Ballo R, Wallis G, Beighton P, Goldblatt J (1988) X-linked mixed deafness with stapes fixation in a Mauritian kindred: linkage to Xq probe pDP34. Genomics 3(4):299–301

Wallis D, Hamblen M, Zhou Y, Venken KJ, Schumacher A, Grimes HL, Zoghbi HY, Orkin SH, Bellen HJ (2003) The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival. Development 130(1):221–232

Warchol ME (2011) Sensory regeneration in the vertebrate inner ear: differences at the levels of cells and species. Hear Res 273(1–2):72–79

Wayne S, Robertson NG, DeClau F, Chen N, Verhoeven K, Prasad S, Tranebjarg L, Morton CC, Ryan AF, Van Camp G, Smith RJ (2001) Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus. Hum Mol Genet 10(3):195–200

Wegner M, Drolet DW, Rosenfeld MG (1993) POU-domain proteins: structure and function of developmental regulators. Curr Opin Cell Biol 5(3):488–498

Weiss S, Gottfried I, Mayrose I, Khare SL, Xiang M, Dawson SJ, Avraham KB (2003) The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity. Mol Cell Biol 23(22):7957–7964

Werth M, Walentin K, Aue A, Schonheit J, Wuebken A, Pode-Shakked N, Vilianovitch L, Erdmann B, Dekel B, Bader M, Barasch J, Rosenbauer F, Luft FC, Schmidt-Ott KM (2010) The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. Development 137(22):3835–3845

Weston MD, Pierce ML, Jensen-Smith HC, Fritzsch B, Rocha-Sanchez S, Beisel KW, Soukup GA (2011) MicroRNA-183 family expression in hair cell development and requirement of microRNAs for hair cell maintenance and survival. Dev Dyn 240(4):808–819

Weston MD, Pierce ML, Rocha-Sanchez S, Beisel KW, Soukup GA (2006) MicroRNA gene expression in the mouse inner ear. Brain Res 1111(1):95–104

White PM, Doetzlhofer A, Lee YS, Groves AK, Segil N (2006) Mammalian cochlear supporting cells can divide and trans-differentiate into hair cells. Nature 441(7096):984–987

Whitehead RH, Robinson PS (2009) Establishment of conditionally immortalized epithelial cell lines from the intestinal tissue of adult normal and transgenic mice. Am J Physiol Gastrointest Liver Physiol 296(3):G455–G460

WHO (2010) Deafness and hearing impairment. Fact sheet #300

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, Plasterk RH (2005) MicroRNA expression in zebrafish embryonic development. Science 309(5732):310–311

Wilanowski T, Tuckfield A, Cerruti L, O'Connell S, Saint R, Parekh V, Tao J, Cunningham JM, Jane SM (2002) A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. Mech Dev 114(1–2):37–50

Winter H, Braig C, Zimmermann U, Engel J, Rohbock K, Knipper M (2007) Thyroid hormone receptor alpha1 is a critical regulator for the expression of ion channels during final differentiation of outer hair cells. Histochem Cell Biol 128(1):65–75

Xiang M, Gan L, Li D, Chen ZY, Zhou L, O'Malley BW Jr, Klein W, Nathans J (1997) Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development. Proc Natl Acad Sci USA 94(17):9445–9450

Xiang M, Gao WQ, Hasson T, Shin JJ (1998) Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells. Development 125(20):3935–3946

Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D (2007) MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. J Biol Chem 282(34):25053–25066

Zhou W, Liu Z, Wu J, Liu JH, Hyder SM, Antoniou E, Lubahn DB (2006) Identification and characterization of two novel splicing isoforms of human estrogen-related receptor beta. J Clin Endocrinol Metab 91(2):569–579

Chapter 9
The Contributions of *RET* Noncoding Variation to Hirschsprung Disease

Zachary E. Stine and Andrew S. McCallion

Abstract First described by Danish pediatrician Harald Hirschsprung, Hirschsprung disease (HSCR) is a disorder of the enteric nervous system characterized by the absence of variable length of the submucous (Meissner's) and myenteric (Auerbach's) plexuses in the distal gut. As a defect in neural crest-derived cell population, Hirschsprung disease is considered a neurocristopathy. While HSCR was originally observed in sporadic cases, the advent of lifesaving surgical intervention has also given rise to the observation of familial forms of HSCR. Subsequently, its presentation in familial, sporadic, and syndromic form illuminated the genetics of HSCR. As this work has progressed the ret proto-oncogene (*RET*), a receptor tyrosine kinase has emerged as a central player in the development of HSCR, most frequently modified in effect by the contributions of risk alleles at other loci. This has been exemplified by the recent characterization of risk variants in a noncoding *RET* regulatory element, establishing it as a model for the study of multigenic disorders.

Keywords *RET* • Hirschsprung • Enhancer • Enteric nervous system • *Cis*-regulatory element • Transcriptional regulation • Disease • *NRG1* • *SOX10* • Neural crest

Z.E. Stine
McKusick – Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, BRB Suite 407, Baltimore, MD 21205, USA
e-mail: zstine@gmail.com

A.S. McCallion (✉)
Department of Molecular and Comparative Pathobiology,
Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
e-mail: amccall2@jhmi.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_9, © Springer Science+Business Media, LLC 2012

9.1 Introduction to Hirschsprung Disease and the Enteric Nervous System

9.1.1 Development of the Enteric Nervous System (ENS)

Hirschsprung disease arises from defects in the enteric nervous system (ENS), which is a part of the parasympathetic nervous system responsible for maintaining proper peristalsis, blood flow, and water and electrolyte secretion (Heanue and Pachnis 2006). The ENS is derived from the neural crest (NC), a transient and migratory group of multipotent cells which gives rise to a large number of structures and cell populations including the ENS, the sympathetic nervous system, Schwann cells, and the connective tissues of the face and neck (Douarin and Kalcheim 1999). Most enteric precursors are derived from the vagal neural crest populations, which originate in the neural tube at somites 1–7 in mammals, with lesser contributions from the sacral neural crest (Douarin and Kalcheim 1999; Heanue and Pachnis 2006; Burns 2005; Burns and Thapar 2006). Enteric neural crest-derived cells enter the foregut at 4 weeks gestation in humans (embryonic 9–9.5 days in mice) and migrate in a rostrocaudal direction to completely populate the gut by 7 weeks gestation in humans (embryonic day 15 in mice) (Newgreen and Young 2002; Druckenbrod and Epstein 2005; Heanue and Pachnis 2006). An exquisite balance between cell survival, migration, and differentiation is critical for the proper colonization of the gut by the enteric nervous system (Holland-Cunz et al. 2003; McCallion and Chakravarti 2001). Alterations in this balance through changes in *RET* dosage are central to HSCR.

9.1.2 HSCR Classification and Epidemiology

The incidence of HSCR is approximately 1 in every 5,000 live births (Bodian and Carter 1963; Amiel et al. 2008). However, incidence varies with ethnicity, ranging from 1 in 10,000 births in Hispanic populations to 1 in 3,700 births in Asian populations (Kenny et al. 2010; Amiel et al. 2008). HSCR can be classified by the length of the enteric aganglionosis. The most common form of isolated HSCR, comprising approximately 80% of cases, is termed classical (or short segment) HSCR (S-HSCR) (Bodian and Carter 1963), involves aganglionosis of the rectum, rectosigmoid colon up to but not including the splenic flexure (Martucciello 2008; Kenny et al. 2010; Amiel et al. 2008; Badner et al. 1990). The remaining ~20% of HCSR cases are termed long-segment HSCR (L-HSCR), where the enteric aganglionosis extends beyond the splenic flexure (Bodian and Carter 1963; Amiel et al. 2008; Kenny et al. 2010) and total colonic aganglionosis (Moore and Zaahl 2009). Long segment and short segment HSCR generally display different modes of inheritance and epidemiological

profiles (Amiel et al. 2008). While long-segment Hirschsprung tends to observe a dominant model of inheritance, short-segment Hirschsprung disease is more compatible with a multifactorial or recessive model of inheritance (Badner et al. 1990; Amiel et al. 2008). Consistent with its multifactorial inheritance, Hirschsprung occurrence also shows a pronounced sex bias, occurring two- to fourfold more frequently in men (Bodian and Carter 1963) dependent on segment length affected (S-HSCR, 4.4:1 male to female L-HSCR, 1.9:1 ratio) (Badner et al. 1990). Similarly, L-HSCR shows higher penetrance (52% males; 40% females) than short-segment HSCR (17% males; 4% females) (Amiel et al. 2008; Kenny et al. 2010). Importantly, recent analyses of variation underlying the multifactorial inheritance of HSCR have revealed the potential contributions made by regulatory mutations, specifically at *RET*.

9.1.3 Syndromic Hirschsprung

While HSCR presents as an isolated trait in 70% of cases, it can also be present with additional anomalies (Amiel et al. 2008; Godbole 2004). Approximately 12% of cases of HSCR have associated chromosomal abnormalities (Amiel et al. 2008), largely accounted for (90%) by trisomy 21(Down syndrome). Children with Down syndrome possess a 50–150-fold higher risk of HSCR than the general population and retain an increased male to female sex ratio (Quinn et al. 1994; Bodian and Carter 1963; Goldberg 1984; Passarge 1967; Amiel et al. 2008). Chromosomal deletions overlapping known HSCR loci constitute a large fraction of remaining cases with chromosomal anomalies (reviewed by Amiel et al. 2008; Kenny et al. 2010 and citations therein), with the remainder of associated structural lesions remaining increasingly rare (reviewed by Amiel et al. 2008; Kenny et al. 2010 and citations therein).

Approximately 18% of HSCR patients have associated anomalies without identified chromosomal abnormalities (Amiel et al. 2008). Syndromic HSCR largely comprises a spectrum of neural crest defect (Amiel et al. 2008; Bolande 1974). These HSCR-associated neurocristopathies include Shah-Waardenburg, Yemenite deaf-blind hypopigmentation, ermine phenotype/BADS, piebaldism, Haddad syndrome, and multiple endocrine neoplasia type 2A (medullary thyroid carcinoma; pheochromocytoma) (reviewed by Amiel et al. 2008 and citations therein). Hirschsprung disease also presents as part of other syndromes including Goldberg, HSCR with limb anomalies, and Mowat-Wilson. HSCR is also less commonly associated with other syndromic disorders (Amiel et al. 2008) including Bardet-Biedl syndrome, cartilage-hair hypoplasia, and Pitt-Hopkins syndrome, and the co-occurrence of HSCR and congenital anomalies of the kidney and urinary tract (CAKUT) (reviewed by Amiel et al. 2008 and the citations within).

9.2 Introduction to *RET*

9.2.1 Genetics Shows RET Centrality to HSCR

Major advances have been made in understanding the genetics of HSCR using linkage and association studies, candidate gene sequencing, and animal models. Below is a brief summary of the genetic and animal model data that illuminates the role of the genes and how they interact in ENS development.

The RET tyrosine kinase signaling pathway has been shown to be central in HSCR, beginning with a report of chromosome 10 interstitial deletions overlapping *RET* in HSCR patients, linkage to *RET* in an HSCR pedigree, and reports of *RET* mutations in HSCR patients (Puliti et al. 1993; Luo et al. 1993; Yin et al. 1994; Angrist et al. 1995). *RET*, first identified as a transforming gene (Takahashi et al. 1985), was also shown to be mutated in many multiple endocrine neoplasia type 2 (MEN2) patients. MEN2 had previously been reported to segregate in families and to copresent in patients along with HSCR (Mahaffey et al. 1990; Verdy et al. 1982; Smith et al. 1994), further implicating *RET* in HSCR. Animal models support a central role of *Ret* in the development of the enteric nervous system, with *Ret* expression detected in the developing enteric nervous system throughout vertebrates. Definitive proof came when *Ret*-deficient mice were shown to exhibit enteric aganglionosis from the stomach to the recto-anal junction (Pachnis et al. 1993; Robertson and Mason 1995; Marcos-Gutiérrez et al. 1997; Tsuzuki et al. 1995; Schuchardt et al. 1995; Enomoto et al. 2001; Schuchardt et al. 1994). Additionally, *Ret* null mice also exhibit renal agenesis and renal dysgenesis, consistent with the co-occurrence of HSCR and congenital anomalies of the kidney and urinary tract (CAKUT) (Pini Prato et al. 2009; Schuchardt et al. 1994; Schuchardt et al. 1996; Myers et al. 1999). Both HSCR patient studies and animal models have also implicated multiple components of the *RET* signaling pathway in enteric nervous system development and pathogenesis. For example, low penetrance mutations in the RET ligand, the glial cell-derived neurotrophic factor (*GDNF*), have also been identified in multiple HSCR patients (Ruiz-Ferrer et al. 2011; Angrist et al. 1996; Salomon et al. 1996).

9.2.2 RET Ligand Binding Activates Downstream Signaling Pathways

During development, *RET* is expressed in the developing central and peripheral nervous system, ENS, and excretory system (McCallion and Chakravarti 2008). RET signaling is activated by the binding of one of its four ligands (GDNF; neurturin, NRTN; persephin, PSPN; artemin, ARTN) mediated by its co-receptors (GDNF family receptor alpha 1–4, GFRA1-4) (McCallion and Chakravarti 2008). Upon the binding of the ligand and co-receptor to the extracellular RET ligand-binding domain, intracellular tyrosine residues are phosphorylated, triggering receptor

dimerization and autophosphorylation setting off various signal transduction pathways (McCallion and Chakravarti 2008; Angrist et al. 1996). RET has been shown to activate NF-kappaB (Ludwig et al. 2001), c-Jun N-terminal kinase (JNK; (Shin et al. 2004; Chiariello et al. 1998)), extracellular signal-regulated kinase (RAS/ERK; (Besset et al. 2000)), p38 mitogen-activated protein kinase (MAPK; (Hayashi et al. 2000; Ohiwa et al. 1997)), and phosphatidylinositol 3-kinase (PI3K/AKT) signaling pathways (Hayashi et al. 2000)).

9.2.3 *Discovery of RET as a Disease Locus with Dosage Sensitivity*

*RET (re*arranged during *t*ransfection) is a critical developmental gene, and as such, its expression is tightly regulated. Spatial or quantitative misexpression of RET can yield oncogenic effects. For example, *RET* rearrangements termed *RET/*PTC that were originally detected through in vitro transfection analysis were then identified in human papillary thyroid carcinoma (Grieco et al. 1990). To date, at least twelve *RET* rearrangements have been identified in papillary thyroid cancer, with *RET/PTC1* and *RET/PTC3* accounting for approximately 90% of rearrangements found in patients (Castellone and Santoro 2008). The *RET/PTC* fusion genes contain the 5′ end of a heterologous gene and the *RET* intracellular domain. While the *RET/PTC* fusion genes can cause ligand-independent *RET* signaling, its oncogenic effect is also caused by the constitutive expression in the thyroid follicular cells by placing it under the control of transcriptional regulatory elements of the 5′ portion of the fusion gene (Castellone and Santoro 2008). However, *RET* overexpression in thyroid cancer can occur in the absence of *RET* rearrangements, suggesting there are other mechanisms that can cause changes in thyroid *RET* expression (Cyniak-Magierska et al. 2011) and further implicating a potential role for changes in the *RET* regulatory landscape in disease.

9.2.4 *Additional Genes in the RET Signaling Pathway in HSCR and Enteric Nervous System Development*

As with *Ret*, *Gdnf* deficiency leads to long-segment enteric aganglionosis in mice, despite an apparently limited role for GDNF mutations in HSCR (Moore et al. 1996; Angrist et al. 1996; Eketjall and Ibanez 2002; Salomon et al. 1996). While *ARTN* mutations have yet to be reported in HSCR patients and *Artn* null mice have a normal enteric nervous system (Honma et al. 2002; Ruiz-Ferrer et al. 2011; Fernandez et al. 2008); mutations in *PSPN* and *NRTN* have been reported in HSCR patients (Doray et al. 1998; Ruiz-Ferrer et al. 2011). Unlike *Ret* or *Gdnf* mutant mice, *Nrtn* and *Pspn* null mice are viable, but *Nrtn* null mice do exhibit decreased enteric plexus density and decreased enteric motility (Heuckeroth et al. 1999; Tomac et al. 2002).

Additionally, although evidence of *GFRA1* and *GFRA2* mutations in HSCR patients is lacking, *Gfra1* plays an important role in enteric nervous system development and survival (Myers et al. 1999), and *Gfra2* null mice also show defects in the cholinergic myenteric plexus in the small intestine (Rossi et al. 1999).

9.2.5 RET Function in the Enteric Nervous System

Mouse models have helped provide insight into the complex role of Ret signaling in the enteric nervous system and how Ret signaling deficits translate to decreased enteric colonization. Decreased *Ret* expression in mice in migrating enteric neural crest-derived cells leads to impaired cell survival (Uesaka et al. 2008). *RET* controls enteric neuronal precursor proliferation (Chalazonitis et al. 1998; Taraviras et al. 1999), survival (Taraviras et al. 1999; Uesaka and Enomoto 2010), migration (Young et al. 2001; Natarajan et al. 2002), and differentiation (Taraviras et al. 1999). *Ret* null enteric neurons showed a delay in migration and non-apoptotic cell death that could be rescued by the expression of Bcl-xL (BCL2L1), an anti-apoptotic protein which had previously been shown to rescue enteric neuron cell death induced by *Gdnf* deficiency (Uesaka and Enomoto 2010; Edlich et al. 2011; Uesaka et al. 2007). However, the Bcl-xL rescued enteric neuron precursors did not undergo proper differentiation, further implicating *Ret* in migration, cell survival, and differentiation (Uesaka and Enomoto 2010). Furthermore, increasing *Ret* signaling by the addition of exogenous Gdnf can alter the structure and function of the enteric nervous system (Wang et al. 2010). Finally, *Ret* has also been shown to play a role in post-migratory enteric neurons, suggesting that *Ret* has an important role beyond just the development of the enteric nervous system (Uesaka et al. 2008).

9.2.6 Upstream Regulators of RET Implicated in HSCR

Several transcription factors that have been implicated in enteric nervous system development appear to play roles in the regulation of *RET* (Burzynski et al. 2009). One of these transcription factors is the paired-like homeobox 2b (*PHOX2B*) that was shown to bind the *RET* promoter and is mutated in central congenital hypoventilation syndrome (CCHS), which often presents with HSCR (Pattyn et al. 1999; de Pontual et al. 2007, 2006; Trang et al. 2005; Leon et al. 2009). In addition, *Phox2b* null enteric-fated NC cells do not express *Ret* (Pattyn et al. 1999; de Pontual et al. 2007, 2006; Trang et al. 2005; Leon et al. 2009). SRY (sex determining region Y)-box 10 (*SOX10*) mutations have also been identified in many HSCR patients with Waardenburg-Shah type 4 (WS4) (Kuhlbrodt et al. 1998), while *Sox10*-deficient mice exhibit enteric aganglionosis and pigment abnormalities as observed in WS4 patients (Touraine et al. 2000; Southard-Smith et al. 1999). A direct role for SOX10 in regulating RET has now been postulated by several groups (Puppo et al. 2002; Lang

and Epstein 2003; Lang et al. 2000; Emison et al. 2010; Leon et al. 2009). *ZFHX1B* (*ZEB2*) mutations have been detected in Mowat-Wilson patients with HSCR (Wakamatsu et al. 2001), while *Zfhx1b*-deficient mice exhibit enteric aganglionosis due to defects in the formation of the vagal neural crest (Van de Putte et al. 2003). Similarly, the achaete-scute complex homolog 1 (ASCL1) transcription factor can activate the *RET* promoter in neuroblastoma cell lines, although *Ascl1* (*Mash1*)-null mice exhibit defects in only a subset of enteric neural crest-derived cells (Blaugrund et al. 1996).

9.3 *RET* Regulatory Element Variation in HSCR

9.3.1 Genetic Evidence of RET Regulatory Mutations in HSCR

While *RET* coding mutations have been implicated as central in HSCR, there has been mounting evidence of noncoding mutations at the *RET* locus in HSCR. *RET* mutations account for approximately 80% of all known HSCR mutations (Amiel et al. 2008; Emison et al. 2010). While, as discussed above, there are a large number of HSCR modifier genes (Trang et al. 2005; Amiel et al. 2008, 2007; Druckenbrod et al. 2008), *RET* is the sole gene implicated in all forms of HSCR risk (Emison et al. 2010). Analysis of HSCR in families supports the central role of *RET in* HSCR. While in 11 of 12 multiplex HSCR families studied by Bolk and colleagues *RET* alleles segregated with the disease, only 50% of patients had identified *RET* coding mutations (Bolk et al. 2000). In other studies, *RET* mutations have been identified in only 15–20% of sporadic HSCR cases and 50% of familial HSCR cases (Amiel et al. 2008; Sancandi et al. 2000; Attié et al. 1995; Angrist et al. 1995; Garcia-Barcelo et al. 2004).

Several studies report the overrepresentation and overtransmission of a synonymous SNP in exon 2 of the *RET* gene (A45A; rs1800858) in HSCR cases compared with controls (Fitze et al. 2003; Borrego et al. 2000, 1999). This SNP is contained in a haplotype comprising six markers in the 5′ region of *RET,* including variants in the *RET* promoter, 5 and 1 bp upstream of the *RET* transcriptional start site (−5G>A, rs10900296; −1C>A, rs10900297; (Pelet et al. 2005; Fitze et al. 2003)). This haplotype was present in approximately 55–60% of European HSCR cases, versus 16–30% of controls (Burzynski et al. 2004; Pelet et al. 2005). Additionally, this haplotype was present in approximately 88% of Chinese cases, versus 47% of controls (Emison et al. 2010). The haplotype also spans 23 kb from the promoter region to intron 2 and was observed to be significantly overtransmitted in cases of sporadic HSCR with no identified *RET* mutation (Pelet et al. 2005). Additionally, many HSCR patients with no observed *RET* mutation were homozygous for the haplotype (Pelet et al. 2005). Importantly, patients with the HSCR risk haplotype exhibited lower levels of *RET* expression in gut tissues (Miao et al. 2010), suggesting a direct genotype-phenotype correlation in *RET* dosage (Emison et al. 2010).

However, the identification of an overrepresented haplotype in cases versus controls did not mean that the causative allele had been identified. Additional studies were initiated to determine if the HSCR risk haplotype contains a causative allele and to establish the mechanism of the causative allele's role in the genesis of HSCR. Questions about the functional role of the risk alleles in HSCR caused additional studies to focus on SNPs within the risk haplotype, with the hypothesis that there may be an ancient low penetrance locus upstream of the exon 2 SNP affecting *RET* transcription (Amiel et al. 2008; Sancandi et al. 2000). Using a comparative genomics strategy to identify putative regulatory elements based on the hypothesis that functional elements are conserved due to negative selection on functional nucleotides (Pennacchio et al. 2006; Visel et al. 2008; Nobrega et al. 2003; Visel et al. 2009; Nobrega and Pennacchio 2004), the Chakravarti group focused on a conserved sequence within the HSCR-associated haplotype that is located in the first intron of *RET* (Emison et al. 2005). Using a transmission disequilibrium test (TDT) on 28 SNPs spanning 175 kb around the *RET* locus, the greatest statistical significance for association with HSCR lay within the previously identified 27.5-kb HSCR risk haplotype (Fig. 9.1a). Importantly, SNPs within the HSCR risk haplotype showed the greatest transmission distortions in HSCR. While resequencing of the patients revealed no *RET* coding sequence mutations, a SNP termed *RET*+3 (rs2435357) lying within a conserved 900 base pair element in the first *RET* intron was identified (Fig. 9.1b, c). This conserved element was termed *RET* MCS (multispecies conserved sequence) +9.7, due to its location 9.7 kb downstream from the *RET* transcriptional start site (Emison et al. 2005) (Fig. 9.1b). *RET* MCS +9.7, which contained two additional variants which are in complete linkage disequilibrium with *RET*+3 (rs2506005 and rs2506004; Fig. 9.1c), showed the highest transmission distortion and statistical significance in HSCR trios of the SNPs in the risk haplotype (Emison et al. 2005). While the *RET*+3:T allele was overtransmitted in HSCR, the *RET*+3:C allele is highly conserved in mammals, suggesting that *RET*+3:T may impact a conserved functional element (Fig. 9.1a–c (Emison et al. 2005)). However, additional functional characterization of *RET*+9.7 was required to test if any SNPs within the element functionally impacted *RET* expression during development of the enteric nervous system (Emison et al. 2005).

9.3.2 Characterization of RET MCS +9.7 Function

The cell type-specific regulatory activity of the *RET* MCS +9.7 element was then tested in vitro (Emison et al. 2005). While *RET* MCS +9.7 directed negligible regulatory activity in a luciferase assay in the nonneuronal HeLa cell line, *RET* MCS +9.7 directed strong regulatory activity in the Neuro-2A neuroblastoma cell line (Fig. 9.2a, (Emison et al. 2005)). Since *RET* MCS +9.7 exhibited neuronal cell activity, it suggests that nucleotide variation in this element could affect *RET* expression in neuronal development (Emison et al. 2005). Additionally, *RET* MCS +9.7 was also capable of binding Neuro-2A nuclear lysate, further supporting its role as a neuroblastoma

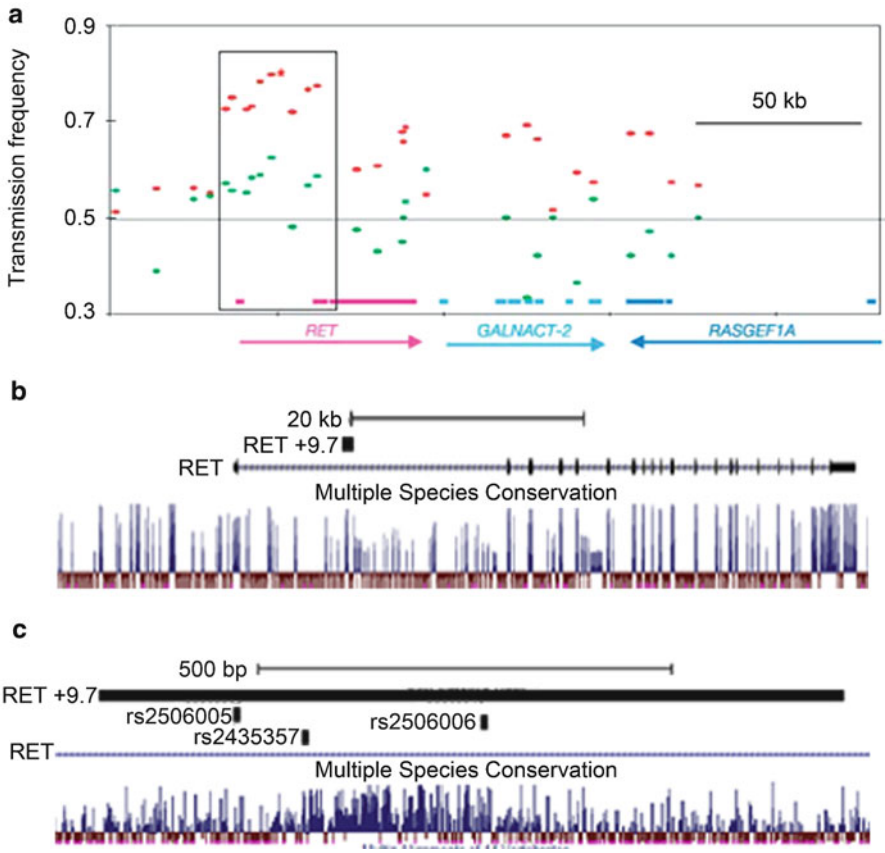

Fig. 9.1 *The overtransmitted HSCR-associated SNP lies within a conserved region*. (**a**) Transmission disequilibrium test identifies an overtransmitted SNP (T is overtransmitted derived/mutant allele, C is ancestral/wild-type allele) part of the HSCR-associated haplotype located within a conserved intronic *RET* element termed *RET* MCS +9.7 (Emison et al. 2005). *Red* = TDT of individual SNPs transmitted to affected offspring, *Green* = TDT of individual SNPs transmitted to non-affected offspring (Adapted with permission from Emison et al. 2005). (**b**) UCSC genome browser (www.genome.ucsc.edu) representation of the human *RET* gene showing the location of the *RET* MCS + 9.7 element and the phastCons mammalian conservation track (Siepel et al. 2005). (**c**) UCSC genome browser representation of the human *RET* MCS + 9.7 element showing the location of the HSCR risk alleles (Emison et al. 2005) and the phastCons mammalian conservation track (Siepel et al. 2005)

enhancer (Grice et al. 2005). Importantly, *RET* MCS +9.7 containing the HSCR risk allele exhibited six- to eightfold lower regulatory activity than the other allele in Neuro-2A cells, indicating that the risk alleles decrease the potential enhancer activity of *RET* MCS +9.7 (Emison et al. 2005) (Fig. 9.2a). Since HSCR is caused by a decrease in *RET* dosage, an allele decreasing *RET* expression through compromised enhancer function is consistent with HSCR biology (Grice et al. 2005).

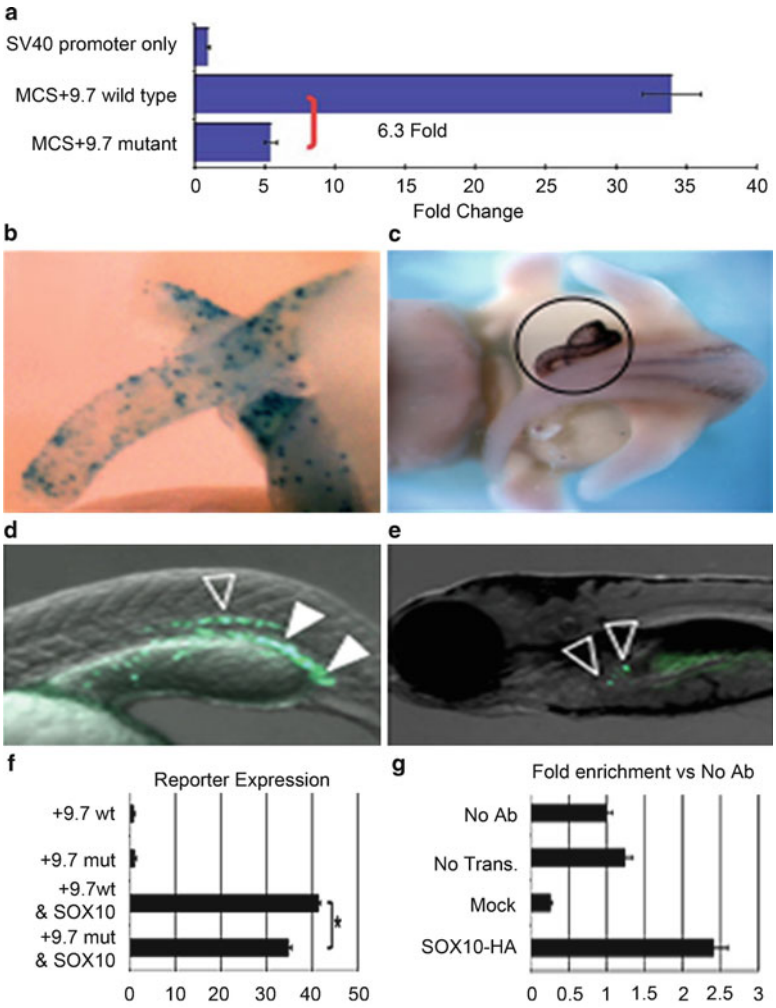

Fig. 9.2 *HSCR-associated SNP compromises in vitro cell-specific regulatory activity of a conserved RET intronic enhancer that directs in vivo enteric nervous system reporter expression.* (**a**) The mutant HSCR allele compromises *RET* MCS +9.7 luciferase activity in Neuro-2A cells compared to wild-type *RET* MCS +9.7 wild-type = C allele, mutant = overtransmitted T allele (Adapted with permission from Emison et al. 2005). (**b**) *RET* MCS +9.7 directs LacZ reporter expression in the external gut loop in vivo in embryonic 12.5 mice (Adapted with permission from Grice et al. (2005)). (**c**) *Ret* expression in the embryonic 12.5 gut loop detected by in situ hybridization (Adapted with permission from Grice et al. 2005). (**d** and **e**) *RET* MCS +9.7 directs *ret* appropriate eGFP reporter expression in vivo in transgenic zebrafish, *open arrow* head = enteric nervous system, *solid white arrow* head = pronephric duct (Adapted with permission from Fisher et al. 2006a). (**f**) Exogenous SOX10 transactivation of *RET* MCS +9.7 directed reporter expression is compromised by the HSCR mutant risk allele in HeLa cells (Adapted with permission from Emison et al. 2010). (**g**) Chromatin immunoprecipitation detects physical interaction between SOX10-HA and *RET* MCS +9.7, fold enrichment versus no antibody control; Ab = antibody, trans = transformed cells (Adapted with permission from Emison et al. 2010)

However, in vitro regulatory activity can give only limited information about a regulatory element's spatial and temporal activity during in vivo development. Using a LacZ reporter vector stably integrated into mice, *RET* MCS +9.7 in vivo enhancer activity was assayed (Grice et al. 2005). *RET* MCS +9.7 directed reporter expression in the developing enteric nervous system at mouse embryonic day 12.5 (E12.5; Fig. 9.2b), consistent with endogenous *Ret* expression (Fig. 9.2c) (Grice et al. 2005). Additionally, *RET* MCS +9.7 directed expression in the dorsal root ganglia and cranial ganglia, other NC-derived tissues (Grice et al. 2005). The in vivo regulatory activity of *RET* MCS +9.7 was also tested in zebrafish, using Tol2-mediated stable transgenesis (Fisher et al. 2006a, b). Once again, *RET* MCS +9.7 directed *ret* appropriate enteric nervous system reporter expression (Fig. 9.2d–e). The *Ret* appropriate regulatory activity of *RET* MCS +9.7 in vivo in the enteric nervous system and other neural crest-derived tissues further support the potential causative role of variation within the *RET* MCS +9.7 in HSCR risk (Fisher et al. 2006a; Grice et al. 2005).

Although the *RET* MCS +9.7 has been shown to be an enteric nervous system enhancer and that HSCR risk variants compromise its in vitro regulatory activity, the identification of the functional variants, the mechanism of the functional effects of *RET* MCS +9.7 variation and how that translates to disease risk remains to be elucidated. While the *RET* +3 was shown to lie between two retinoic acid response elements motifs, no transcription factor binding sites were disrupted (Emison et al. 2005). Additional work has been done to try and identify factors that bind to *RET* MCS +9.7, in particular factors whose binding may be disrupted by the HSCR risk alleles. SOX10, which has a well-documented role in HSCR (Kuhlbrodt et al. 1998; Amiel et al. 2008), is believed to be an upstream regulator of *RET* (Lang and Epstein 2003; Puppo et al. 2002). Within *RET* MCS +9.7, the risk allele rs2435357 was observed to be overlapping a putative *SOX10* transcription factor binding motif (Emison et al. 2010). Ectopic *SOX10* expression in HeLa cells, where *RET* MCS +9.7 was previously shown to lack regulatory activity (Emison et al. 2005; Grice et al. 2005), was sufficient to induce *RET* MCS +9.7 directed luciferase expression (Emison et al. 2010) (Fig. 9.2f). Also, the rs2435357:T risk allele and mutation of the SOX10 binding site compromised *SOX10 RET* MCS +9.7 response to ectopic *SOX10* expression (Fig. 9.2f). Additionally, chromatin immunoprecipitation (ChIP) demonstrated that SOX10 can bind *RET* MCS +9.7 directly in neuroblastoma cells (Emison et al. 2010) (Fig. 9.2g). Furthermore, the rs2506004 HSCR-associated SNP within *RET* MCS +9.7 was reported to disrupt an NXF-ARNT2 and SIM2-ARNT2 binding motif (Sribudiani et al. 2011). Single-minded homolog 2 (*Sim2*),

aryl-hydrocarbon receptor nuclear translocator 2 (*Arnt2*), and *Nxf* (*Npas4*) are all expressed in neural crest stem cells isolated from an embryonic mouse gut and regulate endogenous *RET* expression when transfected into neuroblastoma cells (Sribudiani et al. 2011).

Taken together, these data suggest that one or more variants within *RET* MCS +9.7 may have functional effects on the element's regulatory activity by compromising different transcription factor binding sites.

9.3.3 *Genetic Interactions of the* RET HSCR *Risk Haplotype with Modifier Genes*

Due to its subtle dosage effect on *RET* transcription and low penetrance compared to a loss of function allele (Emison et al. 2010), the *RET* HSCR risk haplotype genotype-phenotype correlation is likely to be highly subjected to modification by alleles at other genes or by other alleles in *trans* to *RET*. After the initial characterization of the *RET* risk haplotype, several studies set out to test its genetic interaction with modifier loci. A range of potential interactions have now been reported. One study demonstrated a role for the *RET* MCS +9.7 risk allele in syndromic HSCR presenting with CCHS, Bardet-Biedl syndrome (BBS), and Down syndrome (de Pontual et al. 2007), but not in Mowat-Wilson syndrome or Waardenburg-Shah syndrome (de Pontual et al. 2007). Epistatic interactions were observed between mutations in various BBS genes and alleles in *RET* intron 1, including a novel 11 bp located near the HSCR-associated alleles in *RET* MCS +9.7. Epistasis between BBS genes and *RET* signaling was also shown using morpholinos in zebrafish (de Pontual et al. 2009). Similarly, genetic interaction between the *RET* +9.7 risk allele and chromosome 21 gene dosage was reported by another group, demonstrating a significant difference in the risk allele frequency between patients with both Down syndrome and HSCR, compared to patients with either Down syndrome alone or HSCR alone (Arnold et al. 2009). Another genome-wide association study in a collection of Chinese sporadic HSCR patients revealed a genetic interaction between the *RET* risk haplotype and two SNPs (rs16879552 and rs7835688) in intron 1 of neuregulin1 (*NRG1*), a gene found to harbor coding sequence mutations in HSCR patients (Garcia-Barcelo et al. 2009; Tang et al. 2011a, b). The genetic interaction between the *RET* risk haplotype increased the odds ratio from 2.3-fold to 19.5-fold in the presence of heterozygous *NRG1* alleles (Garcia-Barcelo et al. 2009). Genetic interactions in Chinese HSCR patients have also been reported between two different *HOX* loci (*HOXA13* and *HOXB7*) and the *RET* risk haplotype (Garcia-Barceló et al. 2007).

The epistatic interactions between genes in human populations are mirrored in mouse models of enteric aganglionosis. While *Ret* heterozygous null mice have a very low penetrance of enteric aganglionosis and *Ret* null homozygous mice do not exhibit sex differences in enteric phenotype, *Ret* null and an endothelin receptor type B (*Ednrb*) allelic series showed a two-locus non-complementation that recapitulated the incomplete penetrance, variance in length of aganglionosis, and

sex variance observed in human HSCR patients (McCallion et al. 2003). Genetic interactions have also been observed between *Sox10* and *Ednrb* and endothelin 3 (*Edn3*) in mouse models (Cantrell et al. 2004; Stanchina et al. 2006). Additionally, the penetrance of the *Sox10* mutant mouse phenotype is modified by the presence of *Sox8* mutations (Maka et al. 2005). *Sox10* also exhibits genetic interactions with the genes *Zfhx1b* and L1 cell adhesion molecule (*L1cam*) (Stanchina et al. 2010; Wallace et al. 2010). *L1cam* also acts as an enteric nervous system development modifier gene for *Ednrb* (Wallace et al. 2011).

9.3.4 *RET* MCS +9.7 Sequence Variation, Distribution, and Genetic Properties

The worldwide distribution of the *RET* MCS +9.7 HSCR-associated risk variant was also studied (Emison et al. 2005). The *RET*+3:T allele frequency was 0.45 in Asia and 0.25 in Europe, but was below 0.01 in Africa. These allele frequencies correlate with a higher frequency of short-segment HSCR in Asia than Europe and a lower rate of short-segment HSCR in Africa (Emison et al. 2005). The *RET* MCS +9.7 risk allele has been found to be associated with HSCR in multiple populations, including Chinese and European populations (Emison et al. 2010, 2005;). Transmission disequilibrium tests across a panel of SNPs in the European and Chinese populations suggested that the disease alleles lie in two identical haplotypes, suggesting that the disease haplotypes have common origin (Emison et al. 2005). The haplotype was also shown to be overtransmitted in Taiwanese and an additional Chinese HSCR population (Zhang et al. 2007; Liu et al. 2008; Wu et al. 2010).

The mechanism causing gender differences in HSCR occurrence is not well understood, with starkly limited evidence for a major role of X-linked mutation in HSCR (Fernandez et al. 2010; Broman et al. 2006; Emison et al. 2005). However, the *RET* MCS +9.7 risk allele does show sex-specific effects. *RET*+3:T allele was transmitted to HSCR affected male offspring more often than to female offspring (Emison et al. 2005). Additionally, transmission of the risk allele to affected offspring caused a larger increase in susceptibility in males than females (Emison et al. 2005). Furthermore, *RET*+3:T shows higher penetrance in males than females. The risk allele accounted for 2.6% and 1.1% of the total susceptibility variance in males and females, respectively, while known coding mutations account for only 0.1% of the total susceptibility variance (Emison et al. 2005). While the *RET*+3:T risk allele penetrance is similar in European and Chinese populations and had a genetic effect on the three lengths of HSCR (short, long, and total colonic aganglionosis), the risk allele penetrance varies across the three lengths of HSCR (Emison et al. 2010). Additionally, the risk allele was observed to be in trans of coding mutations and had a lower allele frequency in HSCR patients with a *RET* coding sequence mutation compared to HSCR patients lacking a *RET* mutation (Emison et al. 2010).

9.3.5 Evidence for Regulatory Variation of Other RET Regulatory Elements

Due to the traditional focus on the characterization of proximal promoter regulatory elements and the −5 and −1 *RET* HSCR risk alleles, additional studies have focused on characterizing the impact of the *RET* promoter SNPs (Griseri et al. 2005; Garcia-Barcelo et al. 2005; Fitze et al. 2003). The *RET* proximal promoter region had previously been shown to direct *RET* appropriate regulatory control in LacZ expressing mice and in vitro in NC-derived cell lines (Sukumaran et al. 2001; Andrew et al. 2000). However, the *RET* proximal promoter region lacked enteric nervous system expression in vivo (Sukumaran et al. 2001). Using an in vitro assay in neuroblastoma cells lines, it was shown that the promoter region containing the risk alleles showed decreased regulatory activity (Fitze et al. 2003). One group reported that the *RET* promoter SNPs showed a significant correlation with HSCR in a Chinese population (Garcia-Barcelo et al. 2005). Additionally, the promoter SNPs overlapped an NK2 homeobox 1 (NKX2-1 or TTF-1) binding site, a factor that can activate the *RET* promoter. Further, they reported a mutation in *NKX2-1,* which has a similar expression pattern to *RET* in the developing human gut, in an HSCR patient that compromised *RET* promoter activation (Garcia-Barcelo et al. 2005). An additional *NKX2-1* mutant that compromised *RET* promoter activation in vitro was identified in two Caucasians with HSCR (Garcia-Barceló et al. 2007). Additionally, lack of *Nkx2.1* expression in the mouse gut shows there could be species-specific effects on enteric development. Nkx2-1 was shown to cooperate with Phox2b and Sox10 to regulate the *RET* promoter in vitro (Leon et al. 2009). However, while *RET* expression was shown to be reduced in lymphocytes in patients homozygous for risk alleles (Griseri et al. 2005), a functional impact on *RET* enteric expression remains to be investigated further. Due to the proposed role of enhancer and promoter interactions in transcriptional regulation (Dekker 2006; Amano et al. 2009), it is possible that the *RET* promoter variants in the HSCR risk haplotype may further compromise *RET* ENS transcription impacted by *RET* MCS +9.7 variants.

9.3.6 RET as a Model Locus of Regulatory Topology

Due to its key role in development and disease, the sequences that control *RET* transcriptional regulation and the transcription factors that bind them have been a focus of many studies. Both in vitro and in vivo studies have been used to identify *RET* locus regulatory elements, including in-depth studies of the *RET* promoter. The *RET* proximal promoter region has been shown to direct *RET* appropriate regulatory activity in vivo, including in the developing excretory system (Zordan et al. 2006). Additionally, the promoter has been found to be regulated by Sp1 and Sp3 in vitro (Andrew et al. 2000). Subsequently, studies began to identify regulatory elements outside of the proximal promoter region. A *RET* enhancer was identified approximately 3.3 kb upstream of the *RET* promoter which is regulated by Sox10 and paired

box 3 (Pax3) (Lang and Epstein 2003; Puppo et al. 2002). Comparative genomic analysis coupled with an in vitro enhancer screen identified five cell type-specific *RET* locus enhancers, including one with in vivo *RET* appropriate regulatory activity (Grice et al. 2005). Several of these cell-specific *RET* enhancers, plus several other sequences from the zebrafish *ret* locus, exhibited *ret* appropriate regulatory control in vivo using a zebrafish-stable transgenic assay (Fisher et al. 2006a). Additionally, potential regulatory elements at the *RET* locus identified by genome-wide ChIP studies have exhibited *ret* appropriate regulatory control in zebrafish (Stine et al. 2011).

9.4 Emerging Connections

9.4.1 RET Modulation by Steroid Hormones in Development and Disease

While genetic interactions between genes in HSCR have been a subject of intense study, much remains to be explained about how diet and exocrine signaling affect the penetrance of HSCR predisposing variance. Recent work showed that the regulation of *RET* by members of the steroid hormone nuclear receptor super family could play an important role in regulating *RET* enteric aganglionosis. Retinoic acid signaling, which is known to be critical in the development and patterning of the nervous system (Maden 2007), is mediated by its receptors (retinoic acid receptors, RAR; and retinoid X receptors (McGrane 2007)). Retinoic acid, a vitamin A derivative, has long been known to regulate *Ret* in kidney development (Moreau et al. 1998; Batourina et al. 2001), while the expression of a dominant negative retinoic acid receptor abolishes *Ret* expression in the developing kidney and *Ret*-mediated ureteric bud formation and branching morphogenesis (Rosselot et al. 2010). *RET* is also directly regulated by retinoic acid in neural crest-derived neuroblastoma cell lines (Bunone et al. 1995; Angrisano et al. 2011; Yamada et al. 2007) and is central to the transcriptional program required for retinoic acid-induced neuroblastoma differentiation (Yamada et al. 2007; Oppenheimer et al. 2007). Furthermore, *Ret* and its co-receptors have been shown to be downregulated in vivo by retinoic acid in developing chick sensory neurons and in rat developing heart neurons (Doxakis and Davies 2005; Shoba et al. 2002). Additionally, *RET* is upregulated by retinoic acid in breast cancer through a retinoic acid response element (Hua et al. 2009; Stine et al. 2011). Despite this extensive evidence of retinoic acid regulation of *RET,* much work remains to understand the interaction between *Ret* and retinoic acid in the developing enteric nervous system. Retinoic acid treatment increased the number of Ret antibody marked cells in ENS precursor primary culture (Sato and Heuckeroth 2008). Disruption of *Raldh2* (*Aldh1a2*), a retinoic acid synthesis protein, led to enteric aganglionosis in mice (Niederreither et al. 2003). Finally, *Ret* heterozygous null mice were shown to interact with vitamin A-depleted mice, increasing the length and severity of intestinal aganglionosis of retinol binding protein 4 (*Rbp4*) null mice (Fu et al. 2010).

Recent evidence also suggests a role for the steroid hormone estrogen in regulating *RET* through its receptor (estrogen receptor alpha, ESR1). Estrogen treatment has been shown to upregulate *RET* in developing mouse kidney explants, with sex-dependent response to GDNF addition (Walker et al. 2009). Since *RET* mutations are associated with renal dysgenesis (Skinner et al. 2008), it is possible that estrogen regulation of *RET* and sex differences in RET signaling could contribute to sex differences in the occurrence of renal agenesis (Parikh et al. 2002). Estrogen signaling, an important regulator of breast cancer, has been shown to upregulate *RET* mRNA levels in breast cancer cell lines (Lin et al. 2007; Carroll et al. 2006). Despite the lack of *RET* mutations in breast cancer (Kan et al. 2010), *RET* mRNA positively correlates with ESR1 expression in breast cancer cell lines (Plaza-Menacho et al. 2010; Esseghir et al. 2007; Boulay et al. 2008; Tozlu et al. 2006). Additionally, *RET* locus estrogen response elements have been identified in breast cancer cell lines (Stine et al. 2011; Tan et al. 2011). Importantly, RET-dependent signaling appears to play a role in estrogen independence and antiestrogen resistance in breast cancer (Plaza-Menacho et al. 2010; Kang et al. 2010). In addition, this *RET*-mediated estrogen independence appears to require ESR1 phosphorylation, causing ligand-independent transcriptional regulation (Plaza-Menacho et al. 2010). This constitutive activation of ESR1 by estrogen responsive *RET* suggests a possible autoregulatory loop (Plaza-Menacho et al. 2010). While there have been reports of potential *RET* regulatory loops (Burzynski et al. 2009), to date, the mechanism of this autoregulatory loop has not been defined. The possibility that ESR1 may be involved in a RET autoregulatory loop remains to be explored. The ESR1 and ESR2 (estrogen receptor beta) receptors are expressed broadly throughout the central and peripheral nervous system, including *RET* expressing neuronal populations such as the dorsal root ganglia (Loven et al. 2010; McCarthy 2008; Bennett et al. 2003; Zoubina and Smith 2001). There have also been reports of estrogen receptor expression in the enteric nervous system (Kawano et al. 2004; Campbell-Thompson et al. 2001), raising the potential that if ESR1 regulates *RET* in the enteric nervous system, it could contribute to sex difference in HSCR. Although ectopic ESR1 expression in neuroblastoma induces neuronal differentiation similar to the *RET*-dependent retinoic acid-induced neuronal differentiation (Loven et al. 2010), the role of *RET* in this estrogen-induced neuronal differentiation remains to be explored. Additionally, cross talk between retinoic acid signaling and estrogen signaling in the regulation of *RET* in breast cancer raises the possibility of estrogen and retinoic acid cross talk in the enteric nervous system (Hua et al. 2009; Stine et al. 2011; Ross-Innes et al. 2010). However, the role of estrogen in the enteric nervous system is speculative and remains to be explored.

9.5 Conclusions

RET is central to HSCR. The discovery of an HSCR-associated haplotype lacking a coding mutation made *RET* a model for the study of noncoding mutations in disease. A SNP lying within a conserved intronic enhancer showing in vivo ENS

regulatory activity compromised the neuronal in vitro regulatory activity of the element. This element is bound and transactivated by SOX10 in vitro, with the risk allele disrupting the *Sox10* binding site and reducing *SOX10* responsiveness. The reports of genetic interactions between the *RET* risk haplotype and other HSCR risk alleles indicate that much remains to be understood about the complex interactions between noncoding regulatory mutations and modifier genes. Furthermore, regulation of *RET* by steroid hormones suggests that diet and environment could further affect HSCR penetrance.

Abbreviations

BBS	Bardet-Biedl syndrome
CAKUT	Congenital anomalies of the kidney and urinary tract
CCHS	Central congenital hypoventilation syndrome
ENS	Enteric nervous system
HSCR	Hirschsprung disease
L-HSCR	Long-segment Hirschsprung disease
MCS	Multispecies conserved sequence
MEN2	Multiple endocrine neoplasia type 2
NC	Neural crest
RET	ret proto-oncogene
S-HSCR	Short-segment or classical Hirschsprung
SNP	Single-nucleotide polymorphism
WS4	Waardenburg-Shah type 4

References

Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T (2009) Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell 16(1):47–57

Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, Plouin P, Carter NP, Lyonnet S, Munnich A, Colleaux L (2007) Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet 80(5):988–993

Amiel J, Sproat-Emison E, Garcia-Barcelo M, Lantieri F, Burzynski G, Borrego S, Pelet A, Arnold S, Miao X, Griseri P, Brooks AS, Antinolo G, de Pontual L, Clement-Ziza M, Munnich A, Kashuk C, West K, Wong KKY, Lyonnet S, Chakravarti A, Tam PKH, Ceccherini I, Hofstra RMW, Fernandez R, Hirschsprung Disease C (2008) Hirschsprung disease, associated syndromes and genetics: a review. J Med Genet 45(1):1–14

Andrew SD, Delhanty PJ, Mulligan LM, Robinson BG (2000) Sp1 and Sp3 transactivate the RET proto-oncogene promoter. Gene 256(1–2):283–291

Angrisano T, Sacchetti S, Natale F, Cerrato A, Pero R, Keller S, Peluso S, Perillo B, Avvedimento VE, Fusco A, Bruni CB, Lembo F, Santoro M, Chiariotti L (2011) Chromatin and DNA methylation

dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. Nucleic Acids Res 39(6):1993–2006. doi:gkq864, [pii] 10.1093/nar/gkq864

Angrist M, Bolk S, Thiel B, Puffenberger EG, Hofstra RM, Buys CH, Cass DT, Chakravarti A (1995) Mutation analysis of the RET receptor tyrosine kinase in Hirschsprung disease. Hum Mol Genet 4(5):821–830

Angrist M, Bolk S, Halushka M, Lapchak PA, Chakravarti A (1996) Germline mutations in glial cell line-derived neurotrophic factor (GDNF) and RET in a Hirschsprung disease patient. Nat Genet 14(3):341–344

Arnold S, Pelet A, Amiel J, Borrego S, Hofstra R, Tam P, Ceccherini I, Lyonnet S, Sherman S, Chakravarti A (2009) Interaction between a chromosome 10 RET enhancer and chromosome 21 in the Down syndrome-Hirschsprung disease association. Hum Mutat 30(5):771–775

Attié T, Pelet A, Edery P, Eng C, Mulligan LM, Amiel J, Boutrand L, Beldjord C, Nihoul-Fékété C, Munnich A (1995) Diversity of RET proto-oncogene mutations in familial and sporadic Hirschsprung disease. Hum Mol Genet 4(8):1381–1386

Badner JA, Sieber WK, Garver KL, Chakravarti A (1990) A genetic study of Hirschsprung disease. Am J Hum Genet 46(3):568–580

Batourina E, Gim S, Bello N, Shy M, Clagett-Dame M, Srinivas S, Costantini F, Mendelsohn C (2001) Vitamin A controls epithelial/mesenchymal interactions through Ret expression. Nat Genet 27(1):74–78

Bennett HL, Gustafsson JA, Keast JR (2003) Estrogen receptor expression in lumbosacral dorsal root ganglion cells innervating the female rat urinary bladder. Auton Neurosci 105(2):90–100. doi:S1566-0702(03)00044-4, [pii] 10.1016/S1566-0702(03)00044-4

Besset V, Scott RP, Ibáñez CF (2000) Signaling complexes and protein-protein interactions involved in the activation of the Ras and phosphatidylinositol 3-kinase pathways by the c-Ret receptor tyrosine kinase. J Biol Chem 275(50):39159–39166

Blaugrund E, Pham TD, Tennyson VM, Lo L, Sommer L, Anderson DJ, Gershon MD (1996) Distinct subpopulations of enteric neuronal progenitors defined by time of development, sympathoadrenal lineage markers and Mash-1-dependence. Development 122(1):309–320

Bodian M, Carter OO (1963) A family study of Hirschsprung's disease. Ann Hum Genet 26(3): 261–277

Bolande RP (1974) The neurocristopathies: a unifying concept of disease arising in neural crest maldevelopment. Hum Pathol 5(4):409–429

Bolk S, Pelet A, Hofstra RM, Angrist M, Salomon R, Croaker D, Buys CH, Lyonnet S, Chakravarti A (2000) A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. Proc Natl Acad Sci USA 97(1):268–273

Borrego S, Sáez ME, Ruiz A, Gimm O, López-Alonso M, Antiñolo G, Eng C (1999) Specific polymorphisms in the RET proto-oncogene are over-represented in patients with Hirschsprung disease and may represent loci modifying phenotypic expression. J Med Genet 36(10):771–774

Borrego S, Ruiz A, Saez ME, Gimm O, Gao X, López-Alonso M, Hernández A, Wright FA, Antiñolo G, Eng C (2000) RET genotypes comprising specific haplotypes of polymorphic variants predispose to isolated Hirschsprung disease. J Med Genet 37(8):572–578

Boulay A, Breuleux M, Stephan C, Fux C, Brisken C, Fiche M, Wartmann M, Stumm M, Lane HA, Hynes NE (2008) The Ret receptor tyrosine kinase pathway functionally interacts with the ER{alpha} pathway in breast cancer. Cancer Res 68(10):3743–3751

Broman KW, Sen S, Owens SE, Manichaikul A, Southard-Smith EM, Churchill GA (2006) The X chromosome in quantitative trait locus mapping. Genetics 174(4):2151–2158. doi:genetics.106.061176, [pii] 10.1534/genetics.106.061176

Bunone G, Borrello MG, Picetti R, Bongarzone I, Peverali FA, de Franciscis V, Della Valle G, Pierotti MA (1995) Induction of RET proto-oncogene expression in neuroblastoma cells precedes neuronal differentiation and is not mediated by protein synthesis. Exp Cell Res 217(1): 92–99

Burns AJ (2005) Migration of neural crest-derived enteric nervous system precursor cells to and within the gastrointestinal tract. Int J Dev Biol 49(2–3):143–150

Burns AJ, Thapar N (2006) Advances in ontogeny of the enteric nervous system. Neurogastroenterol Motil 18(10):876–887

Burzynski GM, Nolte IM, Osinga J, Ceccherini I, Twigt B, Maas S, Brooks A, Verheij J, Plaza Menacho I, Buys CHCM, Hofstra RMW (2004) Localizing a putative mutation as the major contributor to the development of sporadic Hirschsprung disease to the RET genomic sequence between the promoter region and exon 2. Eur J Hum Genet EJHG 12(8):604–612

Burzynski G, Shepherd IT, Enomoto H (2009) Genetic model system studies of the development of the enteric nervous system, gut motility and Hirschsprung's disease. Neurogastroenterol Motil 21(2):113–127

Campbell-Thompson M, Reyher KK, Wilkinson LB (2001) Immunolocalization of estrogen receptor alpha and beta in gastric epithelium and enteric neurons. J Endocrinol 171(1):65–73. doi:JOE04021 [pii]

Cantrell VA, Owens SE, Chandler RL, Airey DC, Bradley KM, Smith JR, Southard-Smith EM (2004) Interactions between Sox10 and EdnrB modulate penetrance and severity of aganglionosis in the Sox10Dom mouse model of Hirschsprung disease. Hum Mol Genet 13(19):2289–2301

Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M (2006) Genome-wide analysis of estrogen receptor binding sites. Nat Genet 38(11):1289–1297

Castellone MD, Santoro M (2008) Dysregulated RET signaling in thyroid cancer. Endocrinol Metab Clin North Am 37(2):363–374, viii-363–374, viii

Chalazonitis A, Rothman TP, Chen J, Gershon MD (1998) Age-dependent differences in the effects of GDNF and NT-3 on the development of neurons and glia from neural crest-derived precursors immunoselected from the fetal rat gut: expression of GFRalpha-1 in vitro and in vivo. Dev Biol 204(2):385–406

Chiariello M, Visconti R, Carlomagno F, Melillo RM, Bucci C, de Franciscis V, Fox GM, Jing S, Coso OA, Gutkind JS, Fusco A, Santoro M (1998) Signalling of the Ret receptor tyrosine kinase through the c-Jun NH2-terminal protein kinases (JNKS): evidence for a divergence of the ERKs and JNKs pathways induced by Ret. Oncogene 16(19):2435–2445

Cyniak-Magierska A, Wojciechowska-Durczyńska K, Krawczyk-Rusiecka K, Zygmunt A, Lewiński A (2011) Assessment of RET/PTC1 and RET/PTC3 rearrangements in fine-needle aspiration biopsy specimens collected from patients with Hashimoto's thyroiditis. Thyroid Res 4(1):5

de Pontual L, Pelet A, Trochet D, Jaubert F, Espinosa-Parrilla Y, Munnich A, Brunet JF, Goridis C, Feingold J, Lyonnet S, Amiel J (2006) Mutations of the RET gene in isolated and syndromic Hirschsprung's disease in human disclose major and modifier alleles at a single locus. J Med Genet 43(5):419–423. doi:jmg.2005.040113, [pii] 10.1136/jmg.2005.040113

de Pontual L, Pelet A, Clement-Ziza M, Trochet D, Antonarakis SE, Attie-Bitach T, Beales PL, Blouin JL, Dastot-Le Moal F, Dollfus H, Goossens M, Katsanis N, Touraine R, Feingold J, Munnich A, Lyonnet S, Amiel J (2007) Epistatic interactions with a common hypomorphic-RET allele in syndromic Hirschsprung disease. Hum Mutat 28(8):790–796

de Pontual L, Zaghloul NA, Thomas S, Davis EE, McGaughey DM, Dollfus H, Baumann C, Bessling SL, Babarit C, Pelet A, Gascue C, Beales P, Munnich A, Lyonnet S, Etchevers H, Attie-Bitach T, Badano JL, McCallion AS, Katsanis N, Amiel J (2009) Epistasis between RET and BBS mutations modulates enteric innervation and causes syndromic Hirschsprung disease. Proc Natl Acad Sci USA 106(33):13921–13926

Dekker J (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods 3(1):17–21

Doray B, Salomon R, Amiel J, Pelet A, Touraine R, Billaud M, Attié T, Bachy B, Munnich A, Lyonnet S (1998) Mutation of the RET ligand, neurturin, supports multigenic inheritance in Hirschsprung disease. Hum Mol Genet 7(9):1449–1452

Douarin NL, Kalcheim C (1999) The neural crest. Cambridge University Press, Cambridge

Doxakis E, Davies AM (2005) Retinoic acid negatively regulates GDNF and neurturin receptor expression and responsiveness in embryonic chicken sympathetic neurons. Mol Cell Neurosci 29(4):617–627

Druckenbrod NR, Epstein ML (2005) The pattern of neural crest advance in the cecum and colon. Dev Biol 287(1):125–133

Druckenbrod NR, Powers PA, Bartley CR, Walker JW, Epstein ML (2008) Targeting of endothelin receptor-B to the neural crest. Genesis (New York, NY: 2000) 46(8):396–400

Edlich F, Banerjee S, Suzuki M, Cleland MM, Arnoult D, Wang C, Neutzner A, Tjandra N, Youle RJ (2011) Bcl-x(L) retrotranslocates Bax from the mitochondria into the cytosol. Cell 145(1):104–116. doi:S0092-8674(11)00186-3 , [pii] 10.1016/j.cell.2011.02.034

Eketjall S, Ibanez CF (2002) Functional characterization of mutations in the GDNF gene of patients with Hirschsprung disease. Hum Mol Genet 11(3):325–329

Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED, Chakravarti A (2005) A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature 434(7035):857–863

Emison ES, Garcia-Barcelo M, Grice EA, Lantieri F, Amiel J, Burzynski G, Fernandez RM, Hao L, Kashuk C, West K, Miao X, Tam PKH, Griseri P, Ceccherini I, Pelet A, Jannot A-S, de Pontual L, Henrion-Caude A, Lyonnet S, Verheij JBGM, Hofstra RMW, Antiñolo G, Borrego S, McCallion AS, Chakravarti A (2010) Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 87(1): 60–74

Enomoto H, Crawford PA, Gorodinsky A, Heuckeroth RO, Johnson EM, Milbrandt J (2001) RET signaling is essential for migration, axonal growth and axon guidance of developing sympathetic neurons. Development 128(20):3963–3974

Esseghir S, Todd SK, Hunt T, Poulsom R, Plaza-Menacho I, Reis-Filho JS, Isacke CM (2007) A role for glial cell derived neurotrophic factor induced expression by inflammatory cytokines and RET/GFR{alpha}1 receptor Up-regulation in breast cancer. Cancer Res 67(24):11732–11741

Fernandez RM, Ruiz-Ferrer M, Lopez-Alonso M, Antiñolo G, Borrego S (2008) Polymorphisms in the genes encoding the 4 RET ligands, GDNF, NTN, ARTN, PSPN, and susceptibility to Hirschsprung disease. J Pediatr Surg 43(11):2042–2047

Fernandez RM, Nunez-Torres R, Gonzalez-Meneses A, Antinolo G, Borrego S (2010) Novel association of severe neonatal encephalopathy and Hirschsprung disease in a male with a duplication at the Xq28 region. BMC Med Genet 11:137. doi:1471-2350-11-137, [pii] 10.1186/1471-2350-11-137

Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006a) Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312(5771):276–279

Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K, McCallion AS (2006b) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc 1(3):1297–1305. doi:nprot.2006.230, [pii] 10.1038/nprot.2006.230

Fitze G, Appelt H, König IR, Görgens H, Stein U, Walther W, Gossen M, Schreiber M, Ziegler A, Roesner D, Schackert HK (2003) Functional haplotypes of the RET proto-oncogene promoter are associated with Hirschsprung disease (HSCR). Hum Mol Genet 12(24):3207–3214

Fu M, Sato Y, Lyons-Warren A, Zhang B, Kane MA, Napoli JL, Heuckeroth RO (2010) Vitamin A facilitates enteric nervous system precursor migration by reducing Pten accumulation. Development 137(4):631–640

Garcia-Barcelo M, Sham MH, Lee WS, Lui VC, Chen BL, Wong KK, Wong JS, Tam PK (2004) Highly recurrent RET mutations and novel mutations in genes of the receptor tyrosine kinase and endothelin receptor B pathways in Chinese patients with sporadic Hirschsprung disease. Clin Chem 50(1):93–100. doi:10.1373/clinchem.2003.022061, clinchem.2003.022061 [pii]

Garcia-Barcelo M, Ganster RW, Lui VC, Leon TY, So MT, Lau AM, Fu M, Sham MH, Knight J, Zannini MS, Sham PC, Tam PK (2005) TTF-1 and RET promoter SNPs: regulation of RET transcription in Hirschsprung's disease. Hum Mol Genet 14(2):191–204. doi:ddi015, [pii] 10.1093/hmg/ddi015

Garcia-Barceló MM, Miao X, Lui VCH, So MT, Ngan ESW, Leon TYY, Lau DKC, Liu TT, Lao X, Guo W, Holden WT, Moore J, Tam PKH (2007) Correlation between genetic variations in Hox clusters and Hirschsprung's disease. Ann Hum Genet 71(4):526–536

Garcia-Barcelo M-M, Tang CS-M, Ngan ES-W, Lui VC-H, Chen Y, So M-T, Leon TY-Y, Miao X-P, Shum CK-Y, Liu F-Q, Yeung M-Y, Yuan Z-W, Guo W-H, Liu L, Sun X-B, Huang L-M, Tou J-F, Song Y-Q, Chan D, Cheung KMC, Wong KK-Y, Cherny SS, Sham P-C, Tam PK-H (2009) Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. Proc Natl Acad Sci 106(8):2694–2699

Godbole K (2004) Many faces of Hirschsprung's disease. Indian Pediatr 41(11):1115–1123

Goldberg EL (1984) An epidemiological study of Hirschsprung's disease. Int J Epidemiol 13(4):479–485

Grice EA, Rochelle ES, Green ED, Chakravarti A, McCallion AS (2005) Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Hum Mol Genet 14(24):3837–3845

Grieco M, Santoro M, Berlingieri MT, Melillo RM, Donghi R, Bongarzone I, Pierotti MA, Della Porta G, Fusco A, Vecchio G (1990) PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. Cell 60(4):557–563

Griseri P, Bachetti T, Puppo F, Lantieri F, Ravazzolo R, Devoto M, Ceccherini I (2005) A common haplotype at the 5′ end of the RET proto-oncogene, overrepresented in Hirschsprung patients, is associated with reduced gene expression. Hum Mutat 25(2):189–195

Hayashi H, Ichihara M, Iwashita T, Murakami H, Shimono Y, Kawai K, Kurokawa K, Murakumo Y, Imai T, Funahashi H, Nakao A, Takahashi M (2000) Characterization of intracellular signals via tyrosine 1062 in RET activated by glial cell line-derived neurotrophic factor. Oncogene 19(39):4469–4475

Heanue TA, Pachnis V (2006) Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. Proc Natl Acad Sci USA 103(18):6919–6924

Heuckeroth RO, Enomoto H, Grider JR, Golden JP, Hanke JA, Jackman A, Molliver DC, Bardgett ME, Snider WD, Johnson EM Jr, Milbrandt J (1999) Gene targeting reveals a critical role for neurturin in the development and maintenance of enteric, sensory, and parasympathetic neurons. Neuron 22(2):253–263

Holland-Cunz S, Krammer HJ, Süss A, Tafazzoli K, Wedel T (2003) Molecular genetics of colorectal motility disorders. Eur J Pediatr Surg Off J Austrian Assoc Pediatr Surg [et Al] = Zeitschrift Für Kinderchirurgie 13(3):146–151

Honma Y, Araki T, Gianino S, Bruce A, Heuckeroth RO, Johnson EM Jr, Milbrandt J (2002) Artemin is a vascular-derived neurotropic factor for developing sympathetic neurons. Neuron 35(2):267–282

Hua S, Kittler R, White KP (2009) Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. Cell 137(7):1259–1271

Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W, Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, Faham M, Seshagiri S (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466(7308):869–873. doi:nature09208, [pii] 10.1038/nature09208

Kang J, Qian PX, Pandey V, Perry JK, Miller LD, Liu ET, Zhu T, Liu DX, Lobie PE (2010) Artemin is estrogen regulated and mediates antiestrogen resistance in mammary carcinoma. Oncogene 29(22):3228–3240. doi:onc201071, [pii] 10.1038/onc.2010.71

Kawano N, Koji T, Hishikawa Y, Murase K, Murata I, Kohno S (2004) Identification and localization of estrogen receptor alpha- and beta-positive cells in adult male and female mouse intestine at various estrogen levels. Histochem Cell Biol 121(5):399–405. doi:10.1007/s00418-004-0644-6

Kenny SE, Tam PKH, Garcia-Barcelo M (2010) Hirschsprung's disease. Semin Pediatr Surg 19(3):194–200

Kuhlbrodt K, Schmidt C, Sock E, Pingault V, Bondurand N, Goossens M, Wegner M (1998) Functional analysis of Sox10 mutations found in human Waardenburg-Hirschsprung patients. J Biol Chem 273(36):23033–23038

Lang D, Epstein JA (2003) Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. Hum Mol Genet 12(8):937–945

Lang D, Chen F, Milewski R, Li J, Lu MM, Epstein JA (2000) Pax3 is required for enteric ganglia formation and functions with Sox10 to modulate expression of c-ret. J Clin Invest 106(8): 963–971

Leon TYY, Ngan ESW, Poon H-C, So M-T, Lui VCH, Tam PKH, Garcia-Barcelo MM (2009) Transcriptional regulation of RET by Nkx2-1, Phox2b, Sox10, and Pax3. J Pediatr Surg 44(10): 1904–1912

Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei C-L, Liu ET (2007) Whole-genome cartography of estrogen receptor α binding sites. PLoS Genet 3(6):e87–e87

Liu C, Jin L, Li H, Lou J, Luo C, Zhou X, Li J-C (2008) RET polymorphisms and the risk of Hirschsprung's disease in a Chinese population. J Hum Genet 53(9):825–833

Loven J, Zinin N, Wahlstrom T, Muller I, Brodin P, Fredlund E, Ribacke U, Pivarcsi A, Pahlman S, Henriksson M (2010) MYCN-regulated microRNAs repress estrogen receptor-alpha (ESR1) expression and neuronal differentiation in human neuroblastoma. Proc Natl Acad Sci USA 107(4):1553–1558. doi:0913517107, [pii] 10.1073/pnas.0913517107

Ludwig L, Kessler H, Wagner M, Hoang-Vu C, Dralle H, Adler G, Böhm BO, Schmid RM (2001) Nuclear factor-kappaB is constitutively active in C-cell carcinoma and required for RET-induced transformation. Cancer Res 61(11):4526–4535

Luo Y, Ceccherini I, Pasini B, Matera I, Bicocchi MP, Barone V, Bocciardi R, Kääriäinen H, Weber D, Devoto M (1993) Close linkage with the RET protooncogene and boundaries of deletion mutations in autosomal dominant Hirschsprung disease. Hum Mol Genet 2(11):1803–1808

Maden M (2007) Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat Rev Neurosci 8(10):755–765

Mahaffey SM, Martin LW, McAdams AJ, Ryckman FC, Torres M (1990) Multiple endocrine neoplasia type II B with symptoms suggesting Hirschsprung's disease: a case report. J Pediatr Surg 25(1):101–103

Maka M, Stolt CC, Wegner M (2005) Identification of Sox8 as a modifier gene in a mouse model of Hirschsprung disease reveals underlying molecular defect. Dev Biol 277(1):155–169

Marcos-Gutiérrez CV, Wilson SW, Holder N, Pachnis V (1997) The zebrafish homologue of the ret receptor and its pattern of expression during embryogenesis. Oncogene 14(8):879–889

Martucciello G (2008) Hirschsprung's disease, one of the most difficult diagnoses in pediatric surgery: a review of the problems from clinical practice to the bench. Eur J Pediatr Surg 18(3): 140–149

McCallion AS, Chakravarti A (2001) EDNRB/EDN3 and Hirschsprung disease type II. Pigment Cell Res Sponsored Eur Soc Pigm Cell Res Int Pigm Cell Soc 14(3):161–169

McCallion AS, Chakravarti A (2008) RET, Hirschsprung disease and multiple endocrine neoplasia type 2. In: Epstein C, Erickson R, Wynshaw-Boris A (eds) Inborn errors of development, 2nd edn. Oxford University Press, San Francisco

McCallion AS, Stames E, Conlon RA, Chakravarti A (2003) Phenotype variation in two-locus mouse models of Hirschsprung disease: tissue-specific interaction between Ret and Ednrb. Proc Natl Acad Sci USA 100(4):1826–1831

McCarthy MM (2008) Estradiol and the developing brain. Physiol Rev 88(1):91–124. doi:88/1/91, [pii] 10.1152/physrev.00010.2007

McGrane MM (2007) Vitamin A regulation of gene expression: molecular mechanism of a prototype gene. J Nutr Biochem 18(8):497–508. doi:S0955-2863(06)00265-8, [pii] 10.1016/j.jnutbio.2006.10.006

Miao X, Leon TY-Y, Ngan ES-W, So M-T, Yuan Z-W, Lui VC-H, Chen Y, Wong KK-Y, Tam PK-H, Garcia-Barceló M (2010) Reduced RET expression in gut tissue of individuals carrying risk alleles of Hirschsprung's disease. Hum Mol Genet 19(8):1461–1467

Moore SW, Zaahl M (2009) Clinical and genetic differences in total colonic aganglionosis in Hirschsprung's disease. J Pediatr Surg 44(10):1899–1903

Moore MW, Klein RD, Fariñas I, Sauer H, Armanini M, Phillips H, Reichardt LF, Ryan AM, Carver-Moore K, Rosenthal A (1996) Renal and neuronal abnormalities in mice lacking GDNF. Nature 382(6586):76–79

Moreau E, Vilar J, Lelièvre-Pégorier M, Merlet-Bénichou C, Gilbert T (1998) Regulation of c-ret expression by retinoic acid in rat metanephros: implication in nephron mass control. Am J Physiol 275(6 Pt 2):F938–945, F938-945

Myers SM, Salomon R, Goessling A, Pelet A, Eng C, von Deimling A, Lyonnet S, Mulligan LM (1999) Investigation of germline GFR alpha-1 mutations in Hirschsprung disease. J Med Genet 36(3):217–220

Natarajan D, Marcos-Gutierrez C, Pachnis V, de Graaff E (2002) Requirement of signalling by receptor tyrosine kinase RET for the directed migration of enteric nervous system progenitor cells during mammalian embryogenesis. Development 129(22):5151–5160

Newgreen D, Young HM (2002) Enteric nervous system: development and developmental disturbances–part 2. Pediatr Dev Pathol 5(4):329–349

Niederreither K, Vermot J, Le Roux I, Schuhbaur B, Chambon P, Dolle P (2003) The regional pattern of retinoic acid synthesis by RALDH2 is essential for the development of posterior pharyngeal arches and the enteric nervous system. Development 130(11):2525–2534

Nobrega MA, Pennacchio LA (2004) Comparative genomic analysis as a tool for biological discovery. J Physiol 554(Pt 1):31–39. doi:jphysiol.2003.050948, [pii] 10.1113/jphysiol.2003.050948

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) Scanning human gene deserts for long-range enhancers. Science (New York, NY) 302(5644):413–413

Ohiwa M, Murakami H, Iwashita T, Asai N, Iwata Y, Imai T, Funahashi H, Takagi H, Takahashi M (1997) Characterization of Ret-Shc-Grb2 complex induced by GDNF, MEN 2A, and MEN 2B mutations. Biochem Biophys Res Commun 237(3):747–751

Oppenheimer O, Cheung N-K, Gerald WL (2007) The RET oncogene is a critical component of transcriptional programs associated with retinoic acid-induced differentiation in neuroblastoma. Mol Cancer Ther 6(4):1300–1309

Pachnis V, Mankoo B, Costantini F (1993) Expression of the c-ret proto-oncogene during mouse embryogenesis. Development 119(4):1005–1017

Parikh CR, McCall D, Engelman C, Schrier RW (2002) Congenital renal agenesis: case–control analysis of birth characteristics. Am J Kidney Dis 39(4):689–694. doi:S0272-6386(02)72952-1, [pii] 10.1053/ajkd.2002.31982

Passarge E (1967) The genetics of Hirschsprung's disease. Evidence for heterogeneous etiology and a study of sixty-three families. N Engl J Med 276(3):138–143

Pattyn A, Morin X, Cremer H, Goridis C, Brunet JF (1999) The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. Nature 399(6734):366–370

Pelet A, de Pontual L, Clément-Ziza M, Salomon R, Mugnier C, Matsuda F, Lathrop M, Munnich A, Feingold J, Lyonnet S, Abel L, Amiel J (2005) Homozygosity for a frequent and weakly penetrant predisposing allele at the RET locus in sporadic Hirschsprung disease. J Med Genet 42(3):e18–e18

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118):499–502

Pini Prato A, Musso M, Ceccherini I, Mattioli G, Giunta C, Ghiggeri GM, Jasonni V (2009) Hirschsprung disease and congenital anomalies of the kidney and urinary tract (CAKUT): a novel syndromic association. Medicine 88(2):83–90

Plaza-Menacho I, Morandi A, Robertson D, Pancholi S, Drury S, Dowsett M, Martin LA, Isacke CM (2010) Targeting the receptor tyrosine kinase RET sensitizes breast cancer cells to tamoxifen treatment and reveals a role for RET in endocrine resistance. Oncogene 29(33):4648–4657. doi:onc2010209, [pii] 10.1038/onc.2010.209

Puliti A, Covone AE, Bicocchi MP, Bolino A, Lerone M, Martucciello G, Jasonni V, Romeo G (1993) Deleted and normal chromosome 10 homologs from a patient with Hirschsprung disease

isolated in two cell hybrids through enrichment by immunomagnetic selection. Cytogenet Cell Genet 63(2):102–106

Puppo F, Griseri P, Fanelli M, Schena F, Romeo G, Pelicci P, Ceccherini I, Ravazzolo R, Patrone G (2002) Cell-line specific chromatin acetylation at the Sox10-Pax3 enhancer site modulates the RET proto-oncogene expression. FEBS Lett 523(1–3):123–127

Quinn FMJ, Surana R, Puri P (1994) The influence of trisomy 21 on outcome in children with Hirschsprung's disease. J Pediatr Surg 29(6):781–783

Robertson K, Mason I (1995) Expression of ret in the chicken embryo suggests roles in regionalisation of the vagal neural tube and somites and in development of multiple neural crest and placodal lineages. Mech Dev 53(3):329–344

Rosselot C, Spraggon L, Chia I, Batourina E, Riccio P, Lu B, Niederreither K, Dolle P, Duester G, Chambon P, Costantini F, Gilbert T, Molotkov A, Mendelsohn C (2010) Non-cell-autonomous retinoid signaling is crucial for renal development. Development 137(2):283–292

Rossi J, Luukko K, Poteryaev D, Laurikainen A, Sun YF, Laakso T, Eerikäinen S, Tuominen R, Lakso M, Rauvala H, Arumäe U, Pasternack M, Saarma M, Airaksinen MS (1999) Retarded growth and deficits in the enteric and parasympathetic nervous system in mice lacking GFR alpha2, a functional neurturin receptor. Neuron 22(2):243–252

Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, Massie CE, Vowler SL, Eldridge M, Carroll JS (2010) Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. Genes Dev 24(2):171–182. doi:24/2/171, [pii] 10.1101/gad.552910

Ruiz-Ferrer M, Torroglosa A, Luzón-Toro B, Fernández RM, Antiñolo G, Mulligan LM, Borrego S (2011) Novel mutations at RET ligand genes preventing receptor activation are associated to Hirschsprung's disease. J Mol Med (Berlin, Germany) 89(5):471–480

Salomon R, Attié T, Pelet A, Bidaud C, Eng C, Amiel J, Sarnacki S, Goulet O, Ricour C, Nihoul-Fékété C, Munnich A, Lyonnet S (1996) Germline mutations of the RET ligand GDNF are not sufficient to cause Hirschsprung disease. Nat Genet 14(3):345–347

Sancandi M, Ceccherini I, Costa M, Fava M, Chen B, Wu Y, Hofstra R, Laurie T, Griffths M, Burge D, Tam PK (2000) Incidence of RET mutations in patients with Hirschsprung's disease. J Pediatr Surg 35(1):139–142, discussion 142–143-139–142; discussion 142–143

Sato Y, Heuckeroth RO (2008) Retinoic acid regulates murine enteric nervous system precursor proliferation, enhances neuronal precursor differentiation, and reduces neurite growth in vitro. Dev Biol 320(1):185–198

Schuchardt A, D'Agati V, Larsson-Blomberg L, Costantini F, Pachnis V (1994) Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature 367(6461):380–383. doi:10.1038/367380a0

Schuchardt A, D'Agati V, Larsson-Blomberg L, Costantini F, Pachnis V (1995) RET-deficient mice: an animal model for Hirschsprung's disease and renal agenesis. J Intern Med 238(4):327–332

Schuchardt A, D'Agati V, Pachnis V, Costantini F (1996) Renal agenesis and hypodysplasia in ret-k-mutant mice result from defects in ureteric bud development. Development 122(6): 1919–1929

Shin E, Hong S-W, Kim SH, Yang W-I (2004) Expression of down stream molecules of RET (p-ERK, p-p38 MAPK, p-JNK and p-AKT) in papillary thyroid carcinomas. Yonsei Med J 45(2):306–313

Shoba T, Dheen ST, Tay SS (2002) Retinoic acid influences the expression of the neuronal regulatory genes Mash-1 and c-ret in the developing rat heart. Neurosci Lett 318(3):129–132. doi:S0304394001024910 [pii]

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15(8):1034–1050. doi:gr.3715005, [pii] 10.1101/gr.3715005

Skinner MA, Safford SD, Reeves JG, Jackson ME, Freemerman AJ (2008) Renal aplasia in humans is associated with RET mutations. Am J Hum Genet 82(2):344–351. doi:S0002-9297(08)00086-4, [pii] 10.1016/j.ajhg.2007.10.008

Smith DP, Eng C, Ponder BA (1994) Mutations of the RET proto-oncogene in the multiple endocrine neoplasia type 2 syndromes and Hirschsprung disease. J Cell Sci Suppl 18:43–49

Southard-Smith EM, Angrist M, Ellison JS, Agarwala R, Baxevanis AD, Chakravarti A, Pavan WJ (1999) The Sox10(Dom) mouse: modeling the genetic variation of Waardenburg-Shah (WS4) syndrome. Genome Res 9(3):215–225

Sribudiani Y, Metzger M, Osinga J, Rey A, Burns AJ, Thapar N, Hofstra RMW (2011) Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. Gastroenterology 140(2):572–582, e572-572-582.e572

Stanchina L, Baral V, Robert F, Pingault V, Lemort N, Pachnis V, Goossens M, Bondurand N (2006) Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development. Dev Biol 295(1):232–249

Stanchina L, Van de Putte T, Goossens M, Huylebroeck D, Bondurand N (2010) Genetic interaction between Sox10 and Zfhx1b during enteric nervous system development. Dev Biol 341(2):416–428

Stine ZE, McGaughey DM, Bessling SL, Li S, McCallion AS (2011) Steroid hormone modulation of RET through two estrogen responsive enhancers in breast cancer. Hum Mol Genet. doi:ddr291, [pii] 10.1093/hmg/ddr291

Sukumaran M, Waxman SG, Wood JN, Pachnis V (2001) Flanking regulatory sequences of the locus encoding the murine GDNF receptor, c-ret, directs lac Z (beta-galactosidase) expression in developing somatosensory system. Dev Dyn Off Pub Am Ass Anatomists 222(3):389–402

Takahashi M, Ritz J, Cooper GM (1985) Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell 42(2):581–588

Tan SK, Lin ZH, Chang CW, Varang V, Chng KR, Pan YF, Yong EL, Sung WK, Cheung E (2011) AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. EMBO J 30(13):2569–2581. doi:emboj2011151, [pii] 10.1038/emboj.2011.151

Tang CS-M, Tang W-K, So M-T, Miao X-P, Leung BM-C, Yip BH-K, Leon TY-Y, Ngan ES-W, Lui VC-H, Chen Y, Chan IH-Y, Chung PH-Y, Liu X-L, Wu X-Z, Wong KK-Y, Sham P-C, Cherny SS, Tam PK-H, Garcia-Barceló M-M (2011a) Fine mapping of the NRG1 Hirschsprung's disease locus. PLoS One 6(1):e16181–e16181

Tang CS, Ngan ES, Tang WK, So MT, Cheng G, Miao XP, Leon TY, Leung BM, Hui KJ, Lui VH, Chen Y, Chan IH, Chung PH, Liu XL, Wong KK, Sham PC, Cherny SS, Tam PK, Garcia-Barcelo MM (2011b) Mutations in the NRG1 gene are associated with Hirschsprung disease. Hum Genet. doi:10.1007/s00439-011-1035-4

Taraviras S, Marcos-Gutierrez CV, Durbec P, Jani H, Grigoriou M, Sukumaran M, Wang LC, Hynes M, Raisman G, Pachnis V (1999) Signalling by the RET receptor tyrosine kinase and its role in the development of the mammalian enteric nervous system. Development 126(12):2785–2797

Tomac AC, Agulnick AD, Haughey N, Chang C-F, Zhang Y, Bäckman C, Morales M, Mattson MP, Wang Y, Westphal H, Hoffer BJ (2002) Effects of cerebral ischemia in mice deficient in Persephin. Proc Natl Acad Sci USA 99(14):9521–9526

Touraine RL, Attié-Bitach T, Manceau E, Korsch E, Sarda P, Pingault V, Encha-Razavi F, Pelet A, Augé J, Nivelon-Chevallier A, Holschneider AM, Munnes M, Doerfler W, Goossens M, Munnich A, Vekemans M, Lyonnet S (2000) Neurological phenotype in Waardenburg syndrome type 4 correlates with novel SOX10 truncating mutations and expression in developing brain. Am J Hum Genet 66(5):1496–1503

Tozlu S, Girault I, Vacher S, Vendrell J, Andrieu C, Spyratos F, Cohen P, Lidereau R, Bieche I (2006) Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. Endocr Relat Cancer 13(4):1109–1120

Trang H, Dehan M, Beaufils F, Zaccaria I, Amiel J, Gaultier C (2005) The French congenital central hypoventilation syndrome registry: general data, phenotype, and genotype. Chest 127(1): 72–79

Tsuzuki T, Takahashi M, Asai N, Iwashita T, Matsuyama M, Asai J (1995) Spatial and temporal expression of the ret proto-oncogene product in embryonic, infant and adult rat tissues. Oncogene 10(1):191–198

Uesaka T, Enomoto H (2010) Neural precursor death is central to the pathogenesis of intestinal aganglionosis in Ret hypomorphic mice. J Neurosci Off J Soc Neurosci 30(15):5211–5218
Uesaka T, Jain S, Yonemura S, Uchiyama Y, Milbrandt J, Enomoto H (2007) Conditional ablation of GFRalpha1 in postmigratory enteric neurons triggers unconventional neuronal death in the colon and causes a Hirschsprung's disease phenotype. Development 134(11):2171–2181
Uesaka T, Nagashimada M, Yonemura S, Enomoto H (2008) Diminished Ret expression compromises neuronal survival in the colon and causes intestinal aganglionosis in mice. J Clin Invest 118(5):1890–1898
Van de Putte T, Maruhashi M, Francis A, Nelles L, Kondoh H, Huylebroeck D, Higashi Y (2003) Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. Am J Hum Genet 72(2):465–470
Verdy M, Weber AM, Roy CC, Morin CL, Cadotte M, Brochu P (1982) Hirschsprung's disease in a family with multiple endocrine neoplasia type 2. J Pediatr Gastroenterol Nutr 1(4):603–607
Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40(2):158–160
Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. Nature 461(7261):199–205
Wakamatsu N, Yamada Y, Yamada K, Ono T, Nomura N, Taniguchi H, Kitoh H, Mutoh N, Yamanaka T, Mushiake K, Kato K, Sonta S, Nagaya M (2001) Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. Nat Genet 27(4):369–370
Walker KA, Caruana G, Bertram JF, McInnes KJ (2009) Sexual dimorphism in mouse metanephroi exposed to 17 beta-estradiol in vitro. Nephron Exp Nephrol 111(2):e42–50. doi:000191104, [pii] 10.1159/000191104
Wallace AS, Schmidt C, Schachner M, Wegner M, Anderson RB (2010) L1cam acts as a modifier gene during enteric nervous system development. Neurobiol Dis 40(3):622–633
Wallace AS, Tan MX, Schachner M, Anderson RB (2011) L1cam acts as a modifier gene for members of the endothelin signalling pathway during enteric nervous system development. Neurogastroenterol Motil. doi:10.1111/j.1365-2982.2011.01692.x
Wang H, Hughes I, Planer W, Parsadanian A, Grider JR, Vohra BPS, Keller-Peck C, Heuckeroth RO (2010) The timing and location of glial cell line-derived neurotrophic factor expression determine enteric nervous system structure and function. J Neurosci Off J Soc Neurosci 30(4):1523–1538
Wu TT, Tsai TW, Chang H, Su CC, Li SY, Lai HS, Li C (2010) Polymorphisms of the RET gene in Hirschsprung disease, anorectal malformation and intestinal pseudo-obstruction in Taiwan. J Formos Med Assoc 109(1):32–38
Yamada S, Nomura T, Uebersax L, Matsumoto K, Fujita S, Miyake M, Miyake J (2007) Retinoic acid induces functional c-Ret tyrosine kinase in human neuroblastoma. Neuroreport 18(4):359–363
Yin L, Barone V, Seri M, Bolino A, Bocciardi R, Ceccherini I, Pasini B, Tocco T, Lerone M, Cywes S (1994) Heterogeneity and low detection rate of RET mutations in Hirschsprung disease. Eur J Hum Genet EJHG 2(4):272–280
Young HM, Hearn CJ, Farlie PG, Canty AJ, Thomas PQ, Newgreen DF (2001) GDNF is a chemoattractant for enteric neural cells. Dev Biol 229(2):503–516
Zhang XN, Zhou MN, Qiu YQ, Ding SP, Qi M, Li JC (2007) Genetic analysis of RET, EDNRB, and EDN3 genes and three SNPs in MCS+9.7 in Chinese Patients with isolated Hirschsprung disease. Biochem Genet 45(7–8):523–527. doi:10.1007/s10528-007-9093-y
Zordan P, Tavella S, Brizzolara A, Biticchi R, Ceccherini I, Garofalo S, Ravazzolo R, Bocciardi R (2006) The immediate upstream sequence of the mouse Ret gene controls tissue-specific expression in transgenic mice. Int J Mol Med 18(4):601–608
Zoubina EV, Smith PG (2001) Sympathetic hyperinnervation of the uterus in the estrogen receptor alpha knock-out mouse. Neuroscience 103(1):237–244. doi:S0306452200005492 [pii]

Chapter 10
Cis-Regulatory Variation and Cancer

Nora F. Wasserman and Marcelo A. Nobrega

Abstract In the traditional model of human disease genetics, mutations in coding regions of the genome were assumed to underlie disease phenotypes. It is only in the recent past that functional noncoding regions – such as promoters, enhancers and silencers – have been implicated in disease states. At its most basic level, cancer is a disease caused by the misexpression of genes normally responsible for regulating cell proliferation. It is therefore logical that mutations and variants within *cis*-regulatory elements controlling the expression of proto-oncogenes and tumor suppressor genes would underlie some tumorigenic gene expression changes. As changes in noncoding functional elements are harder to identify than alterations in protein coding sequences, many of the recent insights into *cis*-regulatory variants involved in cancer etiology have been uncovered by genome-wide association studies (GWAS), highlighting risk variants in non-genic regions. Here, we highlight examples of cancer-associated variation in promoters, enhancers, and silencers, as well as changes to the overall architecture of a gene's regulatory landscape. These functional characterizations bring us closer to understanding the role of *cis*-regulatory mutations and cancer risk/progression.

Keywords Prostate cancer • *MSMB* • Thyroid cancer • *FOXE1* • *MYC* • 8q24 • Breast cancer • *FGFR2* • Colorectal cancer • *SMAD7* • *EIF3H* • Immunoglobulin • Hematologic cancer • *TMPRSS2* • *ETS*

N.F. Wasserman • M.A. Nobrega (✉)
Department of Human Genetics, University of Chicago, 920 E 58th Street,
CLSC 319, Chicago, IL 60637, USA
e-mail: mnobrega@bsd.uchicago.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_10,

10.1 Introduction

Cancer is the uncontrolled proliferation of abnormal cells in the body. At the most basic level, this uncontrolled growth is caused by the misexpression of genes normally responsible for regulating cell division. In a healthy cell, the cell cycle is a tightly controlled process, with numerous checkpoints in place to ensure genomic integrity and functioning cell cycle machinery before allowing a cell to proceed into the next phase of the cycle. If DNA damage (caused either by random replication errors or environmental mutagens) is found, the process of division is either paused to allow time for repair or, if the damage is too great, the cell undergoes apoptosis. When proto-oncogenes – genes that positively regulate proliferation or negatively regulate apoptosis – are overexpressed, or tumor suppressor genes – those that negatively control the cell cycle or promote apoptosis – are underexpressed, the cellular checkpoints necessary for controlled division may be less rigorously executed or bypassed entirely. If the burden of mutations impacting the expression of oncogenes and tumor suppressor genes becomes great enough, uncontrolled proliferation can occur and a potentially cancerous cell is created.

The genetic reasons underlying the misexpression of proto-oncogenes and tumor suppressor genes can vary greatly. For proto-oncogenes to become oncogenes, mutations must result in an overexpression of gene product or expanded expression domain (improper spatial or temporal gene activation). This overexpression can be achieved through an increase in gene copy number – where entire chromosomes or chromosomal segments are duplicated, or localized genic regions are highly amplified – or through mutations in *cis*-regulatory elements involved in the control of gene expression (Fig. 10.1). These *cis*-regulatory elements include promoters and long-range enhancer or repressor elements that function to regulate gene expression in a tissue- and temporal-specific manner. Enhancing mutations or variations within positive regulatory elements (promoters or enhancers) or weakening alterations to negative regulatory elements (repressors) can result in increased gene expression. Variation within or misuse of enhancer and repressor elements can also contribute to the phenomenon of expanded oncogene expression domain; mutations in enhancers could cause them to take on new functional roles, and translocations can result in an enhancer element inappropriately activating a gene near the chromosomal breakpoint. Another mechanistic way for proto-oncogenes to morph into oncogenes is when modifications to protein structure (mutations or deletions) cause them to become constitutively active.

In the inverse scenario, mutations resulting in a decreased level of gene product are necessary for the oncogenic misexpression of tumor suppressor genes. In order for gene expression to be completely silenced, both copies of a tumor suppressor gene must be inactivated. This can be accomplished through any combination of two genetic changes that cause the complete ablation of gene product from one allele, such as the deletion of a gene or entire chromosomal region, a point mutation or frame shift that yields a null allele, or the hypermethylation of a promoter that silences expression. Some tumor suppressor genes also exert oncogenic effects on a

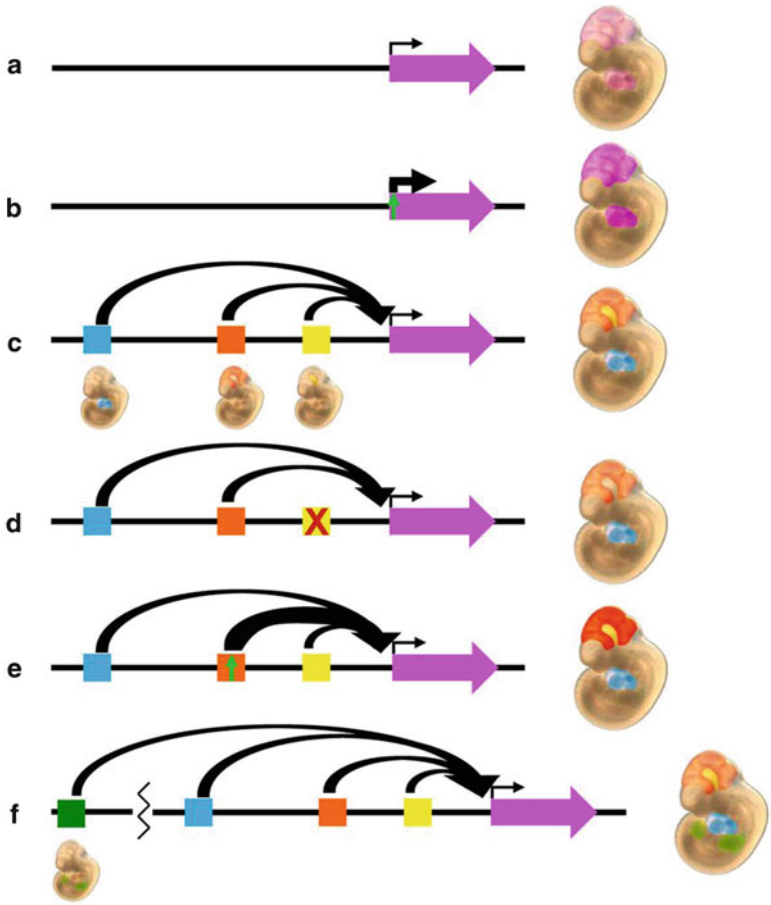

Fig. 10.1 *How cis-regulatory mutations affect gene expression*. (**a**) Endogenous expression pattern of a gene. (**b**) A promoter variant increases overall gene expression levels. (**c**) The long-range enhancer model: three tissue-specific enhancers determine normal gene expression. (**d**) An inactivating mutation in a brain enhancer (*yellow*) results in a reduced expression domain. (**e**) An activating variant in a second brain enhancer (*orange*) results in brain-specific overexpression. (**f**) A translocation juxtaposes a limb enhancer (*green*) into the gene's regulatory landscape, resulting in an expanded expression domain

cell when their expression levels are simply reduced, rather than eliminated. This can be the result of haploinsufficiency – where expression is totally lost from just one allele – or it can be caused by an overall decrease in the amount of transcription from one or both alleles. In the case of decreased expression from a locus, *cis*-regulatory variation in the promoter or long-range enhancer/repressor elements controlling gene expression is often responsible.

In this chapter, we will focus specifically on *cis*-regulatory mutations and common variation underlying cancer etiology or risk. As touched on above, these

cis-regulatory underpinnings to gene misexpression represent just a small subset of known genetic alterations involved in the complexities of cancer biology. In many cases, the same genes have been identified as misexpressed in precancerous or cancerous cells due to a multitude of different mechanisms: a particular tumor suppressor gene that is present in a region frequently deleted in tumors may also be the target of an enhancer element containing common variation that exhibits differential activity in a relevant tissue type. This phenomenon highlights the idea that genes critical to controlling cell proliferation will be focus points for oncogenic mutations, and those mutations may take on many different forms. Many of the more recently discovered examples of *cis*-regulatory changes underlying cancer seem to result in relatively small changes in gene expression levels due to common genetic variation and therefore have relatively small effect sizes. Because of this, most have been discovered in the functional follow-up to GWAS. The case studies presented here will illustrate instances where *cis*-regulatory changes in promoter, enhancer, and repressor elements that function to modify gene expression levels have been implicated in the etiology of cancer risk.

10.2 Promoter Variation

Located directly upstream of their target gene, promoter elements are the easiest of *cis*-regulatory elements to identify (Fig. 10.1b). As the central element involved in controlling gene transcription, their importance and regulatory code have been understood for much longer than long-range *cis*-elements such as enhancers and repressors. As such, countless promoter mutations have been characterized, each altering the expression of a tumor suppressor or proto-oncogene involved in every conceivable type of cancer. Many of these changes – while recurrent in key oncogenic genes – are point mutations unique to a particular individual's tumor. As a whole, they have taught much about tumor biology, but their individual *cis*-regulatory mechanisms of misexpression are not necessarily applicable to a wide range of patients. It has only been with the relatively recent advance of GWAS (Fig. 10.2a) that common variants influencing the regulatory ability of promoters have been identified. Here, we discuss two examples of such GWAS-identified promoter variants, while acknowledging that these represent the very tip of the promoter mutation iceberg.

10.2.1 MSMB and Prostate Cancer Risk

The most straightforwardly interpreted cases of GWAS hits occur when a potentially functional SNP within an ideal functional candidate gene is found to be associated with a disease. Such was the case when two independent GWAS reported an association between SNP rs10993994 on 10q11 and prostate cancer risk

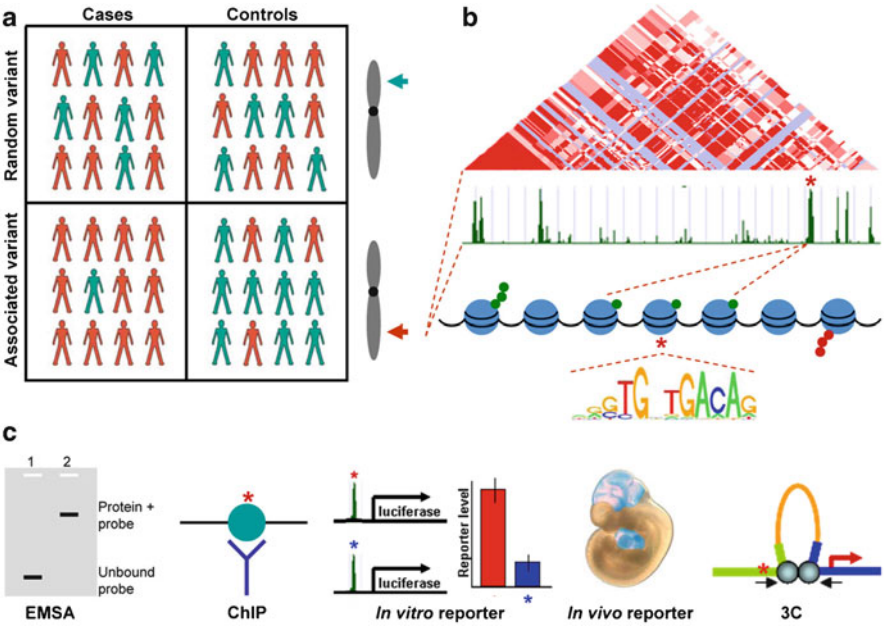

Fig. 10.2 *Strategies to map genetic variation affecting disease traits due to changes in gene expression in human populations*. (**a**) Genome-wide association studies (GWAS) identify genetic variants (SNPs) associated with a disease trait. Differently than most SNPs in the genome, which have similar allele frequencies (*red and green individuals*) in affected (*cases*) and non-affected (*controls*) individuals, an associated variant shows a significant departure from this pattern; in the example shown, there is an overabundance of the "*red*" allele of the associated SNP in cases, compared to "*green*" alleles in controls. (**b**) The associated variant is not necessarily the causal variant underlying the phenotypic difference; rather, multiple SNPs are highly correlated with one another in linkage disequilibrium blocks (LD blocks). Various strategies are used to identify which SNPs (*red asterisk*) within these LD blocks might have a putatively causal role in the phenotype-genotype association. For example, SNPs mapping within evolutionarily conserved noncoding sequences (*green* peaks along the LD block) are good candidates for having a role in phenotypic variation. Further analysis of the genomic context of this candidate SNP can further support the idea that this variant lies within a *cis*-regulatory element, showing, for example, that the local chromatin is compatible with that seen in active *cis*-regulatory elements (single *green balls* on the histones, denoted as *blue balls*). For genome-wide chromatin states in multiple cell lines, see the ENCODE project data at http://genome.ucsc.edu/cgi-bin/hgGateway. More detailed computational analysis may reveal that the SNP lies within a well-defined DNA binding motif for a given transcription factor. This raises the hypothesis that the SNP may alter the binding of proteins to a *cis*-regulatory element, resulting in differential gene expression. (**c**) Multiple experimental strategies can be used to determine that a *cis*-regulatory element controls the expression of a given gene and that a SNP within this regulatory sequence may alter its function. Electromobility shift assays (EMSA) are used to show that a specific protein has the ability to bind to the given stretch of DNA containing the SNP in question (lane 2 of the gel). Chromatin immunoprecipitation (ChIP) detects the binding of a transcription factor to a specific DNA sequence. Reporter assays can be used to test whether a given DNA sequence is a promoter enhancer or silencer and whether an SNP within this element may result in allele-specific functions. These reporter assays can employ in vitro or in vivo experimental models. Chromatin conformation capture (3C) demonstrates long-range interactions in the genome. A putative enhancer (*green*) loops to activate a distant promoter (*blue*) of a gene (*red arrow*). This looping can be captured by cross-linking (*gray balls*) followed by PCR using primers (*black arrows*) for the enhancer and the promoter. PCR amplification using these primers demonstrates that the two distant sequences directly interact, as predicted to occur between enhancers and their distant promoters

(Eeles et al. 2008; Thomas et al. 2008). The SNP is 57 base pairs upstream of the transcriptional start site (TSS) of microseminoprotein beta (*MSMB*), a member of the Ig binding factor family known to be a biomarker for prostate cancer and a suggested prostate cancer tumor suppressor gene (Beke et al. 2007; Reeves et al. 2006). Furthermore, rs10993994 had previously been shown to affect promoter activity levels in embryonic kidney cells (Buckland et al. 2005).

Based on this appealing context, two groups set out to fine map the associated linkage disequilibrium (LD) block (Fig. 10.2b) with the goal of showing that the common variation in the *MSMB* promoter was the underlying reason for the prostate cancer association (Chang et al. 2009; Lou et al. 2009). Using independent populations, both groups determined that the GWAS SNP rs10993994 was most strongly associated with prostate cancer risk. To determine the functional significance of this variant, the *MSMB* promoter region – harboring either the risk (T) or the protective (C) allele of rs10993994 – was cloned into a luciferase vector, and the promoter activity levels were evaluated in prostate cancer cell lines (Fig. 10.2c). Chang et al. found that the promoter element containing the T risk allele drove luciferase expression at 13% compared to the protective C allele in LNCaP prostate cancer cells (Chang et al. 2009); this directionality of affect was expected due to *MSMB*'s status as a tumor suppressor gene. The T risk allele also had decreased promoter activity in PC3 prostate cancer cells as well as in 293T and MCF7 cell lines (Lou et al. 2009).

Once the allele-specific *cis*-regulatory ability of rs10993994 was determined, the question became how the variant exerted its affect on *MSMB* transcriptional activity. As the SNP disrupts a predicted CREB binding site, Lou et al. performed electrophoretic mobility shift assays (EMSA; Fig. 10.2c) on nuclear extracts of a prostate cancer cell line to see whether the differential CREB binding depended on the haplotype (Lou et al. 2009). They showed that CREB bound strongly to the protective T allele of rs10993994, whereas CREB binding was undetectable in the risk allele. This suggests that the prostate cancer risk SNP modulates *MSMB* promoter activity through differential CREB binding (Lou et al. 2009). Strengthening the evidence for rs10993994's role in *MSMB* expression, Lou et al. also showed that cancer cell lines with at least one C allele showed a higher mean *MSMB* mRNA level compared to TT homozygotes (Lou et al. 2009).

To further the link between *MSMB* and prostate cancer tumorigenesis, Pomerantz et al. built on the functional studies and investigated the relationship between rs10993994 and *MSMB* expression in normal prostate and prostate tumor samples (Pomerantz et al. 2010). They determined that rs10993994 genotype correlates with *MSMB* mRNA levels in normal and cancerous human prostate cancer specimens, but not in normal colon or breast tissue. This suggests that rs10993994 shows allele-specific activity in a tissue-specific manner. Furthermore, the authors demonstrated that suppression of *MSMB* in prostate epithelial cells resulted in a significant increase in anchorage-independent colony growth; this affect was not seen in mammary epithelial cells (Pomerantz et al. 2010). Taken together, these results show that the *MSMB* promoter SNP rs10993994 exhibits allele-specific *cis*-regulatory activity, and that its affect on *MSMB* expression appears to be prostate specific, in concordance with its status as a common prostate cancer risk variant.

10.2.2 FOXE1 and Thyroid Cancer Risk

Another example of a promoter *cis*-regulatory variant identified through association studies is the forkhead box E1 (*FOXE1*) variant on chromosome 9q22 that was linked to thyroid cancer risk. First identified in a GWAS (Gudmundsson et al. 2009), variants in *FOXE1* were independently flagged as associated with thyroid cancer in a candidate gene association study (Landa et al. 2009). An ideal candidate gene for misregulation in thyroid cancer, *FOXE1* is at the center of the regulatory network that initiates thyroid differentiation, and increases in FOXE1 expression correlate with dedifferentiation in thyroid carcinomas (Parlato et al. 2004; Sequeira et al. 2001).

Once the thyroid cancer-associated LD block harboring *FOXE1* was located, Landa et al. set about assessing all variants within the interval to prioritize candidate causative SNPs (Landa et al. 2009). Bioinformatic analysis identified SNP rs1867277 – located 283 bases upstream of the *FOXE1* TSS – as disrupting predicted transcription factor binding sites (TFBS); this variant therefore became the lead candidate for functional analysis. In EMSAs performed with the rs1867277 risk or protective allele and nuclear extracts from a thyroid cancer cell line, a lower band was seen forming with both alleles, while an upper band was found only with the A (risk) allele (Landa et al. 2009). After evaluating predicted TFBS, the authors determined that a Kv channel interacting protein 3, calsenilin (KCNIP3; DREAM) antibody supershifted lower EMSA band complex, while an upstream transcription factor (USF) antibody supershifted the A-specific upper band. They therefore concluded that only the risk A allele of SNP rs1867277 is able to bind transcription factors USF1/USF2. While DREAM overexpression has been previously associated with thyroid enlargement (Rivas et al. 2009), an oncogenic role for the ubiquitously expressed USF1/USF2 factors in thyroid cancer has not yet been established. To further understand the role played by DREAM and the USF1/2 transcription factors in *FOXE1* regulation, luciferase reporter constructs containing one of the two *FOXE1* promoter haplotypes were co-transfected into HeLa cells with cDNA plasmids for DREAM or USF1/2 (Landa et al. 2009). While the DREAM co-transfection did not generate variations in promoter activity, co-transfection of the *FOXE1* promoter with USF1/2 yielded an eightfold increase in luciferase expression with the A risk allele, but no change with the G protective variant. These data suggest that the differential binding of USF1/2 to the *cis*-regulatory promoter SNP rs1867277 modulates *FOXE1* expression, explaining the region's association with thyroid cancer risk.

10.3 Common Variation in Long-Range *cis*-Regulatory Elements

Located up to a megabase away from their target gene (Nobrega et al. 2003), long-range *cis*-regulatory elements – such as enhancers and silencers – are functional noncoding elements responsible for controlling tissue- and temporal-specific gene expression.

Many key developmental genes are known to be controlled by an array of enhancers, with each individual *cis*-regulatory element driving a subset of its gene's entire expression profile (Fig. 10.1c). This modular nature makes them ideal candidates for involvement in complex diseases – like cancer – especially, as a functional variant in an individual *cis*-element would result in changes to gene expression levels only in specific organs/tissue types (Fig. 10.1d, e). Less well-characterized are negative *cis*-regulatory elements impacting gene expression; although fewer examples exist, they too are presumed to contain functional variation underlying complex disease etiology. As GWAS routinely implicate variation within gene deserts and other types of noncoding DNA with cancer risk, strategies have been developed for identifying and then characterizing long-range *cis*-regulatory elements potentially harboring cancer-associated variants. The following case studies illustrate examples of successful or in-progress attempts to definitively link noncoding variation with cancer risk.

10.3.1 MYC and the 8q24 Gene Desert Cancer Associations

The best characterized example of *cis*-regulatory variation in long-range enhancer elements underlying cancer risk was found in chromosome 8q24. Numerous GWAS reported associations between multiple types of cancer – including prostate, colorectal, breast, urinary bladder, and chronic lymphocytic leukemia – and variants concentrated within 620 kb of a 1.2-Mb gene desert in this region (Al Olama et al. 2009; Amundadottir et al. 2006; Crowther-Swanepoel et al. 2010; Easton et al. 2007; Ghoussaini et al. 2008; Gudmundsson et al. 2007; Haiman et al. 2007b; Kiemeney et al. 2008; Tomlinson et al. 2007; Turnbull et al. 2010; Yeager et al. 2007; Zanke et al. 2007). Thus far, 14 independent polymorphisms have been associated with various cancers in this region (Grisanzio and Freedman 2010), suggesting that multiple independent functional elements underlie disease risk. Although there are no well-annotated genes within the associated intervals, the independent risk variants (or linked functional elements within the associated regions) may all be involved in regulating the expression pattern of a single gene involved in cancer tumorigenesis and/or progression in various tissue types. The infamous proto-oncogene v-myc myelocytomatosis viral oncogene homolog (*MYC*) lies immediately downstream of this gene desert, raising the possibility that the associated regions of risk harbor long-range *cis*-regulatory elements involved in the tissue-specific transcriptional regulation of *MYC* expression; under this hypothesis, each distinct association interval would harbor a functional noncoding element involved in regulating *MYC* expression in the corresponding tissue type for each implicated cancer. Encoding a well-known transcription factor essential to the regulation of cell proliferation and growth, *MYC* is upregulated at both the mRNA and protein level in each of the 8q24-associated cancers (Chen and Olopade 2008; DeMarzo et al. 2003; Nesbit et al. 1999). Additionally, 8q24 is one of the most common regions for somatic amplification in cancer (Beroukhim et al. 2010). *MYC* misregulation due

to variation within *cis*-regulatory elements would provide yet another path to its oncogenic overexpression.

In the years following the publication of these striking GWAS results, numerous groups using several complimentary methods have shown that the cancer-associated 8q24 risk regions do in fact harbor enhancer elements (Ahmadiyeh et al. 2010; Jia et al. 2009; Pomerantz et al. 2009; Sotelo et al. 2010; Tuupanen et al. 2009; Wasserman et al. 2010; Wright et al. 2010). The most compelling work centers around the cancer risk variant rs6983267, which has independently been associated with prostate and colorectal cancer (Haiman et al. 2007a; Tomlinson et al. 2007; Yeager et al. 2007; Zanke et al. 2007). SNP rs6983267 is not only the actual-typed GWAS variant, but it also disrupts an evolutionarily conserved sequence; this makes it an ideal candidate for functionality. Resequencing and thorough analysis of LD in the cancer-associated region also suggested that rs6983267 itself was the causal risk variant (Yeager et al. 2008). Based on these findings, Pomerantz et al. performed targeted chromatin immunoprecipation (ChIP; Fig. 10.2c) assays on the evolutionary conserved sequence containing rs6983267 with antibodies known to pick out enhancer elements (Pomerantz et al. 2009). These specific epigenetic marks (such as the histone modification H3K4me1) and proteins (like the coactivator p300) have been shown to reliably mark regulatory regions (Heintzman et al. 2007; Visel et al. 2009). Pomerantz et al. found that in the colorectal cancer cell line tested, the rs6983267 element exhibited the classic chromatin signatures for enhancer activity; these findings have since been replicated independently by other groups in both colorectal and prostate cancer cell lines (Ahmadiyeh et al. 2010; Jia et al. 2009; Wright et al. 2010).

While chromatin marks are suggestive of enhancer activity, the regulatory potential of a DNA fragment must be directly assessed using reporter assays. Such experiments ask whether a candidate element is capable of turning on the expression of a reporter gene – usually luciferase for cell-based assays (Fig. 10.2c) or β-galactosidase for in vivo experimentation (Fig. 10.2c) – in the presence of a minimal promoter. The rs6983267-containing element has been shown to exhibit enhancer activity in colorectal (Jia et al. 2009; Pomerantz et al. 2009; Sotelo et al. 2010; Tuupanen et al. 2009) and prostate (Jia et al. 2009; Sotelo et al. 2010) cancer cell lines, as well as in the developing and mature prostate of transgenic mice (Wasserman et al. 2010). Although cell line-based assays are incredibly useful and relevant to the study of misexpression in cancer cells, the full spatial and temporal characterization of an element's endogenous regulatory potential is ideally afforded by in vivo experimentation. It is therefore of particular relevance that the rs6983267-containing enhancer is capable of driving reporter gene expression in the mouse prostate.

If SNP rs6983267 is a *cis*-regulatory modifier of cancer risk, the two alleles would be expected to differentially affect enhancer potential. This allele-specific enhancer activity has in fact been documented in colorectal cancer cell lines (Pomerantz et al. 2009; Tuupanen et al. 2009; Wright et al. 2010) and mouse prostates (Wasserman et al. 2010). In all four cases, the G risk allele was shown to exhibit stronger enhancer activity than the T protective allele in the cancer-relevant cell type. Of note in the in vivo system is the fact that the allele-specific enhancer

potential seemed to be spatially restricted to the prostate and urogenital apparatus; enhancer activity in the genital tubercle and limbs of mouse embryos at embryonic day 14.5 (E14.5) did not exhibit differential activity between the G and T alleles. Given this enhancer's connection to the proto-oncogene *MYC* (detailed below) in prostate and colorectal cancer, the presumed upregulation in the relevant tissue type caused by the presence of the risk variant fits with the model of misexpression needed for oncogenic change.

Once the regulatory potential of the rs6983267-containing element and the allele-specific nature of the SNP itself was determined, the question as to the mechanistic reason for the differential activity was addressed. The cancer risk variant lies within a predicted TCF consensus binding sequence (Pomerantz et al. 2009; Tuupanen et al. 2009). Transcription factor 7-like 2 (*TCF7L2*) is a transcription factor in the Wnt signaling pathway – which is known to target *MYC* – and is activated in most colorectal cancers (Bienz and Clevers 2000; He et al. 1998). Not only was TCF7L2 shown to bind to the rs6983267-containing element in colorectal cancer cell lines, but Pomerantz et al. and Tuupanen et al. both demonstrated allele-specific binding abilities corresponding to the two rs6983267 alleles: TCF7L2 has a higher affinity for the G risk allele and preferentially binds to that haplotype in heterozygous cells (Pomerantz et al. 2009; Tuupanen et al. 2009). It has also been shown that TCF7L2 binds to the rs6983267-containing element in a prostate cancer cell line (Sotelo et al. 2010). These results suggest that the cancer-associated variant mediates risk through differential binding of TCF7L2 to the enhancer element.

The body of work described above convincingly shows that colorectal and prostate cancer-associated SNP rs6983267 is located within an enhancer element and that the SNP confers allele-specific activity to its enhancer through (at least in part) the differential binding of TCF7L2. It does not, however, provide any link – other than circumstantial chromosomal location – between the *cis*-regulatory element and its target gene. In order to definitively associate the enhancer with *MYC*, the ideal candidate gene for misregulation underlying cancer risk, the long-range regulatory element must be shown to physically interact with *MYC*'s promoter. This can be done through the use of the chromosomal conformation capture (3C; Fig. 10.2c) assay, a technique that assesses whether a specific fragment (in this case, the rs6983267-containing element) can loop over large genomic distances to physically connect with another DNA region (such as the *MYC* promoter, approximately 335 kb away) (Dekker et al. 2002). Numerous groups have now demonstrated that the long-range *cis*-regulatory element of interest does in fact interact with *MYC*'s promoter in both colorectal cancer and prostate cancer cell lines, providing very compelling evidence that the rs6983267-containing enhancer is functionally involved in regulating levels of *MYC* expression in these two tissue types (Ahmadiyeh et al. 2010; Pomerantz et al. 2009; Sotelo et al. 2010; Wright et al. 2010). These results provide a crucial link between the *cis*-regulatory risk variant and an infamous proto-oncogene known to be misregulated in the two relevant cancers.

While none of the other 8q24 gene desert risk loci has been as definitively functionally characterized as the LD block harboring the rs6983267-containing element, there is strong evidence for the existence of other long-range tissue-specific *MYC* enhancers within the cancer-associated region boundaries. Two groups have used

chromatin marks to identify candidate regulatory elements located in the different association intervals for cell line-based reporter assay tests, and both reported that several exhibited regulatory potential in the relevant cancer cell line (Jia et al. 2009; Sotelo et al. 2010). In vivo data also exists for a mammary gland enhancer element contained within the breast cancer LD block, but the precise location of the *cis*-regulatory element has not yet been determined (Wasserman et al. 2010). Ahmadiyeh et al. provided additional support for the hypothesis of multiple *MYC* enhancers throughout the 8q24 gene desert by demonstrating that the cancer-associated risk loci physically interact with the *MYC* promoter in a cell type-specific manner. Their 3C results show that the breast cancer locus (but not the prostate or colorectal cancer loci) loops to interact with *MYC* in a breast cancer cell line, and that the multiple prostate cancer loci (but not the breast or colorectal cancer loci) physically interact with *MYC* in a prostate cancer cell line (Ahmadiyeh et al. 2010). Taken together, these observations suggest that each distinct cancer association interval does indeed harbor a functional *cis*-regulatory element involved in modulating *MYC* expression in the corresponding tissue type for each implicated cancer. As has been proven for the rs6983267-containing element, the hypothesis remains that each of the *MYC* enhancers harbors variation that influences *MYC* misregulation and cancer risk.

10.3.2 FGFR2 and Breast Cancer Risk

Another example of *cis*-regulatory variation underlying cancer phenotypes can be seen in the relationship between an intronic region of fibroblast growth factor receptor 2 (*FGFR2*) and breast cancer risk. SNPs within this noncoding LD block exhibited the strongest associations with breast cancer susceptibility in two independent GWAS (Easton et al. 2007; Hunter et al. 2007). Substantiating the strong GWAS association, *FGFR2* – a known breast cancer oncogene – harbors activating missense mutations in some tumors and is somatically amplified in others (Katoh 2008); this makes it an ideal candidate for an additional *cis*-regulatory-driven mechanism of misexpression in breast cancer patients.

Meyer et al. began their inquiries in the locus by determining that *FGFR2* is expressed at higher levels in breast cancer tumors homozygous for the intronic risk alleles than in tumors homozygous for the protective variants (Meyer et al. 2008). They took this correlation as evidence for a *cis*-regulatory variant within the cancer-associated region and focused on identifying differential transcription factor binding abilities for the eight most strongly associated SNPs. EMSA showed that two of the eight candidates' functional SNPs (rs7895676 and rs2981578) displayed an allele-specific binding pattern when assayed with nuclear extracts from a breast cancer cell line. By performing supershift experiments, the authors determined that the protective allele of SNP rs7895676 was binding the CCAAT/enhancer-binding protein beta (C/EBPβ), with the risk allele showing no binding affinity. In the case of SNP rs2981578, only the risk allele was capable of binding the runt-related transcription factor 2 (Runx2) (Meyer et al. 2008). Both C/EBPβ and Runx2 have been previously implicated in breast cancer etiology: C/EBPβ is highly overexpressed in malignant

breast cells (Grigoriadis et al. 2006), and increased Runx2 expression in breast cancer tumors is associated with a more severe clinical outcome (Onodera et al. 2010).

While informative for determining whether DNA-protein complexes are able to form with a given sequence, EMSA cannot establish whether such interactions actually occur within cells. To determine whether the breast cancer risk SNP sites were occupied by the transcription factors of interest in the cellular context, ChIP experiments in breast cancer cell lines homozygous for either the risk or protective haplotype were performed (Meyer et al. 2008). Meyers et al. showed differential binding of Runx2 to SNP rs2981578, with the risk allele binding twice as much protein. For rs7895676, the protective allele was enriched for C/EBP; these results support the EMSA findings. The two variants of both SNPs were tested then for allele-specific regulatory ability in breast cancer cell line luciferase reporter assays. The risk allele of rs2981578 stimulated expression when compared to the protective allele, while rs7895676 showed weaker results in the opposite direction (with the protective allele displaying stronger potential) (Meyer et al. 2008). When the two SNPs were tested together in one haplotype construct – similar to in vivo conditions – the Runx2 SNP prevailed and the risk haplotype showed increased expression. The authors therefore concluded that SNP rs2981578 is likely the functional SNP, as this directionality correlates with increased *FGFR2* expression in tumors harboring risk alleles.

A second study on the same *FGFR2* breast cancer association was performed by Udler et al., using complimentary methods that strengthen the *cis*-regulatory conclusions reached in the previously described work (Udler et al. 2009). Taking advantage of the different haplotype structure present in populations of African descent, the authors fine-mapped the cancer-associated region in African American women and concluded that SNP rs2981578 is most strongly associated with breast cancer risk. They also investigated the chromatin state of the region of interest, reasoning that functional *cis*-regulatory elements must be accessible to transcription factors in order to effectively influence target gene expression. DNase I hypersensitivity assays performed in breast cancer cell lines showed that only two SNPs mapped to open chromatin: rs2981578 was one of them (Udler et al. 2009). As it is also within a region of sequence conservation, they concluded that it is likely to be the functional SNP that is influencing breast cancer risk. Taken together, these two studies provide compelling evidence that SNP rs2981578 lies within an active enhancer element and differentially controls its regulatory potential through allele-specific Runx2 binding. While neither of these studies physically links the rs2981578-containing enhancer element to *FGFR2*, *FGFR2* expression in tumors does correlate with SNP genotype, and it is an ideal functional candidate for *cis*-regulatory oncogenic misregulation in breast cancer.

10.3.3 SMAD7 and Colorectal Cancer Risk

The two previous cases illustrated examples where presumed upregulation of oncogenes due to overactive enhancer element's modulated disease risk. This story represents

the inverse case, where a cancer risk variant decreases the enhancer activity of an apparent tumor suppressor gene. Several GWAS identified colorectal cancer risk variants on 18q21 within a 17-kb LD block in SMAD family member 7 (*SMAD7*) (Broderick et al. 2007; Curtin et al. 2009; Tenesa et al. 2008), an intracellular antagonist of TGF-beta signaling known to influence colorectal cancer progression (Levy and Hill 2006; ten Dijke and Hill 2004). The associated interval spans both exonic and noncoding sequence, but resequencing excluded coding variations (Broderick et al. 2007).

Lower *SMAD7* expression has been shown to be associated with 18q21 risk variants in lymphoblastoid cell lines (LCLs) (Broderick et al. 2007), assuming that the causal variant was therefore asserting its risk effect through *cis*-regulatory means. Pittman et al. resequenced the entire colorectal cancer-associated LD block in a panel of individuals with the goal of identifying all possible variation influencing *SMAD7* expression in the colon (Pittman et al. 2009). The strongest association with disease was provided by a novel SNP dubbed "Novel 1" (rs58920878), which is conserved down to mouse. In vivo *Xenopus* reporter assays performed to determine whether the region surrounding SNP Novel 1 possessed regulatory potential showed GFP expression in the muscle and colorectum of transgenic tadpoles; this strongly suggests that the Novel 1-containing element has enhancer activity (Pittman et al. 2009). Furthermore, the authors demonstrated that the variant confers allele-specific enhancer activity, with the risk allele driving weaker reporter gene expression in the gut compared to the protective haplotype. EMSA results using nuclear extracts from a colorectal cancer cell line revealed the protective allele forming stronger DNA-protein complexes relative to the risk allele, confirming the differential nature of the two alleles (Pittman et al. 2009). The identity of the differentially bound protein remains unknown, and no definitive link has been established between this enhancer element and the presumed target gene *SMAD7*.

10.3.4 EIF3H and Colorectal Cancer Risk

While enhancers and repressors both fall into the category of long-range *cis*-regulatory elements, much more is known about (and many more examples exist of) enhancers. This is largely due to the existence of more developed methodology for identifying and functionally characterizing these positive regulators. One example of variation within a negative regulatory element can be seen in the functional follow-up to several GWAS that identified risk variants for colorectal cancer on 8q23 within a 300-kb region (Houlston et al. 2008; Middeldorp et al. 2009; Tomlinson et al. 2008). After generating a fine-scale map of the region, Pittman et al. determined that a 22-kb block of LD – located 140 kb away from the nearest gene, the eukaryotic translation initiation factor 3, subunit H (*EIF3H*) – showed the highest association with disease (Pittman et al. 2010). Following a similar methodology to the previously described case, they resequenced the associated region in a panel of individuals and prioritized four of the most strongly associated fine-mapped SNPs

(rs16892766, "Novel 28," rs16888589, rs11986063) based on their location within (or flanking) three evolutionally conserved elements. These three conserved elements and their internal/flanking-associated SNPs were cloned and tested for in vivo enhancer activity in *Xenopus*, zebrafish, and mouse reporter gene transgenic assays. To the authors' surprise, none of the elements exhibited enhancer activity (Pittman et al. 2010). Luciferase reporter assays in colorectal cancer cell lines, however, showed that one of the conserved elements – dubbed "island 2" – functioned as an allele-specific repressor: the protective allele A (but not the risk allele G) of SNP rs16888589 repressed luciferase expression below the level seen with the promoter-only reporter construct.

Working on the assumption that the rs16888589-containing repressor element targets the nearest gene *EIF3H*, Pittman et al. conducted experiments aimed at elucidating the effect of differential *EIF3H* expression in colorectal cancer cell lines. They found that knocking down gene expression reduced cell proliferation and colony formation in a soft agar assay, and that overexpressing *EIF3H* increased cell proliferation. This suggests the possible role of a colorectal cancer oncogene for *EIF3H*. To further support its relevance to the functional *cis*-regulatory variant rs16888589, 3C experiments demonstrated that the island 2 repressor physically interacts with the *EIF3H* promoter in colorectal cancer cell lines (Pittman et al. 2010). Taken together, these data imply that the risk G allele of rs16888589 destroys the functionality of its long-range *EIF3H* repressor element, likely increasing *EIF3H* expression and possibly influencing colorectal cancer risk.

10.4 Misuse of Enhancer Elements at Translocation Breakpoints

Translocations are mutations where two nonhomologous chromosomes become joined. Genomic instability – a characteristic of many tumors – results in an increased number of translocations, some of which can have oncogenic effects on cells. These recurrent abnormal karyotypes were among the first genetic alterations to be identified in cancer cells, as they were visible using classic cytogenetic approaches. As technology progressed, it became clear that the specific chromosomal breakpoints of a translocation were key to determining its potential impact of cell growth and differentiation. Some oncogenic translocations join the coding sequence of two different genes, generating a fusion protein capable of promoting tumorigenesis. Others result from the juxtaposition of one gene's regulatory landscape (long-range *cis*-regulatory element/s) with the coding sequence of another gene (Fig. 10.1f). Enhancers are promiscuous elements, capable of interacting with any promoter that enters their range of influence. This promiscuity allows for the improper activation of a gene outside its normal spatial range; this second example falls within the bounds of *cis*-regulatory variation underlying cancer etiology, as it involves the change to a gene's expression pattern due to alterations in its regulatory control.

10.4.1 Immunoglobulin Translocations and Hematologic Cancers

Recurrent translocations between the immunoglobulin (Ig) loci and assorted oncogene-partners are hallmark of many leukemia and lymphoma cancers and a seminal example of aberrant oncogene transactivation due to chromosomal translocation (Nambiar et al. 2008; Willis and Dyer 2000). During normal B cell development, the Ig heavy- and light-chain genes (IgH and IgL) undergo a process of rearrangement to produce a functional surface antigen receptor. These rearrangements are mediated by carefully controlled double-stranded DNA breaks (Kuppers 2005; Willis and Dyer 2000). While the mechanisms vary between cancer types and in many cases the precise pathogenesis of Ig translocations remain unclear, it is thought that many of the oncogenic translocations occur as mistakes during V(D)J recombination or during class-switching recombination (Kuppers 2005). Regardless of their mechanistic origins, these recurrent chromosomal rearrangements result in the juxtaposition of the active Ig *cis*-regulatory landscape and the coding portion of a given proto-oncogene, causing the production of a deregulated constitutively active oncogene in B cells.

The t(14;18)(q32;q21) translocation is the most common chromosomal rearrangement in low-grade lymphomas (Duan et al. 2008). Its consequence is to bring the anti-apoptotic proto-oncogene B cell CLL/lymphoma 2 (*bcl-2*) from chromosome 18q21 to the IgH locus on 14q32, yielding a deregulated and overexpressed *bcl-2* gene. Prolonged cell survival due to this misexpression has been shown to contribute to the development of lymphomas (Desoize 1994). While this common translocation was originally identified using cytogenetic approaches decades ago, work performed during the last several years has been crucial to uncovering the *cis*-regulatory elements and mechanisms through which the IgH regulatory landscape influences *bcl-2* misexpression.

The IgH locus harbors a cluster of long-range enhancer elements (the 3′ IgH enhancers) comprised of four DNase I hypersensitive sites; these elements have been shown to function as a locus control region in B cells (Khamlichi et al. 2000). Direct evidence for the 3′ IgH enhancers' involvement in misregulating *bcl-2* first came from reporter gene assays in cell lines linking the 3′ IgH enhancers directly to the *bcl-2* promoter. These constructs recapitulated the deregulation observed in lymphomas, with the Ig *cis*-elements driving high levels of expression and mimicking a *bcl-2* promoter usage shift seen in vivo (Duan et al. 2007). The enhancer elements are 350 kb away from the translocation breakpoint in vivo, leaving the question of how they mediated *bcl-2* expression still open.

With the advent of 3C technology, Duan et al. asked whether the 3′ IgH enhancers were capable of looping to physically interact with the *bcl-2* promoter in t(14;18)(q32;q21) cells (Duan et al. 2008). Using two lymphoma cell lines – one with the translocation and one without – the authors looked for interactions between probes at the *bcl-2* promoter and those located in and around the 3′ IgH enhancer cluster. They found that the two loci do indeed physically interact in the lymphoma line

harboring the translocation, and that the interaction signal dropped off quickly outside of the enhancer cluster. Furthermore, they demonstrated that treatment with a drug known to decrease *bcl-2* transcription from the translocated locus (trichostatin A) dramatically decreased the IgH enhancer/*bcl-2* promoter interaction as measured by 3C (Duan et al. 2005, 2008). This correlation between 3′IgH enhancer looping and *bcl-2* expression provides strong evidence for the enhancers' direct role in modulating *bcl-2* deregulation.

The gold standard for any functional hypothesis is to create a mouse model that recapitulates the desired phenotype. Xiang et al. were able to do just that by showing that the introduction of the 3′ IgH enhancers into the endogenous mouse *bcl2* locus caused *bcl-2* deregulation and the formation of follicular lymphomas (Xiang et al. 2011). Using mouse embryonic stem (ES) cells, they knocked in the sequence surrounding the 3′ IgH enhancers into the 3′ region of the *bc-l2*, approximately 170 kb downstream of the *bcl-2* promoter. The authors then characterized the mice, demonstrating an increase in B cell-specific *bcl-2* overexpression, extended B cell survival, and a physical interaction between the endogenous *bcl-2* promoter and the knocked-in 3′ IgH enhancers. Finally, they showed that the mice developed B cell lymphomas (Xiang et al. 2011). These results conclusively prove that the 3′ IgH enhancers are the *cis*-regulatory elements functionally responsible for the misregulation of *bcl-2* seen in the t(14;18)(q32;q21) translocation.

10.4.2 TMPRSS2/ETS Transcription Factor Translocations and Prostate Cancer

The oncogenic misexpression of proteins due to translocation is a signature of hematologic cancers, and very few recurrent chromosomal arrangements have been identified in solid tumors (Mitelman 2000). One exception is a translocation commonly seen in prostate cancers that juxtaposes the 5′ untranslated region of the chromosome 21q22.2 gene transmembrane protease serine 2 (*TMPRSS2*) – and all of the *cis*-regulatory elements contained within – with members of the *ETS* transcription factor gene family (Kumar-Sinha et al. 2008). *ETS* transcription factors are key proto-oncogenes involved in the control of cell growth, cell cycle regulation, and apoptosis and are known to be overexpressed in numerous cancers (Hsu et al. 2004). Tomlins et al. first identified this translocation by searching for "outlier" genes characterized by relatively low expression in most prostate cancer microarray profiles but highly overexpressed in a small percent of samples (Tomlins et al. 2005). Two *ETS* family transcription factors, v-ETS erythroblastosis virus E26 oncogene homolog (*ERG*) and ETS variant 1 (*ETV1*), appeared in their analysis. The authors investigated the nature of the *ERG* and *ETV1* overexpression in prostate cancer cell lines and specimens by performing exon-walking qPCR, where the expression level of each exon was interrogated individually. They noted that for both genes, the 5′ exon(s) were expressed at a reduced level compared to the rest of the protein; this suggested the presence of a translocation breakpoint between the

normally expressed exon(s) and the downstream overexpressed neighbors. By using 5′ RNA ligase-mediated rapid amplification of cDNA ends (RACE) technology, they were able to discover that the 5′ exon(s) of *ERG* and *ETV1* had been replaced with the 5′ untranslated region of *TMPRSS2* (Tomlins et al. 2005). These two translocations were confirmed using fluorescence in situ hybridization (FISH), a technique that allows for the visualization of marked chromosomal locations in interphase cell spreads.

TMPRSS2 is a prostate-specific, androgen-responsive gene that is expressed in both normal and neoplasic prostate tissue (Lin et al. 1999). The *ETS* gene translocations result in a fused transcript consisting of the 5′ untranslated first exon of *TMPRSS2* and the *ERG* or *ETV1* gene body; so while this translocation technically creates gene fusion products, there is no actual coding contribution from *TMPRSS2* (Kumar-Sinha et al. 2008). Instead, it is the *TMPRSS2* promoter and other *cis*-regulatory elements contained within the 5′ untranslated region and further upstream that cause the misexpression of the *ERG* or *ETV1* transcripts.

Work in cell lines and transgenic mice suggests that the *ETS* gene overexpression may result in increased invasiveness, suggesting a mechanism through which the translocation could mechanistically influence prostate cancer progression (Kumar-Sinha et al. 2008). *ERG* is the most commonly overexpressed oncogene in prostate cancer (Petrovics et al. 2005), and the *TMPRSS2* translocation was found to be present in 90% of cases exhibiting overexpression of *ERG* or *ETV1* (Tomlins et al. 2005). Therefore, this *cis*-regulatory gene fusion may underlie *ETS* oncogenic overexpression in the majority of prostate cancer cases.

10.5 Summary

Cancer, a disease of uncontrolled cellular proliferation, occurs when the genes normally responsible for regulating cell growth and division become misexpressed and cells gain the ability to bypass crucial cell cycle checkpoints. This overexpression of growth-promoting proto-oncogenes or underexpression of growth-curbing tumor suppressor genes can be caused by a plethora of different genetic mechanisms, and often, the same key genes are subject to a variety of independent alterations. One means of tumorigenic misexpression is through mutations or variations affecting *cis*-regulatory elements. As described here, such *cis*-regulatory changes are involved in the etiology of many different cancers and may help to explain the genetic underpinnings of these complex diseases. Recently, GWAS have been instrumental in identifying common risk variants in noncoding regions; functional follow-ups to these associations have resulted in the characterization of alternations in many *cis*-regulatory elements affecting the expression of nearby tumorigenic genes. Whether in the promoter, long-range elements such as enhancers or silencers or in the overall architecture of a gene's *cis*-regulatory landscape, these mutations and variants have taught us much about the role of noncoding changes to cancer risk and progression.

While these *cis*-regulatory changes can have profound effects on gene expression, they are only one component of tumorigenic gene misexpression. Previously touched upon were other mechanisms that alter DNA sequence or structure: mutations to coding sequence, large-scale deletions or duplications, or translocations that create fusion proteins. Another facet of gene regulation – namely, epigenetic marks and their dynamics – will also prove critical to understanding cancer etiology. While this type of variation has no impact on DNA sequence, it is likely to be at least as crucial as variation in noncoding DNA as a causative agent in tumorigenesis and may help provide a link between the environmental factors known to play a role in cancer risk and actual gene expression changes. It is already well understood that cancer cells have profound methylation changes at many promoters (Esteller 2008; Feinberg and Tycko 2004), and the chromatin marks that help to define active and closed *cis*-regulatory elements and domains will also likely be linked to oncogenic misexpression. Future research will likely uncover the mechanisms linking epigenetics and cancer, enriching our understanding of the full impact *cis*-regulatory alterations have on tumorigenesis.

Abbreviations

3C	Chromosomal conformation capture
bcl-2	B cell CLL/lymphoma 2
C/EBPβ	CCAAT/enhancer-binding protein beta
ChIP	Chromatin immunoprecipitation
E14.5	Mouse embryonic day 14.5
EIF3H	Eukaryotic translation initiation factor 3 subunit H
EMSA	Electrophoretic mobility shift assay
ERG	v-ETS erythroblastosis virus E26 oncogene homolog
ES	Embryonic stem
ETV1	ETS variant 1
FGFR2	Fibroblast growth factor receptor 2
FISH	Fluorescence in situ hybridization
FOXE1	Forkhead box E1
GFP	Green fluorescent protein
GWAS	Genome-wide association studies
Ig	Immunoglobulin
KCNIP3	Kv channel interacting protein 3 calsenilin
LCLs	Lymphoblastoid cell lines
LD	Linkage disequilibrium
MSMB	Microseminoprotein beta
MYC	Proto-oncogene v-myc myelocytomatosis viral oncogene homolog
RACE	RNA ligase-mediated rapid amplification of cDNA ends
RUNX2	Runt-related transcription factor 2
SMAD7	SMAD family member 7

SNP	Single nucleotide polymorphism
TCF7L2	Transcription factor 7-like 2
TMPRSS2	Transmembrane protease serine 2
TSS	Transcriptional start site
USF	Upstream transcription factor

References

Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M et al (2010) 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc Natl Acad Sci USA 107:9742–9746

Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E et al (2009) Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet 41:1058–1060

Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J et al (2006) A common variant associated with prostate cancer in European and African populations. Nat Genet 38:652–658

Beke L, Nuytten M, Van Eynde A, Beullens M, Bollen M (2007) The gene encoding the prostatic tumor suppressor PSP94 is a target for repression by the polycomb group protein EZH2. Oncogene 26:4590–4595

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M et al (2010) The landscape of somatic copy-number alteration across human cancers. Nature 463:899–905

Bienz M, Clevers H (2000) Linking colorectal cancer to Wnt signaling. Cell 103:311–320

Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S et al (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39:1315–1317

Buckland PR, Hoogendoorn B, Coleman SL, Guy CA, Smith SK, O'Donovan MC (2005) Strong bias in the location of functional promoter polymorphisms. Hum Mutat 26:214–223

Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, Smith S, Pruett K, Romero LM, Wiley KE et al (2009) Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. Hum Mol Genet 18:1368–1375

Chen Y, Olopade OI (2008) MYC in breast tumor progression. Expert Rev Anticancer Ther 8:1689–1698

Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, Ruiz-Ponte C, Enjuanes A, Rosenquist R, Carracedo A et al (2010) Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet 42:132–136

Curtin K, Lin WY, George R, Katory M, Shorto J, Cannon-Albright LA, Bishop DT, Cox A, Camp NJ (2009) Meta association of colorectal cancer confirms risk alleles at 8q24 and 18q21. Cancer Epidemiol Biomarkers Prev 18:616–621

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306–1311

DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI (2003) Pathological and molecular aspects of prostate cancer. Lancet 361:955–964

Desoize B (1994) Anticancer drug resistance and inhibition of apoptosis. Anticancer Res 14:2291–2294

Duan H, Heckman CA, Boxer LM (2005) Histone deacetylase inhibitors down-regulate bcl-2 expression and induce apoptosis in t(14;18) lymphomas. Mol Cell Biol 25:1608–1619

Duan H, Heckman CA, Boxer LM (2007) The immunoglobulin heavy-chain gene 3′ enhancers deregulate bcl-2 promoter usage in t(14;18) lymphoma cells. Oncogene 26:2635–2641

Duan H, Xiang H, Ma L, Boxer LM (2008) Functional long-range interactions of the IgH 3′ enhancers with the bcl-2 promoter region in t(14;18) lymphoma cells. Oncogene 27:6720–6728

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447:1087–1093

Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J et al (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40:316–321

Esteller M (2008) Epigenetics in cancer. N Engl J Med 358:1148–1159

Feinberg AP, Tycko B (2004) The history of cancer epigenetics. Nat Rev Cancer 4:143–153

Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, Pooley KA, Ramus SJ, Kjaer SK, Hogdall E et al (2008) Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst 100:962–966

Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ, Jongeneel CV, Valgeirsson H, Fenwick K, Iravani M et al (2006) Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. Breast Cancer Res 8:R56

Grisanzio C, Freedman ML (2010) Chromosome 8q24-Associated Cancers and MYC. Genes Cancer 1:555–559

Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M et al (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nat Genet 41:460–464

Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A et al (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 39:631–637

Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE (2007a) A common genetic risk factor for colorectal and prostate cancer. Nat Genet 39:954–956

Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ et al (2007b) Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 39:638–644

He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW (1998) Identification of c-MYC as a target of the APC pathway. Science 281:1509–1512

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39:311–318

Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S et al (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 40:1426–1435

Hsu T, Trojanowska M, Watson DK (2004) Ets proteins in biological control and cancer. J Cell Biochem 91:896–903

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A et al (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 39:870–874

Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, Reich D, Yan C, Khalid O, Kantoff P, Oh W et al (2009) Functional enhancers at the gene-poor 8q24 cancer-linked locus. PLoS Genet 5:e1000597

Katoh M (2008) Cancer genomics and genetics of FGFR2 (Review). Int J Oncol 33:233–237

Khamlichi AA, Pinaud E, Decourt C, Chauveau C, Cogne M (2000) The 3′ IgH regulatory region: a complex structure in a search for a function. Adv Immunol 75:317–345

Kiemeney LA, Thorlacius S, Sulem P, Geller F, Aben KK, Stacey SN, Gudmundsson J, Jakobsdottir M, Bergthorsson JT, Sigurdsson A et al (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nat Genet 40:1307–1312

Kumar-Sinha C, Tomlins SA, Chinnaiyan AM (2008) Recurrent gene fusions in prostate cancer. Nat Rev Cancer 8:497–511

Kuppers R (2005) Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer 5:251–262
Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, Pita G, Milne R, Maravall J, Ramos I et al (2009) The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet 5:e1000637
Levy L, Hill CS (2006) Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. Cytokine Growth Factor Rev 17:41–58
Lin B, Ferguson C, White JT, Wang S, Vessella R, True LD, Hood L, Nelson PS (1999) Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 59:4180–4184
Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, Yu K, Hutchinson A, Jacobs KB, Kraft P et al (2009) Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc Natl Acad Sci USA 106:7933–7938
Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C, Ponder BA (2008) Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol 6:e108
Middeldorp A, Jagmohan-Changur S, van Eijk R, Tops C, Devilee P, Vasen HF, Hes FJ, Houlston R, Tomlinson I, Houwing-Duistermaat JJ et al (2009) Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 18:3062–3067
Mitelman F (2000) Recurrent chromosome aberrations in cancer. Mutat Res 462:247–253
Nambiar M, Kari V, Raghavan SC (2008) Chromosomal translocations in cancer. Biochim Biophys Acta 1786:139–152
Nesbit CE, Tersak JM, Prochownik EV (1999) MYC oncogenes and human neoplastic disease. Oncogene 18:3004–3016
Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) Scanning human gene deserts for long-range enhancers. Science 302:413
Onodera Y, Miki Y, Suzuki T, Takagi K, Akahira J, Sakyu T, Watanabe M, Inoue S, Ishida T, Ohuchi N et al (2010) Runx2 in human breast carcinoma: its potential roles in cancer progression. Cancer Sci 101:2670–2675
Parlato R, Rosica A, Rodriguez-Mallon A, Affuso A, Postiglione MP, Arra C, Mansouri A, Kimura S, Di Lauro R, De Felice M (2004) An integrated regulatory network controlling survival and migration in thyroid organogenesis. Dev Biol 276:464–475
Petrovics G, Liu A, Shaheduzzaman S, Furusato B, Sun C, Chen Y, Nau M, Ravindranath L, Dobi A, Srikantan V et al (2005) Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene 24:3847–3852
Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B, Lloyd A, Vijayakrishnan J, Qureshi M, Broderick P et al (2010) Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. PLoS Genet 6(9):E1001126
Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P et al (2009) The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res 19:987–993
Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M et al (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 41:882–884
Pomerantz MM, Shrestha Y, Flavin RJ, Regan MM, Penney KL, Mucci LA, Stampfer MJ, Hunter DJ, Chanock SJ, Schafer EJ et al (2010) Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. PLoS Genet 6:e1001204
Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM (2006) Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. Clin Cancer Res 12:6018–6022
Rivas M, Mellstrom B, Torres B, Cali G, Ferrara AM, Terracciano D, Zannini M, Morreale de Escobar G, Naranjo JR (2009) The DREAM protein is associated with thyroid enlargement and nodular development. Mol Endocrinol 23:862–870

Sequeira MJ, Morgan JM, Fuhrer D, Wheeler MH, Jasani B, Ludgate M (2001) Thyroid transcription factor-2 gene expression in benign and malignant thyroid lesions. Thyroid 11:995–1001

Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJ, Munroe DJ, Wu X (2010) Long-range enhancers on 8q24 regulate c-Myc. Proc Natl Acad Sci USA 107:3001–3005

ten Dijke P, Hill CS (2004) New insights into TGF-beta-Smad signalling. Trends Biochem Sci 29:265–273

Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N et al (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 40:631–637

Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A et al (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40:310–315

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R et al (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644–648

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W et al (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39:984–988

Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K et al (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40:623–630

Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS et al (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42:504–507

Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymaki I et al (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet 41:885–890

Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, Doody DR, MacArthur S, Tyrer J, Pharoah PD et al (2009) FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 18:1692–1703

Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. Nature 461:199–205

Wasserman NF, Aneas I, Nobrega MA (2010) An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res 20:1191–1197

Willis TG, Dyer MJ (2000) The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. Blood 96:808–822

Wright JB, Brown SJ, Cole MD (2010) Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol Cell Biol 30:1411–1420

Xiang H, Noonan EJ, Wang J, Duan H, Ma L, Michie S, Boxer LM (2011) The immunoglobulin heavy chain gene 3′ enhancers induce Bcl2 deregulation and lymphomagenesis in murine B cells. Leukemia 2011:24

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N et al (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39:645–649

Yeager M, Xiao N, Hayes RB, Bouffard P, Desany B, Burdett L, Orr N, Matthews C, Qi L, Crenshaw A et al (2008) Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. Hum Genet 124:161–170

Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E et al (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39:989–994

Chapter 11
Cohesin and Human Diseases

Dongbin Xu and Ian D. Krantz

Abstract Cohesin is a four-protein complex capable of tethering sister chromatid strands together. With the help of multiple facilitating proteins, cohesin plays essential cellular functions in sister chromatid cohesion during mitosis and meiosis, DNA repair, gene expression, and maintaining 3-D genome organization. Cohesin is required for cell division, maintaining pluripotency of stem cells and ensuring normal organ development. Defective cohesin genes have been associated with several rare human developmental disorders including Cornelia de Lange syndrome and Roberts/SC phocomelia syndrome, as well as several malignancies. Here, we provide an overview of cohesion biology and our current understanding as to how cohesin defects cause human disorders.

Keywords Cohesin • *NIPBL* • *SMC1A* • *SMC3* • Cornelia de Lange syndrome • ATRX • Roberts syndrome • Mitosis • Meiosis • DNA double-strand breaks • CTCF • Warsaw breakage syndrome • Malignancies

11.1 Overview of Cohesion

In order to maintain genomic stability and integrity, cells need to ensure precise separation of identical chromosome sets into daughter cells during cell division. During mitosis and meiosis II, sister chromatid cohesion (SCC), tight alignment of sister chromatids, is indispensible to establish attachment of sister chromatids to bipolar metaphase spindles and thus to allow equal separation of sister chromatids upon release of cohesion. In contrast, SCC ensures sister chromatids are tethered together and distributed appropriately during meiosis I. The protein complex required

D. Xu • I.D. Krantz (✉)
Division of Human Genetics, The Children's Hospital of Philadelphia,
3615 Civic Center Boulevard, Philadelphia, PA 19104, USA
e-mail: XUD@email.chop.edu; ian2@mail.med.upenn.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_11, © Springer Science+Business Media, LLC 2012

for cohesion is called cohesin, an SMC (structural maintenance of chromosome) complex in eukaryotic cells. The other two types of SMC complexes are condensin and the Smc5/Smc6 complex (Losada and Hirano 2005). Cohesin includes four core protein components: Smc1, Smc3, Scc1 (Rad21), and Scc3 (stromalin/SA1/SA2/STAG3). The four-protein complex assembles on chromosomes and stably associates with DNA in cells from G1 to metaphase (Gerlich et al. 2006; McNairn and Gerton 2009). Mitotic and meiotic cohesion requires a different protein complex but with similarities to the cohesin complex (Table 11.1). In addition to these four core proteins, several facilitating proteins such as Scc2 (NIPBL), Scc4 (Mau-2), ECO1, Wapal, Pds5, separase, and securin are required for dynamic loading and establishment or stabilization of the cohesin complex through cell cycle progression (Feeney et al. 2010). Beyond its canonical role in sister chromatid cohesion, more recently cohesin has been found to play important roles in repairing double-strand DNA breaks, regulating gene expression, and maintaining higher-order chromatin structure. Involvement of cohesin in these essential molecular functions determines its indispensible roles in cell proliferation, maintaining stem cell pluripotency and ensuring normal organism development. Defects in cohesin genes have been associated with human diseases such as Cornelia de Lange syndrome (Krantz et al. 2004; Tonkin et al. 2004) and Roberts/SC phocomelia syndrome (Schule et al. 2005; Vega et al. 2005), and cohesion defects have been implicated in other disorders including α-thalassemia/mental retardation syndrome, X-linked (ATRX) (Gibbons et al. 1995; Ritchie et al. 2008) and the Warsaw breakage syndrome (WBS) (van der Lelij et al. 2010). Additionally, several types of cancer have been found to have mutations in cohesin-associated structural and regulatory genes (see below for details).

11.2 Cohesin Subunits

11.2.1 Core Cohesin Components

It was initially thought that the DNA strands of the sister chromatids were intertwined with each other, and it was the DNA catenation that kept sister chromatids tethered together during metaphase (Murray and Szostak 1985). A later study found that intertwined DNA strands had been decatenated by topoisomerase II before attachment of sister chromatids to the bipolar spindle poles (Koshland and Hartwell 1987), indicating that other factors are responsible for holding chromatids together. Following this study, several essential cohesin genes such as *Smc1*, *Smc3*, and *Scc1* (*MCD1*) were identified from genetic studies performed in the yeast, *S. cerevisiae* (Michaelis et al. 1997; Guacci et al. 1997). This discovery was followed by the identification of several more cohesin and cohesin-associated genes from various species. All these cohesin structural proteins and associated facilitating components are highly conserved across species from single cell organisms (yeast) to complex organisms (humans). Although mitotic and meiotic cohesin complexes are not

Table 11.1 Cohesin genes in various species. (**a**) Genes encoding cohesin core components that are mitosis specific are highlighted in *blue*, and meiosis specific are highlighted in *yellow*. Modified from Nasmyth and Haering (2009). (**b**) Genes encoding cohesin regulatory proteins

a

Species	SMC1		SMC3	SCC1		SCC3	
	Mitosis	Meiosis		Mitosis	Meiosis	Mitosis	Meiosis
S. cerevisiae	*SMC1*		*SMC3*	*SCC1=MCD1*	*REC8*	*SCC3=IRR1*	
S. pombe	*psm1*		*psm3*	*rad21*	*rec8*	*psc3*	*rec11*
Caenorhabditis elegans	*him-1*		*smc-3*	*scc-1* *coh-1*	*rec-8* *coh-3*	*scc-3*	
Drosophila melanogaster	*SMC1*		*Cap*	*Rad21=vtd*	*c(2)M*	*SA* *SA-2*	
Danio rerio	*smc1a*	*smc1b*	*smc3*	*rad21*	*Zgc:136888*	*LOC564533* *LOC563669* *wu:fc17g12*	
Xenopus laevis	*smc1a*	*smc1b*	*smc3*	*rad21*	*rec8*	*stag1* *stag2*	
Mus musculus	*Smc1a*	*Smc1b*	*Smc3*	*Rad21*	*Rec8* *Rad21L*	*Stag1* *Stag2*	*Stag3*
Homo sapiens	*SMC1A*	*SMC1B*	*SMC3*	*RAD21*	*REC8*	*STAG1* *STAG2*	*STAG3*

b

Species	Cohesin Loading		Cohesion Establishment	Anti-establishent of Cohesion		Cleavage of Scc1	Protection of Separase cleavage
	SCC2	SCC4	ESCO	Wapal	Pds5	Separase	Securin
S. cerevisiae	*SCC2*	*SCC4*	*ECO1*	*RAD61*	*PDS5*	*ESP1*	*PDS1*
S. pombe	*mis4*	*ssl3*	*eso1*	*wpl1*	*pds5*	*cut1*	*cut2*
Caenorhabditis elegans	*pqn-85*	*mau-2*	*F08F8.4*	*wapl-1*	*evl-14*	*sep-1*	*ify-1*
Drosophila melanogaster	*Nipped-B*	*CG4203*	*San* *eco*	*wapl*	*pds5*	*Sse thr*	*pim*
Danio rerio	*Nipblb*	*LOC794129*	*LOC100333282*	*KIAA0261*	*pds5a* *pds5b*	*espl1*	*pttg1*
Xenopus laevis	*Nipbl*	*kiaa0892*	*esco1*	*Wapal*	*pds5a* *pds5b*	*espl1*	*LOC398156*
Mus musculus	*Nipbl*	*Mau2*	*Esco1*	*Wapal*	*pds5a* *pds5b*	*Espl1*	*Pttg1*
Homo sapiens	*NIPBL*	*MAU2*	*ESCO1* *ESCO2*	*WAPAL*	*PDS5A* *PDS5B*	*ESPL1*	*PTTG1*

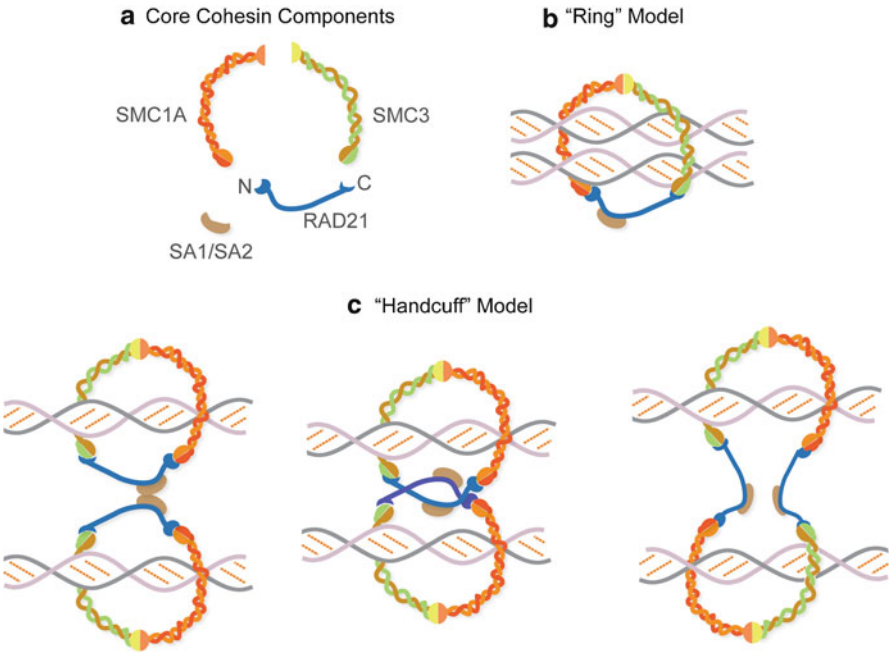

Fig. 11.1 *Cohesin structure and models for interaction with DNA.* (**a**) SMC1, SMC3, RAD21 and SA1/SA2 are cohesin core components, being able to form a ring-shaped structure. SMC1 and SMC3 are depicted as a rod shape of a long coiled-coil capped with a globular hinge domain and an ATP-binding head domain at both ends. *N*- and *C*-termini of RAD21 bind to ATP head domains of SMC3 and SMC1A, respectively. SA1/SA2 binds with RAD21 at its cleavage site by separase. (**b**) The "Ring" model. A single cohesin ring embraces the two sister chromatids together. (**c**) The "Handcuff" model. Each cohesin ring structure embraces a single chromatid and two rings associate to effect cohesion of the sister chromatids

identical, both contain two SMC subunits (e.g., SMC1 and SMC3 in budding yeast), a kleisin subunit (e.g., SCC1 and REC8 in budding yeast) which functions to bridge the two SMC proteins and a HEAT repeat containing subunit (e.g., SCC3 in budding yeast) (Table 11.1). HEAT domains were initially found in four eukaryotic proteins: *H*untingtin, *E*longation factor 3, protein phosphatase 2*A*, and the yeast PI3-kinase *T*OR1 (Andrade and Bork 1995), and consist of repeats of two antiparallel alpha-helices and two turns that form around a common axis.

SMC1 and SMC3 have similar protein architecture (Fig. 11.1a). Generally, an SMC protein is mainly an alpha-helix peptide with two nucleotide-binding motifs (called Walker A and Walker B motifs) at both ends and a globular hinge domain lying in the middle of the alpha-helix peptide. Folding at the hinge domain brings the two halves of the alpha-helix peptides together to form a long, antiparallel coiled-coil domain (Melby et al. 1998). Thus, the Walker A and Walker B motifs associate together to form an ATP-binding head domain. Overall, an SMC protein displays a dumbbell-shaped architecture with a rod shape of long coiled-coil capped

with a globular hinge domain and an ATP-binding head domain at both ends (Fig. 11.1a). SMC1 and SMC3 form a V-shaped heterodimer through a stable hydrophobic interaction between two hinge domains. SCC1 (MCD1/Rad21) serves to link ATPase domains in SMC complexes (Schleiffer et al. 2003). The N- and C-termini of SCC1 associate with head domains of SMC3 and SMC1, respectively. Although the SMC1 and SMC3 heterodimer can bind with DNA by itself in vitro, addition of SCC1 dramatically increases the binding affinity with DNA, especially with cruciform DNA (Sakai et al. 2003). These biochemical interactions among SMC1, SMC3, and SCC1 support the role of these three proteins which is to form a tripartite ring-like structure to encircle sister chromatids together topologically (Fig. 11.1b). Electron micrographs revealed that the antiparallel coiled-coil arm of the SMC1/SMC3 dimer is about 50 nm in length and can form a ring structure with SCC1 with a diameter of 30–35 nm which is large enough to embrace 10-nm DNA fibers (Haering et al. 2002). SCC1 is cleaved by a cysteine protease called separase/Esp1 (Uhlmann et al. 1999). It was shown that cohesin is released from DNA upon cleavage of SCC1 or SMC3 (Uhlmann et al. 2000). This further supports that cohesin associates with DNA topologically rather than by physical binding (Gruber et al. 2003). It remains uncertain how the ring structure embraces sister chromatids although several models have been proposed (see below for details).

SCC3 is a HEAT repeat protein that directly binds with SCC1 to complete the core cohesin complex. The binding site of SCC3 to SCC1 appears to occur within the separase cleavage region. SA1, SA2, and STAG3 are mammalian homologues of SCC3 with distinctive functions. It was suggested that SA1 is essential for telomere cohesion and SA2 is required for centromere cohesion (Canudas and Smith 2009). STAG3 is a germinal-cell-specific protein that functions during meiotic cohesion (Prieto et al. 2001).

11.2.2 Cohesin Facilitating Factors

The distribution of cohesin on the chromosomes is highly dynamic in different cell cycle phases and in different developing tissues. In addition to the aforementioned four core components, several cohesin auxiliary factors, e.g., SCC2 (Nipped-B/NIPBL), SCC4 (MAU2), ECO1 (Eco/ESCO1/ESCO2), Pds5 (PDS5A/PDS5B), Rad61 (Wpl/Wapl/WAPAL), ESP1 (separase), and PDS1 (securin), play indispensible roles in the spatial and temporal regulation of loading, establishment, and protecting cohesin (Table 11.1).

SCC2 is a HEAT repeat-containing protein (Neuwald and Hirano 2000). SCC2 and SCC4 form a tight stoichiometric complex. The SCC2/SCC4 complex (NIPBL/MAU2 complex in humans) as well as the ATPase activity of SMC1/SMC3 is required for loading cohesin onto the chromatids (Ciosk et al. 2000). The exact mechanism of how the Scc2/Scc4 complex loads cohesion onto DNA remains unclear. It has been observed that only a small part of the SCC2/SCC4 complex associates with cohesin (Arumugam et al. 2003) and the SCC2/SCC4 complex is

not required for maintenance of cohesion (Ciosk et al. 2000; Bernard et al. 2006), supporting the hypothesis that the SCC2/SCC4 complex transiently associates with cohesin and facilitates cohesin loading by adjusting the open/close status of the cohesin rings. Nipped-B and NIPBL were found to be the homologues of SCC2 in *Drosophila* and humans, respectively (Krantz et al. 2004; Tonkin et al. 2004; Rollins et al. 1999). *Nipped-B* was identified from a genetic screen to find genes mediating long-range interaction between distant enhancer and promoter regions of a homeobox gene called *cut* (Rollins et al. 1999), indicating a novel function of the cohesin complex in regulating gene expression. Metazoans are very sensitive to dosage of Nipped-B/NIPBL. Reduction of 30% of NIPBL in humans with heterozygous *NIPBL* mutation causes a severe multisystem developmental disorder called Cornelia de Lange syndrome (CdLS) (Borck et al. 2006). The observation that sister chromatid separation and cell division are not severely affected in CdLS patient cells added supporting evidence to a new role for cohesin in gene expression and development (see below for details). It was observed by fluorescence recovery after photobleaching (FRAP) in *Drosophila* salivary glands that cohesin binds the chromosomes with a weak (20 s duration) and a stable (340 s duration) mode. Decreasing Nipped-B transcript levels to about 30% of normal levels in a heterozygous fly mutant result in one third reduction in the amount of stable cohesin binding to the chromosome (Gause et al. 2010). This is consistent with the function of Nipped-B as a cohesin loader and the dosage sensitivity of Nipped-B/NIPBL in fly and humans observed from the aforementioned studies.

Another HEAT repeat protein, precocious dissociation of sisters 5 (PDS5), as well as Wings apart-like 1 (Wapl1), presents a weak association with tripartite cohesion rings (Neuwald and Hirano 2000; Panizza et al. 2000; Rowland et al. 2009). PDS5 and Wapl1 form a complex to inhibit cohesion establishment. This antiestablishment effect can be antagonized by acetylation of Smc3 by ECO1while DNA is replicating (Gandhi et al. 2006; Kueng et al. 2006). Pds5 is important for animal development based on the severe defects observed in the *Pds5* knockout mouse. *Pds5* homozygous knockout mice are lethal after birth and have features that overlap with those seen in CdLS such as developmental delay, congenital heart defects, and limb defects (Zhang et al. 2007, 2009).

In yeast, ESP1 (separase) is required for the release of cohesin from chromosomes when cells transit from metaphase to anaphase. After sister chromatids are attached to the mitotic spindle apparatus and aligned at the equatorial plate, the E3 ubiquitin ligase APC/C (anaphase-promoting complex/cyclosome) is activated, and it degrades securin (PDS1), a protector of cohesin which binds with separase. Thus, separase is released from securin's inhibitory binding and actively cleaves its target SCC1 resulting in the release of cohesin from the chromatids (Hauf et al. 2001; Uhlmann 2001). However, in human cells, the protease-dependent cleavage of RAD21 is only responsible for releasing a small part of cohesin from the pericentromeric region in anaphase. Cohesin that is localized to the chromosome arms is removed by a separase-independent approach (Hauf et al. 2001; Waizenegger et al. 2000; Hauf et al. 2005) (see below for details).

11.3 Cohesins in Mitosis and Meiosis

From the introduction above, we can see that different species contain similar but not identical cohesin subunits. Moreover, some cohesin components are mitotic or meiotic specific (Table 11.1). For instance, most species contain two SCC1 homologues. One forms mitotic cohesin and another one belongs to meiotic cohesin. Human and *Xenopus* have SMC1A that is active during mitosis and SMC1B that is active during meiosis. STAG3 is a human homologue specific for meiotic cohesion, and its counterparts for mitotic cohesion are SA1 and SA2 (Nasmyth and Haering 2009). Rad21L, a paralog of Rad21, is a recently identified meiosis-specific cohesin component (Ishiguro et al. 2011; Lee and Hirano 2011; Gutierrez-Caballero et al. 2011). It is conceivable that meiosis and mitosis require different cohesin complexes with distinct SCC1 homologues because, during mitosis, SCC1 needs to be cleaved leading to separation of sister chromatids in anaphase, while during meiosis I, sister chromatids need to be tethered tightly and not be pulled apart by the bipolar spindle microtubule fibers. Differences are also observed for cohesin in different chromosomal regions. It has been reported that cohesin is removed from chromosome arms during meiosis I and remains in the centromeric regions until metaphase of meiosis II (Klein et al. 1999; Watanabe and Nurse 1999). It has also been demonstrated that SA1/STAG1 and SA2/STAG2 are more active in telomere and centromere cohesin, respectively (Canudas and Smith 2009), demonstrating distinctive activities and regulation of centromeric and telomeric cohesion.

11.4 Interaction Between Cohesin and Chromatin

11.4.1 Models of Cohesin Binding with Sister Chromatids

Several models have been proposed to illustrate how the cohesin complex holds sister chromatids together including a "single ring" model and several types of "handcuff" models (Fig. 11.1b, c). The "ring" model proposes that a tripartite ring structure consisting of SMC1, SMC3, and SCC1 encircles both sister chromatids together (Fig. 11.1b). The ring model is supported by several lines of evidence. First, cleavage of cohesin components or linearization of DNA causes release of cohesin from DNA (Uhlmann et al. 2000; Ivanov and Nasmyth 2007). Second, different cohesin complex units do not associate with each other based on results from co-immunoprecipitation (Haering et al. 2002) and fluorescence resonance energy transfer (FRET) experiments (Mc Intyre et al. 2007). Lastly, cross-linking of a tripartite structure consisting of SMC1, SMC3, and SCC1 produces minichromosome dimers which are resistant to protein denaturation by the detergent SDS (Haering et al. 2008). This latter study further supported the single ring model by showing that the fraction of DNA dimers is equivalent to the fraction of cross-linked cohesin rings and 50% of DNA dimers survive after cleaving half of the rings. However, it has also been shown that cohesin subunits SMC1, SMC3, and Rad21/SCC1 interact

with themselves in an SA/SCC3-dependent manner, supporting the "handcuff" model which proposes that a cohesin ring structure encircles a single DNA strand and two rings associate or interconnect each other (Zhang et al. 2008a) (Fig. 11.1c). This model is also supported, at least for some heterochromatic regions, by the finding that the larger diameter of heterochromatin regions would seem incapable of fitting two DNA fibers in a single cohesin ring as proposed by the "ring" model (Chang et al. 2005). Nevertheless, the common point of these models is that they all demonstrate that the cohesin complex tethers sister chromatids together through topological interactions instead of physical binding.

11.4.2 Cohesin Loading

The association of cohesin with chromatids starts at variable cell cycle phases in different species. Cohesin associates with chromatids after nuclear envelope formation in telophase in mammalian cells (Darwiche et al. 1999) and at the end of G1 phase in yeast (Michaelis et al. 1997; Guacci et al. 1997; Lengronne et al. 2004). ATP hydrolysis and the SCC2/SCC4 (NIPBL/MAU2) complex are required for loading of cohesin on to chromatids (Fig. 11.2a). It is not clear if opening of the hinge domain of the SMC dimer or transient removal of SCC1/Rad21 is the mechanism by which cohesin rings entrap DNA. Studies in yeast have shown that SCC2 does not co-localize with cohesin on the chromosomes and cohesin relocates to convergent transcription regions on the chromosomes after loading (Lengronne et al. 2004). Interestingly, studies using *Drosophila* cells revealed a different pattern. *Drosophila* Nipped-B and cohesin co-localize throughout the genome, and they preferentially bind to actively transcribed regions (Misulovin et al. 2008). It is not clear why this difference between species exists. However, overlapping of Nipped-B and cohesin with RNA polymerase II in *Drosophila* supports a function of Nipped-B and cohesin in regulating gene expression, a likely mechanistic role that leads to CdLS when cohesin regulation or function is disrupted.

As noted above, cohesin rings are loaded onto DNA fibers through the action of the SCC2/SCC4 complex and ATP hydrolysis by the SMC subunits prior to DNA replication. De novo loading of cohesin also occurs after DNA replication when double-stranded break-induced DNA repair occurs (see below for details). Preloaded cohesin becomes cohesive during DNA replication. Establishment of sister chromatid cohesion (SCC) generally occurs along with PCNA (proliferating cell nuclear antigen)-dependent DNA replication in S phase (Moldovan et al. 2006) and is dependent on ECO1 (ESCO1/ESCO2), an acetyltransferase (Moldovan et al. 2006; Toth et al. 1999). Thus, cohesion generation and DNA replication fork progression are closely associated processes. It was noted that some cohesion is generated at replication forks (Lengronne et al. 2006). ECO1 has been found to be a critical factor during SCC establishment and DNA replication. ECO1 acetylates SMC3 at two conserved lysine sites (K112 and K113 in yeast SMC3) in the ATPase domain, and the acetylation only occurs at the onset of S phase. Several studies have demonstrated that SMC3

acetylation by ECO1 is required for cohesion establishment (Rolef Ben-Shahar et al. 2008; Unal et al. 2008; Zhang et al. 2008b) (Fig. 11.2a). Given the essential function in cohesion generation, ECO1 has been found to physically interact with several other proteins involved in both SCC establishment and DNA replication including chromosome transmission fidelity 18 (Ctf18), enhanced level of genomic instability 1 (Elg1) (Kenna and Skibbens 2003; Parnas et al. 2009), PCNA, and chromosome loss 1 (Chl1) (Moldovan et al. 2006; Mayer et al. 2001, 2004; Inoue et al. 2007).

In contrast with the function of Eco1 in establishing cohesion, PDS5 and RAD61, two cohesin-associated proteins, form a complex and manifest an antiestablishment function (Rowland et al. 2009; Sutani et al. 2009; Peters and Bhaskara 2009) (Fig. 11.2a). Acetylation of SMC3 by ECO1 can act against this antiestablishment effect, probably by changing the ATPase activity of SMC3 and affecting interactions among SCC3, PDS5, and RAD61 (Heidinger-Pauli et al. 2010). Given the function of Wpl1 in establishing cohesion, loss of its homologue in lower organisms and higher organisms (e.g., *RAD61* in *S. cerevisiae* and *WAPAL* in humans) surprisingly results in an opposite effect. Depletion of *WAPAL* in human cells results in increased cohesin binding to DNA (Gandhi et al. 2006). However, mutations in *RAD61* result in decreased cohesin binding (Sutani et al. 2009). The reason for these distinct phenotypes remains unclear.

11.4.3 Cohesin Dissolvement

Dissolving cohesin before cytokinesis is as equivalently important as establishing cohesion for ensuring normal cell division. Cohesin needs to be dissolved to allow chromosome separation. As previously noted, the APC/C (anaphase-promoting complex or cyclosome) and the separase pathway is required for removing cohesin in yeast. However, in metazoans, more than 90% of cohesin is released through a non-proteolytic process (also known as the prophase pathway) which is promoted by phosphorylation of RAD21 and SA2 by a polo-like kinase and aurora B. PDS5 and WAPAL are shown to be required for the releasing process, possibly though altering the ring conformation (Gandhi et al. 2006; Kueng et al. 2006; Shintomi and Hirano 2009, 2010) (Fig. 11.2a). Shugoshin (Sgo) and protein phosphatase type A (PP2A) inhibit phosphorylation of SA and prevent the remaining 10% of cohesin, which assembles in the pericentromeric region, to be dissolved from the chromosomes until the onset of anaphase. The pericentromeric cohesins are released in anaphase after cleavage of RAD21 by separase (Oliveira et al. 2010). In short, in lower organisms, most of cohesin remains bound to the DNA until metaphase and is released after SCC1 is cleaved by separase upon APC/C activation. In metazoans, cohesin is dissolved in a two-step process in which cohesin on the chromosome arms is removed by a separase-independent process in prophase and subsequently cohesin in the pericentromeric regions is released in anaphase with RAD21 being cleaved by separase. These processes are tightly regulated by controlling phosphorylation of cohesin components by polo-like kinases and aurora B.

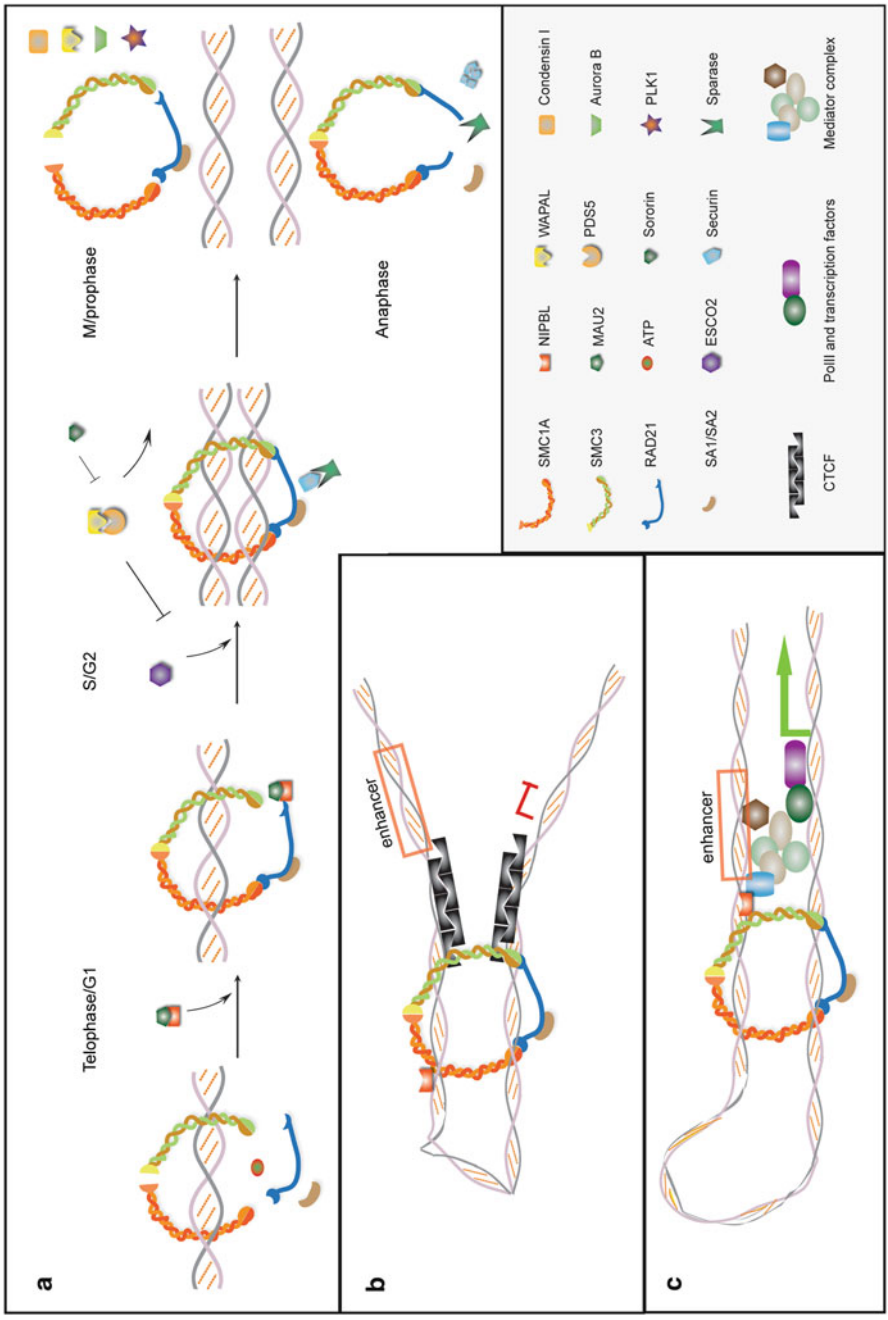
a
Telophase/G1
S/G2
M/prophase
Anaphase
b
enhancer
c
enhancer
SMC1A
SMC3
RAD21
SA1/SA2
NIPBL
MAU2
ATP
ESCO2
WAPAL
PDS5
Sororin
Securin
Condensin I
Aurora B
PLK1
Sparase
CTCF
PolII and transcription factors
Mediator complex

Fig. 11.2 *Cohesin loading and releasing and models of how cohesin regulates gene expression through the organization of 3-D chromatin architecture.* (**a**) Cohesin components are loaded onto chromatin during telophase or G1 phase. The cohesin loading requires the NIPBL/MAU2 complex and the ATPase activity of SMC proteins. Cohesion establishment is coupled with DNA replication in S phase. To promote cohesion establishment, ESCO2 acetylates SMC3 and antagonizes the inhibitory effect of the PDS5/WAPAL complex. Sororin maintains cohesion by displacing WAPAL from the PDS5/WAPAL complex. Securin prevents cleavage of RAD21 by separase through binding with separase. The majority of cohesin is removed through the prophase pathway which is regulated by polo-like kinase 1 (PLK1), aurora B kinase, and condensin I. PDS5 and WAPAL complex also promotes this release process. Cohesin in the centromeric region and residual cohesin on the chromosome arms are removed through a proteolysis-dependent anaphase pathway triggered by activation of the anaphase-promoting complex (APC). APC degrades securin leading to release and activation of separase which subsequently cleaves RAD21 and releases the cohesin complex from chromatin (Modified from Liu and Krantz 2009). (**b**) CTCF-mediated intra-chromatin cohesion facilitates chromatin loop formation between enhancer and promoter and thus blocks gene expression (e.g., suppression of maternal *IGF2* expression). Conversely, cohesin rings can facilitate establishment of boundary elements to prevent spreading of the silencing effect by forming chromatin loop structures. (**c**) Association of cohesin complex and mediator promotes gene expression by facilitating long-range communication between enhancers and promoter (e.g., activation of pluripotency genes) (Modified from Newman and Young 2010)

11.5 Cohesin's Function in Double-Strand DNA Break Repair

Evidence supporting cohesin's function in DNA double-strand break (DSB) repair has been noted even before the identification of its function in SCC. It was shown that mutation of *rad21* in *S. pombe* results in defects in repairing DNA damage caused by radiation (Birkenbihl and Subramani 1992). More evidence supporting cohesin's function in DSB repair has been accumulated since.

The two main approaches for eukaryotic cells to repair DSB in DNA include non-homologous end joining (NHEJ) and homologous recombination (Hartlerode and Scully 2009). The NHEJ pathway usually results in loss of genetic material and causes chromosomal instability due to directly fused DNA breaks. The homologous recombination pathway uses a homologous template, such as the DNA strand of a sister chromatid, to accurately repair damaged DNA. Homologous recombination requires the impaired chromatid and its sister chromatid to be closely tethered. Thus, it is conceivable that cohesin, which tethers sister chromatids together, is required for DSB during the G2 phase.

Although most of cohesion has been established during S phase, repairing DNA DSB requires the generation of de novo cohesion after DNA replication. The cohesin loading protein SCC2 (Strom et al. 2004; Unal et al. 2004) and establishment protein ECO1 (Sjogren and Nasmyth 2001) are indispensible for DSB repair. SCC2 and ECO1 generate new cohesion near breakpoints when DSB occurs. The newly established cohesion is believed to further tightly tether damaged DNA and its sister chromatid and maintain the chromosomal structure.

SCC establishment during both DNA replication and DSB reparation requires the same cohesin loading and establishment factors. However, distinct protein modifications have been observed for cohesion required for repairing DSB. It has been shown that cohesin components need to be phosphorylated to efficiently carry out DNA damage repair function. For instance, SMC1 and SCC1are phosphorylated by the kinases, ataxia telangiectasia mutated (ATM) and checkpoint kinase 1 (Chk1), respectively, in response to DSB DNA damage (Kim et al. 2002; Yazdi et al. 2002; Heidinger-Pauli et al. 2009). Chk1 phosphorylates Ser83 of Scc1, which promotes acetylation of SCC1 by ECO1. The acetylation of SCC1 inhibits the anti-establishment effect of Wpl1 (RAD61) (Heidinger-Pauli et al. 2009). In contrast, SMC3 is acetylated by ECO1 to antagonize RAD61 as cohesion is established during DNA replication.

Most of the early evidence supporting cohesin's role in DNA damage repair came from studies performed in yeast. Later, it was shown that the SMC1/SMC3 complex is also recruited to DSB regions in human cells (Potts et al. 2006). The recruitment of the cohesin complex is promoted by SMC5/SMC6, which is known to function in DNA damage repair. Interestingly, in response to DNA damage, PDS1 in yeast is stabilized and, in contrast, securin, the human homologue of PDS1, is degraded. The phosphorylation of SMC1 at Ser957 and Ser966 by ATM and ATR (ATM- and Rad3-related) is critical for DNA damage responses (Kim et al. 2002; Yazdi et al. 2002; Garg et al. 2004; Kitagawa et al. 2004). More recently, it was

reported that human NIPBL is recruited to DNA DSB sites that are dependent on mediator of DNA damage checkpoint 1 (MDC1), ring finger protein 168 (RNF168), and heterochromatin protein 1 γ (HP1γ) (Oka et al. 2011), indicating that DSB repair requires de novo recruitment of cohesin. MDC1 and RNF168 are known to accumulate at DSB sites. It was also observed that the ATM-dependent phosphorylation of SMC1 is critical for the mobilization of cohesin in response to ionizing radiation-induced DSB responses (Bauerschmidt et al. 2011).

In summary, DNA DSB stimulates generation of de novo post-replication SCC in G2 phase by activating several cell cycle checkpoint-related kinases. The specific modification of cohesin components by these checkpoint proteins is important for cohesin's recruitment to DSB sites and for repairing these DNA breaks.

11.6 Noncanonical Functions of Cohesin

Sister chromatid cohesion and double-strand break DNA repair are two universal functions of cohesin. The observation that cohesin associates with chromosomes in interphase indicated that cohesin might have sister chromatid cohesion-independent roles (Darwiche et al. 1999). Some noncanonical functions of cohesin such as regulating gene expression, controlling the epigenetic states of chromatids, forming higher-order chromatin structures, and maintaining normal subnuclear organization have been described in recent years.

11.6.1 Cohesin's Function in Regulating Gene Expression

The function of cohesin in regulating gene expression has been established by several observations. The cohesin core components Smc1 and Smc3 were shown to facilitate the boundary element at the yeast Hidden MAT Right (HMR) mating-type locus to prevent the spreading of the silencing effect from the HMR domain to its neighboring chromatin region, possibly through mediating the formation of chromosome loop structure in this region (Fig. 11.2b) (Donze et al. 1999). Additional and critical supporting evidence was the discovery of *Drosophila* Nipped-B as an essential factor facilitating long-range communication between distant enhancers and promoters of essential developmental genes such as *cut* and *Ultrabithorax* homeobox genes (Rollins et al. 1999). Later on, the identification of heterozygous mutations in *NIPBL*, *SMC1A*, or *SMC3* as the cause of CdLS in humans further supported the role of cohesin in regulating gene expression. This is due to the observation that SCC and cell division are not significantly affected in CdLS probands' cells and a conserved pattern of genome-wide expression disruption was observed in these cells that was significantly distinct from controls (Krantz et al. 2004; Tonkin et al. 2004; Liu et al. 2009; Deardorff et al. 2007; Kaur et al. 2005; Musio et al. 2006). Interestingly, reduction of *SMC1* and *SA* in *Drosophila* results in opposite effects on *cut* expression and phenotypes as observed in *Nipped-B* mutants in which

long-range activation of *cut* expression is suppressed and the wing margin nick effect is enhanced (displaying more nicks along the wing margin) (Rollins et al. 2004; Schaaf et al. 2009; Dorsett et al. 2005). These contrasting results might reflect a Nipped-B-dependent dynamic interaction among cohesin, DNA, and other transcriptional regulators.

Cohesin's function in regulating gene expression has been intensively investigated in recent years. A series of chromatin immunoprecipitation (ChIP) studies for defining genome-wide distribution of cohesin in various species has greatly improved our understanding of this noncanonical role of cohesin. These studies revealed that cohesin localization on chromosomes is distinct in different species and also in different cell types (see review by Merkenschlager 2010).

In *S. cerevisiae*, cohesin relocates to convergent transcription chromosomal regions after initial loading at transcription sites. In *S. pombe*, cohesin also locates at convergent transcription sites. Cohesin, in *S. pombe*, is actively involved in regulating transcription termination between convergent genes and the heterochromatin-associated RNA interference machinery which recruits a series of proteins, including cohesin components, to sites of bidirectional transcription so as to define transcription termination sites (Gullerova and Proudfoot 2008).

In *Drosophila*, Nipped-B and cohesin co-localize throughout the genome, and they preferentially bind to the promoter and coding regions of actively transcribed genes (Misulovin et al. 2008). These studies with *Drosophila* cell lines showed that cohesin and Nipped-B's binding on DNA overlaps with RNA polymerase II and is excluded from silenced regions marked with histone H3 lysine 27 trimethylation (H3K27me3). Interestingly, cohesin's binding is tightly associated with the on/off state of gene expression. For instance, cohesin and Nipped-B binding to the Abd-B homeobox gene locus is present in cells in which this gene is expressed and absent in cells in which it is silenced (Misulovin et al. 2008). Although cohesin binding is excluded from silenced regions in *Drosophila* cells, rare exceptions exist. Some genes such as *Enhancer of split* [*E(spl)-C*] and *invected-engrailed* complexes are bound with both cohesin and polycomb group (PcG) silencing protein. These genes usually express at low or moderate level, suggesting cohesin and PcG act together to suppress gene expression in these cases. Interestingly, these gene regions display a bivalent histone modification pattern, having both the silencing histone mark H3K27me3 and activating mark H3K4me3. This bivalent feature might help to explain the dramatic shift of these genes' expression from suppression to activation upon knocking down of cohesin. A similar bivalent phenomenon is also observed for the most upregulated genes upon knocking down cohesin in mouse embryonic cells (Kagey et al. 2010).

A biphasic response to cohesin levels was observed for some cohesin and PcG co-bound genes while knocking down cohesin in *Drosophila* cells. These genes decrease their expression in the first 3 days when cohesin is reduced by 30% but increase their expression by day 6 when cohesin is further reduced up to 80%. This observation reveals that some genes' expression is highly sensitive to cohesin levels. It was hypothesized that this might relate to a balance between cohesin and PcG. Small changes in cohesin levels at a threshold point might break this balance

and shift expression state efficiently from one direction to another. A "weak" and a "stable" form of cohesin binding with chromosomes has also been described in *Drosophila* salivary glands. By FRAP assay, the two types of binding displayed a 20- and 340-second duration, respectively. The amount of stable form is significantly reduced in heterozygous Nipped-B mutants in which Nipped-B expression is decreased by approximately 30% (Gause et al. 2010), suggesting that stable binding of cohesin is crucial for gene regulation.

In mammalian cells, cohesin binding is enriched in DNase I hypersensitive sites and conserved noncoding sequences and co-localizes with CCCTC-binding factor (CTCF) (Parelho et al. 2008; Wendt et al. 2008; Rubio et al. 2008). CTCF is thought to associate with insulators. Insulators are stretches of DNA sequence that serve as blocking elements to prevent the spreading of chromosomal architecture from one chromosomal region to its neighboring region or to inhibit communication between enhancers/silencers and promoters(Ohlsson et al. 2010). As an insulator-related protein, CTCF plays roles in defining imprinting and heterochromatin regions and facilitates chromosome loop formation (Fig. 11.2b) (Ohlsson et al. 2010). Around 1,800 and 8,000 cohesin/CTCF co-localization sites have been identified in non-repetitive regions of the mouse and human genome, respectively (Parelho et al. 2008; Wendt et al. 2008). About 89% of cohesin binding sites co-localize with CTCF sites in humans. These cohesin/CTCF sites are preferentially enriched within a few kilobases of genes, suggesting their function in regulating gene expression (Wendt et al. 2008). It is unlikely that cohesin itself can recognize a specific DNA sequence and selectively bind to this DNA region. It is more likely that cohesin interacts with the CTCF protein and its binding position on DNA is determined by this binding partner. This is supported by experiments in which knockdown of CTCF abolishes cohesin binding and the direct interaction between CTCF and cohesin subunit Scc3. Moreover, knocking down CTCF affects the expression of several hundred genes but does not affect the overall amount of cohesin binding to DNA (Wendt et al. 2008). Thus, it is thought that CTCF helps to recruit cohesin to CTCF sites and co-localization of cohesin and CTCF is critical for normal gene expression.

Co-localization of CTCF and cohesin implies that cohesin mediates CTCF function. Wendt et al. showed that the enhancer blocking effect of CTCF in the H19 imprinting control region is dependent on cohesin (Fig. 11.2b) (Wendt et al. 2008). Hadjur et al. further showed that RAD21 and cohesin promote interferon-gamma (*IFNG*) expression by regulating CTCF-dependent DNA loop formation (Hadjur et al. 2009) (see below for details). Considering the broad spectrum of CTCF function, the close association of cohesin and CTCF might provide hints for understanding cohesin's potential function in regulating gene expression and the epigenetic state of chromatin.

Additional evidence clearly supporting the SCC-independent role of cohesin in gene regulation came from studies of postmitotic neuronal development in the *Drosophila* mushroom body (Pauli et al. 2008; Schuldiner et al. 2008). The mushroom body, an essential structure in *Drosophila* brain, is involved in olfactory learning and memory. Axon pruning is a general natural developmental process to form

mature neuronal circuits. During *Drosophila* mushroom body development, γ neurons initially extend excess dendrites and axons before pupae formation followed by an extensive pruning process to remove all the dendrites and most of the axon branches after pupae formation (Lee et al. 1999). By analyzing mushroom body development in *SMC1* mutants, Schuldiner et al. found that SMC1 is required for γ neuron pruning partially through the regulation of the steroid hormone ecdysone receptor B1 (*EcR-B1*) gene expression (Schuldiner et al. 2008). More interestingly, the γ neuron pruning defect of the mutant can be rescued by expressing *SMC1* or *EcR-B1* in these postmitotic neurons. In another independent study, Pauli et al. generated a transgenic *Drosophila* with a modified cohesin subunit *RAD21* which is cleavable by the tobacco etch virus (TEV) protease (Pauli et al. 2008). They observed that expression of TEV protease in the postmitotic neurons (effectively removing cohesin from these cells) causes a γ neuron pruning defect and lethality. Considering γ neuron pruning is a postmitotic process and cohesin subunits are not disturbed before this process, these results clearly reveal SCC-independent roles of cohesin in regulating gene (e.g., *EcR-B1*) expression. This is further supported by an additional study in the postmitotic *Drosophila* salivary gland using the transgenic TEV cleavable *Rad21* (Pauli et al. 2010), demonstrating that cleavage of RAD21 induces changes in many genes' expression including some genes in the ecdysone steroid signaling pathway. Ecdysone signaling is critical for morphogenesis and molting during *Drosophila* development. It involves many essential cellular processes such as apoptosis, cell division, cell polarity, and cell differentiation (Galikova et al. 2011). Cohesin might also be associated with these cellular processes through the regulation of ecdysone receptor expression. Cohesin has also been found to co-localize with the estrogen receptor in a CTCF-independent manner in response to estrogen in human MCF-7 breast cancer cells (Schmidt et al. 2010), further suggesting that cohesin plays a role in regulating gene expression. This function is critical for the developing organisms and cells to be able to respond to hormone stimulation and seems to be evolutionally conserved and mitosis independent.

11.6.2 Cohesin Facilitates DNA Looping and Higher-Order Genome Architecture

As mentioned above, cohesin co-localizes with CTCF sites in mammalian cells and correlates with gene expression. The exact mechanism of how cohesin regulates gene expression is not fully understood. One proposed model is that cohesin might facilitate DNA looping, so as to regulate communication between distal regulatory elements and promoters and thus to affect the state of gene expression. This has been supported by several recent studies discussed below.

A CTCF-associated insulator plays an important role in controlling reciprocal imprinting of the *IGF2*/*H19* locus on the paternal and maternal alleles (Bell and Felsenfeld 2000; Hark et al. 2000; Szabo et al. 2000). *IGF2* and *H19* are located

within ~100 Kb on the same chromosome and are separated by an insulator sequence upstream of *H19*. This insulator is also called the ICR (imprinting control region). The ICR is methylated on the paternal allele, preventing CTCF binding and inhibiting the enhancer blocking effect of CTCF. Thus, *IGF2* is activated by the enhancer on the paternal allele. In contrast, the ICR is not methylated on the maternal allele, allowing CTCF binding. The CTCF binding blocks communication between the enhancer and *IGF2*. Thus, maternal *IGF2* is inactivated. The paternal and maternal chromatins at this region were observed to form different chromatin structures. CTCF-dependent DNA loops were observed on the maternal allele (Engel et al. 2008; Kurukuti et al. 2006; Li et al. 2008; Qiu et al. 2008; Yoon et al. 2007; Murrell et al. 2004) in mice. Cohesin also binds to this CTCF site and is required for this blocking effect (Fig. 11.2b) (Wendt et al. 2008). In order to test if cohesin contributes to the enhancer blocking effect through the facilitation of DNA looping at this locus, Nativio et al. analyzed the higher-order chromatin structure using the quantitative chromosome conformation capture (3C) technique after knocking down *RAD21* by RNAi in normal breast epithelial cells (Nativio et al. 2009). They found that the overall chromatin association between CTCF sites is significantly reduced upon knocking down *RAD21* and the chromatin association with CTCF sites between the *IGF2* and *H19* locus is reduced by 30% in G2 phase of the cell cycle. The association between *IGF2* and the enhancer is not reduced, and it changes from monoallelic to biallelic association following knockdown, consistent with the observed activation of *IGF2* transcription. These results support the idea that cohesin, recruited by CTCF to the insulator sequence, regulates gene expression by stabilizing higher-order chromatin conformation. The chromatin association is also affected in G1 cells in which no cohesion exists, suggesting that this function is independent of SCC (Nativio et al. 2009).

Cohesin facilitated DNA looping was also reported to be required for the activation of gene expression during development. Hadjur et al. showed that Rad21 and cohesin promote *IFNG* expression by regulating CTCF-dependent DNA loop formation while naive CD4 T cells are induced to form specialized T helper (T_H) 1 cells (Hadjur et al. 2009). Knocking down *RAD21* reduced long-range chromatin interactions at the *IFGN* locus and the level of inducible transcripts of *IFGN* in T_H1 cells. These results indicate that cohesin is involved in long-range chromosomal associations and facilitates cell-type-specific gene activation or suppression.

Additional evidence indicates that CTCF sites and DNA looping play an essential function in regulating gene expression in the β-globin locus (Splinter et al. 2006; Hou et al. 2010; Chien et al. 2011) and the human IL-3/GM-CSF (interleukin-3/granulocyte-macrophage colony-stimulating factor) region (Bowers et al. 2009). It is possible that cohesin also contributes to the higher-order chromosomal conformation at these loci. In summary, there is increasing evidence supporting cohesin's noncanonical role in DNA looping. However, it remains unclear whether cohesin stabilizes intrachromosomal confirmation through additional protein interactions or by direct "entrapment" of DNA strands using the single ring or two-ring handcuff model.

11.6.3 Cohesin Interacts with Mediator to Maintain Pluripotency of Stem Cells by Facilitating DNA Looping

Through an RNAi screen to identify important regulators of mouse embryonic stem cell state maintenance, many cohesin and mediator subunits have been identified (Kagey et al. 2010). Knocking down cohesin core components (*Smc1A*, *Smc3*, and *Stag2*) and the cohesin loading factor *Nipbl* causes deceased expression of essential pluripotency genes such as *Oct4*, *Sox2*, and *Nanog* and increased expression of some differentiation markers, leading to loss of stem cell morphology. Genome-wide localization of Smc1 and Smc3 in mouse embryonic stem cells was investigated by using ChIP-Seq assays (chromatin immunoprecipitation combined with massively parallel DNA sequencing). The results indicated that two types of cohesin binding exist in embryonic stem cells. One is co-bound with CTCF and is unrelated to RNA polymerase II (Fig. 11.2b), and the other binds with mediator around enhancer and core promoter regions and is associated with RNA polymerase II (Fig. 11.2c). Interestingly, the cohesin loading factor Nipbl only co-binds to enhancer and core promoter regions with the mediator complex, indicating that Nipbl preferentially binds to actively transcribed genes. The physical interaction between mediator, cohesin, and Nipbl at enhancer and core promoter regions implies that these proteins might contribute to the formation of DNA loops between the enhancer and core promoter region. Indeed, DNA looping was confirmed between the enhancer and promoter of several pluripotency genes such as *Oct4*, *Nanog*, *Phc1*, and *Lefty1* by 3C assays in ES cells but not in mouse embryonic fibroblast cells in which these genes are silenced. These results strongly supported the idea that mediator and cohesin act to facilitate the association of enhancers and promoters of active genes by forming DNA loops in a cell-type-specific manner (Fig. 11.2c). Thus, cohesin and mediator function to maintain pluripotency of ES cells by activating pluripotent gene expression.

Cohesin has also been reported to be essential for subnuclear localization of genetic elements such as tRNAs, heterochromatin, and telomeres (see details in the review by Bose and Gerton 2010). Proper subnuclear localization of these genetic elements is critical for maintaining their integrity and active/silenced state.

11.7 Cohesin Studies from Multiple Animal Models

11.7.1 Cohesin Studies in Fruit Flies

Drosophila Nipped-B mutants are among the earliest animal models developed to study the cohesin complex and provided a fundamental base for exploring the role of cohesin in regulating gene expression and animal development (Rollins et al. 1999, 2004; Dorsett et al. 2005). Studies from *Drosophila* cohesin mutants suggest that cohesin plays a SCC-independent function, mediating long-range chromosomal

interactions between distant enhancers and promoters of target genes (Dorsett 2007; Hallson et al. 2008) (see above discussion).

11.7.2 Cohesin Studies in Zebrafish

Cohesin has also been shown to be required for *Runx* gene expression in zebrafish. *Runx* genes are transcription factors essential for differentiation of multiple cell lineages during early embryogenesis and are involved in hematopoiesis, osteogenesis, neurogenesis, and gastric epithelial cell growth control (Blyth et al. 2005; Ito 2004). Horsfield et al. showed that a *Rad21* mutation, or knockdown of *Smc3*, impaired *Runx* expression and led to a series of developmental defects including failure of blood cell differentiation (Horsfield et al. 2007). *Runx* expression defects were also observed in the heterozygous *Rad21* mutants. Subsequently, Rhodes et al. revealed that *myca* (zebrafish *myc*) is positively regulated by cohesin, consistent with the observation in *Drosophila* (Rhodes et al. 2010). *Myc* is also involved in multiple critical cellular processes during development. In contrast to the observations from the *Rad21* mutant and *Smc3*-depleted fish, *Runx*1 is normally expressed and *myca* is upregulated in *Esco2*-depleted zebrafish embryos (Monnich et al. 2011). Biallelic mutations in *Esco2* in humans cause Roberts syndrome (RBS) (Schule et al. 2005; Vega et al. 2005). The *Esco2*-depleted zebrafish embryos presented RBS-like features including craniofacial and limb defects. Cell cycle blocking at G2/M phase and high levels of cell death are also observed in these embryos. Expression profile studies further revealed that genes involved in cell cycle and apoptosis are affected in the *Esco2*-depleted embryos, suggesting that proliferation and apoptosis abnormalities might be responsible for the features seen in RBS.

11.7.3 Cohesin Studies in Mice

More than half of the probands with Cornelia de Lange syndrome (CdLS) carry heterozygous mutations in *NIPBL*. In order to establish a mouse model of CdLS, Kawauchi et al. generated mice with a heterozygous gene-trap mutation of *Nipbl* (Kawauchi et al. 2009). These heterozygous mice display several features similar to human CdLS probands including prenatal growth delay, craniofacial anomalies, microbrachycephaly, congenital heart defects, failure to thrive, and hearing loss. These mice also had delayed bone maturation, irregular behavior, and high ratio of early lethality. Similar to the CdLS probands (and *Nipped-B* mutant *Drosophila*), the mutant heterozygous mice demonstrate *Nipbl* transcript levels that are about 70% of wild-type levels, and sister chromatid cohesion was not affected. Gene expression studies revealed the misregulation of many genes, suggesting that the underlying pathogenic mechanism is likely related to cohesin's role in regulating gene expression.

Pds5 along with Wp11 associates with the cohesin ring and acts to inhibit cohesion establishment (Neuwald and Hirano 2000; Panizza et al. 2000; Rowland et al. 2009). Interestingly, homozygous *Pds5* knockout mice have multiple defects that also overlap with features seen in CdLS (Zhang et al. 2007, 2009). Pds5A and Pds5B are two homologues of Pds5 in vertebrates. Both homozygous *Pds5A* and *Pds5B* knockout mice display early mortality and many other developmental malformations including congenital heart defects, cleft palate, skeletal defects, and growth retardation. In contrast, renal agenesis is observed in *Pds5A* null mice, and limb defects are observed in *Pds5B* null mice. Double homozygous mutants of *Pds5A* and *Pds5B* are lethal in the very early embryonic stages, while loss of three alleles of *Pds5* causes a later embryonic lethality, suggesting that appropriate dosage of *Pds5* is critical for development. Cohesion defects were not observed in these mutant mice either, suggesting that the developmental defects of these mice are not related to the cohesion function of cohesin. *Rad21* heterozygous mice are sensitive to irradiation treatment and display defects in DNA repair (Xu et al. 2010).

As mentioned before, some cohesin components are unique to mitosis or meiosis. In many species, Rec8 substitutes SCC1/Rad21 in meiotic cohesion. Similarly, Smc1B and Stag3 are meiosis-specific cohesin components. *Rad21L* is a newly identified member of the meiosis-specific cohesin complex (Ishiguro et al. 2011; Lee and Hirano 2011; Gutierrez-Caballero et al. 2011). Mice with mutations in several meiosis-specific cohesin genes such as *Rec8*, *Smc1B*, and *Rad21L* have been generated in recent years (Bannister et al. 2004; Xu et al. 2005; Revenkova et al. 2004; Herran et al. 2011). Not surprisingly, all these mutant mice display infertility/sterility phenotypes and various meiotic defects.

11.8 Cohesin and Human Disorders

As outlined above, cohesin plays important roles in sister chromatid cohesion, double-strand DNA break repair, gene expression, and ensuring higher order of chromatin structure. A number of human disorders have been found to be caused by the disruption of structural and regulatory cohesin-associated genes and have collectively been termed the "cohesinopathies." The two most well characterized of these disorders are Cornelia de Lange syndrome (CdLS) and Roberts syndrome (RBS).

11.8.1 Cornelia de Lange Syndrome (CdLS)

Cornelia de Lange syndrome (CdLS) (OMIM #122470, #300590, and #610759), also referred to as Brachmann–de Lange syndrome, was initially reported by Vrolik in 1839 and Brachmann in 1916 (Vrolik 1849; Brachmann 1916). Cornelia de Lange reported two unrelated individuals with strikingly similar features and proposed the diagnostic criteria for this condition in 1933 (Lange 1933).

CdLS is a dominantly inherited genetically heterogeneous diagnosis characterized by multiple organ system differences including typical facial features, somatic growth delay, intellectual disability, limb defects (primarily affecting the upper limbs), congenital heart defects, hirsutism, and gastrointestinal and other visceral system involvement (Liu and Krantz 2009, 2008). The prevalence of CdLS is estimated to be approximately 1 in 10,000 (Opitz 1985). However, this is likely an underestimate as the clinical presentations can be quite variable and milder cases are likely not recognized as CdLS.

The facial features are the most clinically consistent and recognizable findings in CdLS (Fig. 11.3). Most individuals have a short neck, low posterior hairline, hirsute forehead, arched eyebrows, synophrys, ptosis, thick and long eyelashes, low-set ears, flattened midface, short nose, long philtrum, a thin upper lip with downturned corners, a high (or cleft) palate, widely spaced teeth, and micrognathia (Jackson et al. 1993; Kline et al. 2007a, b). Typical extremity findings range from small hands and small feet to more severe reduction defects (primarily affecting the ulnar structures) of the upper limbs (seen in approximately one third of probands) (Fig. 11.3). Disproportional shortening of the first metacarpal with resulting proximally placed thumb, brachydactyly, clinodactyly, and single palmar creases are common findings (Jackson et al. 1993). Probands can also have radial head dislocation with radioulnar synostosis and incomplete elbow extension (Jackson et al. 1993). Hypertrichosis is mainly on the face, back, and extremities. Cutis marmorata can also be seen in half of the probands (Jackson et al. 1993). Multiple organ systems are involved in CdLS. Gastroesophageal reflux disease (GERD) is almost universal (Luzzani et al. 2003). Pyloric stenosis, diaphragmatic hernia, malrotation, and increased risk for volvulus formation have also been frequently reported (Masumoto et al. 2001). A quarter of probands also have a congenital heart defects, the most common being ventricular or atrial septal defects, although other lesions are also seen (Mehta and Ambalavanan 1997; Tsukahara et al. 1998). Renal malformations and dysfunction can be seen as well (e.g., vesicoureteral reflux, pelvic dilatation, and renal dysplasia) (Selicorni et al. 2005). Ophthalmologically peripapillary pigmentation, high myopia, ptosis, blepharitis, mild forms of microcornea, and nasolacrimal duct obstruction are more commonly described (Wygnanski-Jaffe et al. 2005). Auditory and vestibular anomalies include both sensorineural and conductive hearing loss, recurrent otitis media, and sinusitis (Sataloff et al. 1990; Kaga et al. 1995). Orthopedic manifestations, beyond the upper limb deficiencies, include hip dislocation or dysplasia, scoliosis, tight Achilles tendons, and delayed maturation of bone (Roposch et al. 2004; Russell et al. 2001). Genitalia are in general hypoplastic with cryptorchidism, micropenis, and hypospadias being commonly seen in males and small labia majora in females. Fertility is normal among less severely affected probands (Russell et al. 2001). Premature aging has been suggested (Jackson et al. 1993; Kline et al. 2007b). There is no obvious increased risk of cancer. Thrombocytopenia has also been consistently reported (Froster and Gortner 1993).

Probands have proportionate small stature that occurs prenatally, usually manifesting late in the second trimester. At birth, all measurement parameters tend to be below the 10th percentiles, and fall to below the fifth percentiles by early childhood,

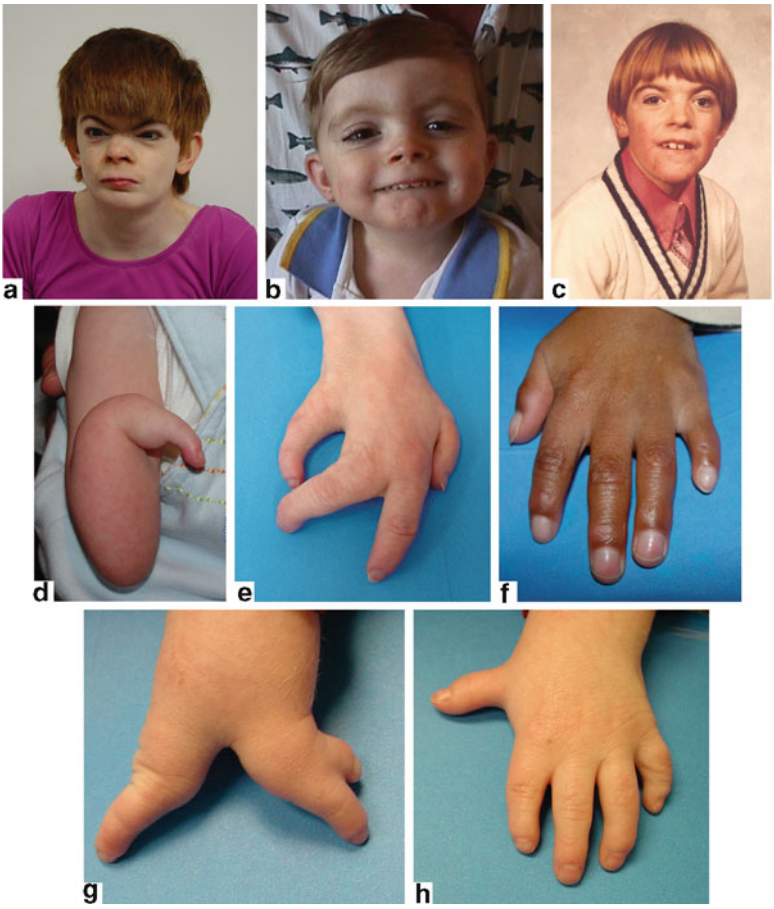

Fig. 11.3 *Clinical features in CdLS.* (**a**–**c**) Facial features in probands with (**a**) a truncating *NIPBL* mutation, (**b**) an *SMC1A* mutation, and (**c**) an *SMC3* mutation. Note characteristic facial appearance in all probands (arched eyebrows, flat nasal root, short upturned nasal tip, long philtrum, and thin upper lip); however, the features are much more pronounced in the child with the truncating *NIPBL* mutation as compared to the children with *SMC1A* and *SMC3* mutation. (**d**–**h**) Photographs of the variable involvement of the hands and forearms in children with CdLS. (**d**) Depicts the more severe end of the spectrum with complete absence of the ulnar structures and severely hypoplastic radius with the only digit formed being the thumb, (**e**) an intermediate form of oligodactyly, where the radial structures are relatively preserved. (**f**) Mild involvement of the hand with micromelia (*small hands*) and fifth finger clinodactyly and hypoplasia and (**g**–**h**) the hands of the same proband demonstrating the asymmetrical involvement that is typical in CdLS

with growth paralleling the standard growth curves. CdLS-specific growth curves are available (Kline et al. 1993). In adulthood, both the average height and weight are below the third percentiles, with a mean head circumference of 49 cm that is consistent with significant microcephaly (Kline et al. 1993).

Developmental delay and intellectual disability are typically observed. Speech and language are most significantly affected, while perceptual organization and

visual–spatial memory are more preserved. The average IQ ranges from mild to moderate intellectual disability; however, both borderline normal intelligence and severe intellectual disability are commonly reported. Learning continues throughout life without evidence of regression (Kline et al. 1993).

Almost every proband has behavioral issues that may be caused or aggravated by physical complications, including self-injurious behavior, obsessive–compulsive behaviors, attention deficit disorder with or without hyperactivity, short attention span, sleep disturbances, depression, and autistic features (Luzzani et al. 2003; Berney et al. 1999; Hyman et al. 2002; Moss et al. 2005). Seizures are the primary neuropathological manifestation. No specific electroencephalography (EEG) pattern has been described, and the seizures can generally be well managed with standard medical intervention. Both hypertonicity and hypotonia occur. Probands tend to have a high pain threshold probably due to poorly characterized peripheral neuropathy (Kline et al. 2007a).

11.8.2 CdLS due to NIPBL, SMC1A, or SMC3 Mutations

About 60% of CdLS probands have a heterozygous mutation in *NIPBL*. Genotype–phenotype correlations among a large number of probands indicate that haploinsufficient *NIPBL* mutations (protein-truncating mutations such as nonsense mutations, splice site mutations, and out-of-frame deletions or insertions) usually result in a more severe cognitive and structural phenotype than missense mutations (Gillis et al. 2004). Approximately 5% of probands with a clinical diagnosis of CdLS were found to have missense or small in-frame deletion mutations in *SMC1A*, and one individual was found to have an in-frame 3 bp deletion in the *SMC3* gene (Deardorff et al. 2007). The *SMC1A* and *SMC3* cases have mild to moderate intellectual disability without significant impairments in growth or structural abnormalities of the limb or other organ systems (Deardorff et al. 2007). Notably, probands with *SMC1A* or *SMC3* mutations demonstrated some clinical features that are in contrast to the "classical" form of CdLS (Deardorff et al. 2007). This cohort tends to have a more prominent nasal bridge than is typically seen in CdLS (Musio et al. 2006), and the majority of them had birth weights within normal parameters with normal head circumferences and growth measures later in life as well (Fig. 11.3). For the most, walking and speech are often acquired, and overall, they exhibit a milder level of cognitive involvement (Deardorff et al. 2007). The molecular etiology of the remaining 35% of probands is unknown at this time.

11.8.3 Roberts/SC Phocomelia Syndrome

Roberts/SC phocomelia syndrome (RBS OMIM #268300; SC OMIM #269000) is an autosomal recessive developmental disorder caused by homozygous or compound heterozygous mutations in the *ESCO2* gene (Schule et al. 2005; Vega et al. 2005).

The clinical features of Roberts syndrome are distinct from CdLS but with some overlap (Schule et al. 2005; Vega et al. 2010) and include growth retardation, symmetric mesomelic shortening of the limbs (in which the upper limbs were more commonly and severely affected than the lower limbs), and characteristic facies with microcephaly (Vega et al. 2010). The severity of malformations of the facies tends to correlate with the severity of limb reduction (Vega et al. 2010).

The facial feature in RBS is characterized by microcephaly, hypoplastic nasal alae, malar hypoplasia, hypertelorism, micrognathia, hemangiomas, exophthalmos, down-slanting palpebral fissures, and cleft lip and palate (Vega et al. 2010). Limb reduction affects the distal–proximal and anterior–posterior axes, resulting in a mesomelic reduction with a hand-specific affection pattern in which the thumb is always the first finger being affected (Vega et al. 2010). In the upper limbs, the radius is always affected, followed in frequency by the ulna (97.6%) and the humerus (78.1%). Hands were characteristically affected, with 97.8% of the cases affected with either aplasia (66.7%) or hypoplasia (31.1%) of the thumbs (Vega et al. 2010). Other fingers were affected at a lower frequency. In the lower limbs, the fibula was the bone most commonly and severely affected (73.8%), followed by the tibia (69%) and the femur (57.5%) (Vega et al. 2010). Other organ system involvement includes congenital heart defects (primarily atrial and ventricular septal defects), genitourinary anomalies (hypospadias, cryptorchidism, bicornuate uterus), structural renal anomalies, and variable intellectual disability ranging from normal intelligence to significant impairment, but milder on average than that seen in CdLS.

As previously described, *Esco2* is the human homologue of yeast *Eco1* which encodes an acetyltransferase involved in the acetylation of SMC3 and is essential for the establishment of sister chromatid cohesion (Hou and Zou 2005). Cell lines derived from Roberts probands show heterochromatic repulsion (HR) which was demonstrated by premature sister chromatid separation primarily at the heterochromatic regions on prophase and metaphase chromosomes (German 1979). Cytogenetically, HR appears in 100% of RBS–SC phocomelia probands, is highly correlated with the phenotype and ESCO2 mutations (Schule et al. 2005), and has been used for prenatal diagnosis (Schulz et al. 2008). It has been shown that most mutations in the *ESCO2* gene identified in RBS probands result in disruption of the acetyltransferase domain. This results in faulty cohesion, and other cellular events in RBS cell lines, indicating that acetyltransferase activity contributes to the development of the major organ systems affected in RBS (Gordillo et al. 2008).

Despite the fact that the clinical presentations seen in CdLS and RBS have some overlap and the molecular mechanisms are similar, these two congenital disorders are quite distinct. CdLS is a dominant disorder: 60% of probands carry heterozygous mutations in *NIPBL*, 5% carry heterozygous mutations in *SMC1A*, and one proband carries an *SMC3* mutation (19). RBS/SC phocomelia is an autosomal recessive disorder with all probands caused by either homozygous or compound heterozygous mutations in *ESCO2*. No significant genotype–phenotype correlations have been described.

11.9 Other Disorders Demonstrating Cohesion Defects

Additional disorders have been identified that have associated cohesion defects, but the causative proteins are not known to directly interact with the cohesin complex or its regulation at this time.

11.9.1 α-Thalassemia/Mental Retardation Syndrome, X-Linked

α-Thalassemia/mental retardation syndrome, X-linked (ATRX) (OMIM #301040), is a multisystem disorder characterized by postnatal growth and mental deficiency, microcephaly, dysmorphic craniofacial features (hypertelorism, midface hypoplasia, anteverted nares, and full lips with protruding tongue), lack of speech, seizures, and abnormal genitalia in males (Gibbons et al. 1995). Affected individuals usually have a mild form of hemoglobin H (Hb H) disease. ATRX is caused by mutations in the *ATRX* gene on the X chromosome (Gibbons et al. 1995). The *ATRX* gene encodes a chromatin remodeling enzyme that associates with the chromo shadow domain of HP1α (as does NIPBL) and preferentially localizes to the pericentromeric heterochromatin in mouse and human cells (Ritchie et al. 2008). ATRX was suggested to have a role in loading cohesin onto chromatin during S phase and recruiting cohesin to specific chromosomal loci (Ritchie et al. 2008). Defective sister chromatid cohesion and impaired chromosome congression was observed in cultured human cells depleted for ATRX, indicating a disruption of mitotic progression. Similar findings were seen in embryonic mouse brains with no ATRX protein (Ritchie et al. 2008). The impaired cohesin targeting or transportation due to mutations in the *ATRX* gene may therefore contribute to the clinical phenotypes in ATRX syndrome.

11.9.2 Warsaw Breakage Syndrome

Van der Lelij et al. (van der Lelij et al. 2010) reported a single male child with severe microcephaly, pre- and postnatal growth retardation, and abnormal skin pigmentation that was found to have mitomycin C (MMC)-induced chromosomal breakage in fresh T-lymphocyte cultures as well as in EBV-immortalized B lymphoblasts. Centromeric cohesion ("railroading") and premature chromatid separation (PCS) defects were seen in 50–60% of cells. Reduced levels of the DEAD/H box polypeptide 11 (DDX11) helicase were identified, and compound heterozygous mutations in this gene were subsequently identified. DDX11 is the ortholog of yeast Chl1 and siRNA experiments in human cells point to a role for DDX11 in sister chromatid cohesion.

11.10 Human Malignancies

There is increasing evidence that links disruption of the cohesin complex or the cohesion pathway to many forms of human cancer. The tumor suppressor gene breast cancer 1 early onset (*BRCA1*) associates with many factors that function in the sister chromatid cohesion pathway, indicating a role in BRCA1 tumorigenesis (Mayer et al. 2004; Kobayashi et al. 2004; Petronczki et al. 2004). BRCA1 and Eco1/Ctf7 family members share overlapping partners, and cells harboring mutations in either BRCA1- or ESCO-related pathways exhibit similar chromosomal abnormalities including cohesion defects, especially along heterochromatic and centromeric regions (Skibbens 2005; Skibbens et al. 2007). As previously described, WAPL and PDS5 form a complex weakly associating with cohesion rings and inhibit cohesion establishment. Interestingly, there is increasing evidence supporting an association between WAPL and malignancy. Human WAPL protein overexpression was found in cervical cancers and significantly correlated to the grades of the malignancy (Oikawa et al. 2004, 2008). NIH 3T3 cells overexpressing WAPL produced tumors in 100% of injected nude mice (Oikawa et al. 2004). Human papillomavirus E6 and E7 oncoproteins are able to induce the expression of human WAPL (Kwiatkowski et al. 2004). Downregulated WAPL inhibited the growth of tumors derived from cervical cancer cell lines; therefore, WAPL was proposed as a therapeutic target for cervical cancer. In addition, a splice variant of WAPL may also associate with other types of human neoplasia because it interacts with the Epstein–Barr virus transformation-related protein EBNA2 in human cells (Oikawa et al. 2008; Kuroda et al. 2005). The contribution of dysregulated WAPL to cervical carcinogenesis may be partially due to chromosomal instability (CIN) (Ohbayashi et al. 2007).

Separase digests RAD21 at the beginning of anaphase to release cohesin from the sister chromatids, and it has been identified as a potential tumor suppressor gene in zebrafish (Horsfield et al. 2007). Heterozygous mutations of separase contribute to the initiation and progression of epithelial tumors, partially due to genome instability (Shepard et al. 2007). In *Drosophila*, epithelial organization and integrity seems to be affected the most by loss of separase (Pandey et al. 2005). Separase has been postulated to act as an oncogene as significant overexpression of separase was detected in human breast tumors, most of which are infiltrating ductal carcinomas (Zhang et al. 2008c). Overexpression of separase alone is sufficient to induce aneuploid tumors in mouse mammary epithelial cells under a p53 mutant background. Cohesion defects (premature sister chromatid separation) were manifested in separase-induced cell lines.

Defective sister chromatid cohesion was suggested to play a major role in human colorectal cancers (Barber et al. 2008). A systematic study to identify somatic mutations in potential CIN genes in 132 colorectal cancer samples has identified 11 somatic mutations distributed among five genes: *SMC1A*, *NIPBL*, *SMC3*, *STAG3*, and *RNF20*. Many other regulatory factors of cohesin complex have been discussed in cancer research as well. Human securin which inhibits separase's enzyme activity before the onset of anaphase is actually the human proto-oncogene pituitary tumor-transforming gene (*PTTG*) (Zhang et al. 1999a). The protein levels of securin

were reported to correlate to the invasiveness of pituitary tumors. Securin is able to transform cultured cells, and its expression is elevated in human cancer cell lines (Zou et al. 1999). Human cancer cells with securin loss-of-function mutations show high levels of CIN (Jallepalli et al. 2001), and cells overexpressing securin produce tumors in nude mice (Zhang et al. 1999b). In addition, the cohesion establishment factor EFO2/ESCO2 is highly upregulated in aggressive melanoma cells (Ryu et al. 2007).

11.11 Summary

Discoveries associating disruption of structural and regulatory components of the cohesin complex with human developmental disorders and cancers have greatly stimulated research interest in cohesin biology. Through the use of multiple model systems ranging from yeast to human cells, fundamental functions of cohesin complex in sister chromatid cohesion, double-strand DNA repair, regulation of gene expression, and structural organization of genomic architecture have been identified. Cohesin has also been found to be indispensible in cell division, maintaining pluripotency of stem cells and ensuring normal organ development. These novel advances in cohesin research have greatly improved our understanding of the molecular mechanism leading to the cohesinopathies and have laid the groundwork toward the identification of potential targets for therapeutic interventions.

Abbreviations

ATM	Ataxia telangiectasia mutated
ATR	ATM- and Rad3-related
ATRX	Thalassemia/mental retardation syndrome X-linked
BRCA1	Breast cancer 1 early onset 1
CdLS	Cornelia de Lange syndrome
Chk1	Checkpoint kinase 1
Chl1	PCNA and chromosome loss 1
CIN	Chromosomal instability
CTCF	CCCTC-binding factor
Ctf18	Chromosome transmission fidelity 18
DDX11	DEAD/H box polypeptide 11
DSB	Double-strand break
EcR-B1	Ecdysone receptor B1
EEG	Electroencephalography
Elg1	Enhanced level of genomic instability 1
FRAP	Fluorescence recovery after photobleaching
GERD	Gastroesophageal reflux disease

Hb H	Hemoglobin H
HMR	Hidden MAT Right
HP1γ	Heterochromatin protein 1 γ
HR	Heterochromatic repulsion
IFNG	Interferon-gamma
MDC1	Mediator of DNA damage checkpoint 1
NHEJ	Nonhomologous end-joining
NIPBL	Nipped-B homologue
PDS5	Precocious dissociation of sisters 5
PP2A	Protein phosphatase type A
PTTG	Proto-oncogene pituitary tumor-transforming gene
RBS	Roberts syndrome
RNF168	Ring finger protein 168
SCC	Sister chromatid cohesion
SMC	Structural maintenance of chromosome
T_H	T helper
Wapl1	Wings apart-like 1
WBS	Warsaw breakage syndrome

References

Andrade MA, Bork P (1995) HEAT repeats in the Huntington's disease protein. Nat Genet 11(2):115–116

Arumugam P et al (2003) ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol 13(22):1941–1953

Bannister LA et al (2004) Positional cloning and characterization of mouse mei8, a disrupted allele of the meiotic cohesin Rec8. Genesis 40(3):184–194

Barber TD et al (2008) Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc Natl Acad Sci USA 105(9):3443–3448

Bauerschmidt C et al (2011) Cohesin phosphorylation and mobility of SMC1 at ionizing radiation-induced DNA double-strand breaks in human cells. Exp Cell Res 317(3):330–337

Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405(6785):482–485

Bernard P et al (2006) A screen for cohesion mutants uncovers Ssl3, the fission yeast counterpart of the cohesin loading factor Scc4. Curr Biol 16(9):875–881

Berney TP, Ireland M, Burn J (1999) Behavioural phenotype of Cornelia de Lange syndrome. Arch Dis Child 81(4):333–336

Birkenbihl RP, Subramani S (1992) Cloning and characterization of rad21 an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. Nucleic Acids Res 20(24):6605–6611

Blyth K, Cameron ER, Neil JC (2005) The RUNX genes: gain or loss of function in cancer. Nat Rev Cancer 5(5):376–387

Borck G et al (2006) Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5′ untranslated region of the NIPBL Gene. Hum Mutat 27(8):731–735

Bose T, Gerton JL (2010) Cohesinopathies, gene expression, and chromatin organization. J Cell Biol 189(2):201–210

Bowers SR et al (2009) A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes. Mol Cell Biol 29(7):1682–1693
Brachmann W (1916) Ein fall von symmetrischer monodaktylie durch Ulnadefekt, mit symmetrischer flughautbildung in den ellenbeugen, sowie anderen abnormitaten (zwerghaftogkeit, halsrippen, behaarung). Jarb Kinder Phys Erzie 84:225–235
Canudas S, Smith S (2009) Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. J Cell Biol 187(2):165–173
Chang CR et al (2005) Targeting of cohesin by transcriptionally silent chromatin. Genes Dev 19(24):3031–3042
Chien R et al (2011) Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. J Biol Chem 286(20):17870–17878
Ciosk R et al (2000) Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5(2):243–254
Darwiche N, Freeman LA, Strunnikov A (1999) Characterization of the components of the putative mammalian sister chromatid cohesion complex. Gene 233(1–2):39–47
Deardorff MA et al (2007) Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. Am J Hum Genet 80(3):485–494
Donze D et al (1999) The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. Genes Dev 13(6):698–708
Dorsett D (2007) Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. Chromosoma 116(1):1–13
Dorsett D et al (2005) Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development 132(21):4743–4753
Engel N et al (2008) Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. Hum Mol Genet 17(19):3021–3029
Feeney KM, Wasson CW, Parish JL (2010) Cohesin: a regulator of genome integrity and gene expression. Biochem J 428(2):147–161
Froster UG, Gortner L (1993) Thrombocytopenia in the Brachmann-de Lange syndrome. Am J Med Genet 46(6):730–731
Galikova M et al (2011) Steroid hormone regulation of *C. elegans* and Drosophila aging and life history. Exp Gerontol 46(2–3):141–147
Gandhi R, Gillespie PJ, Hirano T (2006) Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. Curr Biol 16(24):2406–2417
Garg R et al (2004) Chromatin association of rad17 is required for an ataxia telangiectasia and rad-related kinase-mediated S-phase checkpoint in response to low-dose ultraviolet radiation. Mol Cancer Res 2(6):362–369
Gause M et al (2010) Dosage-sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5, and Wapl. Mol Cell Biol 30(20):4940–4951
Gerlich D et al (2006) Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Curr Biol 16(15):1571–1578
German J (1979) Roberts' syndrome. I. Cytological evidence for a disturbance in chromatid pairing. Clin Genet 16(6):441–447
Gibbons RJ et al (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 80(6):837–845
Gillis LA et al (2004) NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet 75(4):610–623
Gordillo M et al (2008) The molecular mechanism underlying Roberts syndrome involves loss of ESCO2 acetyltransferase activity. Hum Mol Genet 17(14):2172–2180
Gruber S, Haering CH, Nasmyth K (2003) Chromosomal cohesin forms a ring. Cell 112(6):765–777
Guacci V, Koshland D, Strunnikov A (1997) A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. Cell 91(1):47–57

Gullerova M, Proudfoot NJ (2008) Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. Cell 132(6):983–995
Gutierrez-Caballero C et al (2011) Identification and molecular characterization of the mammalian alpha-kleisin RAD21L. Cell Cycle 10(9):1477–1487
Hadjur S et al (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 460(7253):410–413
Haering CH et al (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell 9(4):773–788
Haering CH et al (2008) The cohesin ring concatenates sister DNA molecules. Nature 454(7202):297–301
Hallson G et al (2008) The Drosophila cohesin subunit Rad21 is a trithorax group (trxG) protein. Proc Natl Acad Sci USA 105(34):12405–12410
Hark AT et al (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405(6785):486–489
Hartlerode AJ, Scully R (2009) Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J 423(2):157–168
Hauf S, Waizenegger IC, Peters JM (2001) Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science 293(5533):1320–1323
Hauf S et al (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS Biol 3(3):e69
Heidinger-Pauli JM, Unal E, Koshland D (2009) Distinct targets of the Eco1 acetyltransferase modulate cohesion in S phase and in response to DNA damage. Mol Cell 34(3):311–321
Heidinger-Pauli JM, Onn I, Koshland D (2010) Genetic evidence that the acetylation of the Smc3p subunit of cohesin modulates its ATP-bound state to promote cohesion establishment in *Saccharomyces cerevisiae*. Genetics 185(4):1249–1256
Herran Y et al (2011) The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. EMBO J 30(15):3091–3105
Horsfield JA et al (2007) Cohesin-dependent regulation of Runx genes. Development 134(14):2639–2649
Hou F, Zou H (2005) Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. Mol Biol Cell 16(8):3908–3918
Hou C, Dale R, Dean A (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. Proc Natl Acad Sci USA 107(8):3651–3656
Hyman P, Oliver C, Hall S (2002) Self-injurious behavior, self-restraint, and compulsive behaviors in Cornelia de Lange syndrome. Am J Ment Retard 107(2):146–154
Inoue A et al (2007) Loss of ChlR1 helicase in mouse causes lethality due to the accumulation of aneuploid cells generated by cohesion defects and placental malformation. Cell Cycle 6(13):1646–1654
Ishiguro K et al (2011) A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. EMBO Rep 12(3):267–275
Ito Y (2004) Oncogenic potential of the RUNX gene family: 'overview'. Oncogene 23(24):4198–4208
Ivanov D, Nasmyth K (2007) A physical assay for sister chromatid cohesion in vitro. Mol Cell 27(2):300–310
Jackson L et al (1993) de Lange syndrome: a clinical review of 310 individuals. Am J Med Genet 47(7):940–946
Jallepalli PV et al (2001) Securin is required for chromosomal stability in human cells. Cell 105(4):445–457
Kaga K et al (1995) Auditory brainstem responses in children with Cornelia de Lange syndrome. Int J Pediatr Otorhinolaryngol 31(2–3):137–146
Kagey MH et al (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314):430–435
Kaur M et al (2005) Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. Am J Med Genet A 138(1):27–31

Kawauchi S et al (2009) Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/−) mouse, a model of Cornelia de Lange Syndrome. PLoS Genet 5(9):e1000650
Kenna MA, Skibbens RV (2003) Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. Mol Cell Biol 23(8):2999–3007
Kim ST, Xu B, Kastan MB (2002) Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. Genes Dev 16(5):560–570
Kitagawa R et al (2004) Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. Genes Dev 18(12):1423–1438
Klein F et al (1999) A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell 98(1):91–103
Kline AD, Barr M, Jackson LG (1993) Growth manifestations in the Brachmann-de Lange syndrome. Am J Med Genet 47(7):1042–1049
Kline AD et al (2007a) Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. Am J Med Genet A 143A(12):1287–1296
Kline AD et al (2007b) Natural history of aging in Cornelia de Lange syndrome. Am J Med Genet C Semin Med Genet 145C(3):248–260
Kobayashi J et al (2004) NBS1 and its functional role in the DNA damage response. DNA Repair (Amst) 3(8–9):855–861
Koshland D, Hartwell LH (1987) The structure of sister minichromosome DNA before anaphase in *Saccharomyces cerevisiae*. Science 238(4834):1713–1716
Krantz ID et al (2004) Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 36(6):631–635
Kueng S et al (2006) Wapl controls the dynamic association of cohesin with chromatin. Cell 127(5):955–967
Kuroda M et al (2005) The human papillomavirus E6 and E7 inducible oncogene, hWAPL, exhibits potential as a therapeutic target. Br J Cancer 92(2):290–293
Kurukuti S et al (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci USA 103(28):10684–10689
Kwiatkowski BA et al (2004) Identification and cloning of a novel chromatin-associated protein partner of Epstein-Barr nuclear protein 2. Exp Cell Res 300(1):223–233
Lange D (1933) Sur un type nouveau de d'eg'en'eration (typus Amstelodamensis). Arch Med Enfants 36:713–719
Lee J, Hirano T (2011) RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. J Cell Biol 192(2):263–276
Lee T, Lee A, Luo L (1999) Development of the Drosophila mushroom bodies: sequential generation of three distinct types of neurons from a neuroblast. Development 126(18):4065–4076
Lengronne A et al (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430(6999):573–578
Lengronne A et al (2006) Establishment of sister chromatid cohesion at the *S. cerevisiae* replication fork. Mol Cell 23(6):787–799
Li T et al (2008) CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol 28(20):6473–6482
Liu J, Krantz ID (2008) Cohesin and human disease. Annu Rev Genomics Hum Genet 9:303–320
Liu J, Krantz ID (2009) Cornelia de Lange syndrome, cohesin, and beyond. Clin Genet 76(4): 303–314
Liu J et al (2009) Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol 7(5):e1000119
Losada A, Hirano T (2005) Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev 19(11):1269–1287
Luzzani S et al (2003) Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical symptoms. Am J Med Genet A 119A(3):283–287

Masumoto K, Izaki T, Arima T (2001) Cornelia de Lange syndrome associated with cecal volvulus: report of a case. Acta Paediatr 90(6):701–703
Mayer ML et al (2001) Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. Mol Cell 7(5):959–970
Mayer ML et al (2004) Identification of protein complexes required for efficient sister chromatid cohesion. Mol Biol Cell 15(4):1736–1745
Mc Intyre J et al (2007) In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae. EMBO J 26(16):3783–3793
McNairn AJ, Gerton JL (2009) Intersection of ChIP and FLIP, genomic methods to study the dynamics of the cohesin proteins. Chromosome Res 17(2):155–163
Mehta AV, Ambalavanan SK (1997) Occurrence of congenital heart disease in children with Brachmann-de Lange syndrome. Am J Med Genet 71(4):434–435
Melby TE et al (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. J Cell Biol 142(6):1595–1604
Merkenschlager M (2010) Cohesin: a global player in chromosome biology with local ties to gene regulation. Curr Opin Genet Dev 20(5):555–561
Michaelis C, Ciosk R, Nasmyth K (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell 91(1):35–45
Misulovin Z et al (2008) Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma 117(1):89–102
Moldovan GL, Pfander B, Jentsch S (2006) PCNA controls establishment of sister chromatid cohesion during S phase. Mol Cell 23(5):723–732
Monnich M et al (2011) A zebrafish model of Roberts syndrome reveals that Esco2 depletion interferes with development by disrupting the cell cycle. PLoS One 6(5):e20051
Moss J et al (2005) The association between environmental events and self-injurious behaviour in Cornelia de Lange syndrome. J Intellect Disabil Res 49(Pt 4):269–277
Murray AW, Szostak JW (1985) Chromosome segregation in mitosis and meiosis. Annu Rev Cell Biol 1:289–315
Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet 36(8): 889–893
Musio A et al (2006) X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38(5):528–530
Nasmyth K, Haering CH (2009) Cohesin: its roles and mechanisms. Annu Rev Genet 43:525–558
Nativio R et al (2009) Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet 5(11):e1000739
Neuwald AF, Hirano T (2000) HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res 10(10):1445–1452
Newman JJ, Young RA (2010) Connecting transcriptional control to chromosome structure and human disease. Cold Spring Harb Symp Quant Biol 75:227–235
Ohbayashi T et al (2007) Unscheduled overexpression of human WAPL promotes chromosomal instability. Biochem Biophys Res Commun 356(3):699–704
Ohlsson R, Bartkuhn M, Renkawitz R (2010) CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. Chromosoma 119(4):351–360
Oikawa K et al (2004) Expression of a novel human gene, human wings apart-like (hWAPL), is associated with cervical carcinogenesis and tumor progression. Cancer Res 64(10):3545–3549
Oikawa K et al (2008) Expression of various types of alternatively spliced WAPL transcripts in human cervical epithelia. Gene 423(1):57–62
Oka Y et al (2011) Recruitment of the cohesin loading factor NIPBL to DNA double-strand breaks depends on MDC1, RNF168 and HP1gamma in human cells. Biochem Biophys Res Commun 411(4):762–767

Oliveira RA et al (2010) Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. Nat Cell Biol 12(2):185–192
Opitz JM (1985) The Brachmann-de Lange syndrome. Am J Med Genet 22(1):89–102
Pandey R, Heidmann S, Lehner CF (2005) Epithelial re-organization and dynamics of progression through mitosis in Drosophila separase complex mutants. J Cell Sci 118(Pt 4):733–742
Panizza S et al (2000) Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Curr Biol 10(24):1557–1564
Parelho V et al (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132(3):422–433
Parnas O et al (2009) The ELG1 clamp loader plays a role in sister chromatid cohesion. PLoS One 4(5):e5497
Pauli A et al (2008) Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. Dev Cell 14(2):239–251
Pauli A et al (2010) A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. Curr Biol 20(20):1787–1798
Peters JM, Bhaskara V (2009) Cohesin acetylation: from antiestablishment to establishment. Mol Cell 34(1):1–2
Petronczki M et al (2004) Sister-chromatid cohesion mediated by the alternative RF-CCtf18/Dcc1/Ctf8, the helicase Chl1 and the polymerase-alpha-associated protein Ctf4 is essential for chromatid disjunction during meiosis II. J Cell Sci 117(Pt 16):3547–3559
Potts PR, Porteus MH, Yu H (2006) Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J 25(14):3377–3388
Prieto I et al (2001) Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. Nat Cell Biol 3(8):761–766
Qiu X et al (2008) A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal Igf2 allele. Mol Endocrinol 22(6):1476–1488
Revenkova E et al (2004) Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. Nat Cell Biol 6(6):555–562
Rhodes JM et al (2010) Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. Dev Biol 344(2):637–649
Ritchie K et al (2008) Loss of ATRX leads to chromosome cohesion and congression defects. J Cell Biol 180(2):315–324
Rolef Ben-Shahar T et al (2008) Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science 321(5888):563–566
Rollins RA, Morcillo P, Dorsett D (1999) Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152(2):577–593
Rollins RA et al (2004) Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol 24(8):3100–3111
Roposch A et al (2004) Orthopaedic manifestations of Brachmann-de Lange syndrome: a report of 34 patients. J Pediatr Orthop B 13(2):118–122
Rowland BD et al (2009) Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. Mol Cell 33(6):763–774
Rubio ED et al (2008) CTCF physically links cohesin to chromatin. Proc Natl Acad Sci USA 105(24):8309–8314
Russell KL et al (2001) Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. Am J Med Genet 104(4):267–276
Ryu B et al (2007) Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. PLoS One 2(7):e594
Sakai A et al (2003) Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. EMBO J 22(11):2764–2775

Sataloff RT et al (1990) Cornelia de Lange syndrome. Otolaryngologic manifestations. Arch Otolaryngol Head Neck Surg 116(9):1044–1046
Schaaf CA et al (2009) Regulation of the Drosophila enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. PLoS One 4(7):e6202
Schleiffer A et al (2003) Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. Mol Cell 11(3):571–575
Schmidt D et al (2010) A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 20(5):578–588
Schuldiner O et al (2008) piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. Dev Cell 14(2):227–238
Schule B et al (2005) Inactivating mutations in ESCO2 cause SC phocomelia and Roberts syndrome: no phenotype-genotype correlation. Am J Hum Genet 77(6):1117–1128
Schulz S et al (2008) Prenatal diagnosis of Roberts syndrome and detection of an ESCO2 frameshift mutation in a Pakistani family. Prenat Diagn 28(1):42–45
Selicorni A et al (2005) Anomalies of the kidney and urinary tract are common in de Lange syndrome. Am J Med Genet A 132(4):395–397
Shepard JL et al (2007) A mutation in separase causes genome instability and increased susceptibility to epithelial cancer. Genes Dev 21(1):55–59
Shintomi K, Hirano T (2009) Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. Genes Dev 23(18):2224–2236
Shintomi K, Hirano T (2010) Sister chromatid resolution: a cohesin releasing network and beyond. Chromosoma 119(5):459–467
Sjogren C, Nasmyth K (2001) Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. Curr Biol 11(12):991–995
Skibbens RV (2005) Unzipped and loaded: the role of DNA helicases and RFC clamp-loading complexes in sister chromatid cohesion. J Cell Biol 169(6):841–846
Skibbens RV, Maradeo M, Eastman L (2007) Fork it over: the cohesion establishment factor Ctf7p and DNA replication. J Cell Sci 120(Pt 15):2471–2477
Splinter E et al (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20(17):2349–2354
Strom L et al (2004) Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. Mol Cell 16(6):1003–1015
Sutani T et al (2009) Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. Curr Biol 19(6):492–497
Szabo P et al (2000) Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol 10(10):607–610
Tonkin ET et al (2004) NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet 36(6):636–641
Toth A et al (1999) Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. Genes Dev 13(3):320–333
Tsukahara M et al (1998) Brachmann-de Lange syndrome and congenital heart disease. Am J Med Genet 75(4):441–442
Uhlmann F (2001) Secured cutting: controlling separase at the metaphase to anaphase transition. EMBO Rep 2(6):487–492
Uhlmann F, Lottspeich F, Nasmyth K (1999) Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature 400(6739):37–42
Uhlmann F et al (2000) Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103(3):375–386
Unal E et al (2004) DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Mol Cell 16(6):991–1002
Unal E et al (2008) A molecular determinant for the establishment of sister chromatid cohesion. Science 321(5888):566–569
van der Lelij P et al (2010) Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1. Am J Hum Genet 86(2):262–266

Vega H et al (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet 37(5):468–470
Vega H et al (2010) Phenotypic variability in 49 cases of ESCO2 mutations, including novel missense and codon deletion in the acetyltransferase domain, correlates with ESCO2 expression and establishes the clinical criteria for Roberts syndrome. J Med Genet 47(1):30–37
Vrolik (1849) Tabulae ad illustrandam embryogenesin hominiset mammalium tam naturalem quam abnormem
Waizenegger IC et al (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell 103(3):399–410
Watanabe Y, Nurse P (1999) Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature 400(6743):461–464
Wendt KS et al (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451(7180):796–801
Wygnanski-Jaffe T et al (2005) Ophthalmologic findings in the Cornelia de Lange Syndrome. J AAPOS 9(5):407–415
Xu H et al (2005) Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. Dev Cell 8(6):949–961
Xu H et al (2010) Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. PLoS One 5(8):e12112
Yazdi PT et al (2002) SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. Genes Dev 16(5):571–582
Yoon YS et al (2007) Analysis of the H19ICR insulator. Mol Cell Biol 27(9):3499–3510
Zhang X et al (1999a) Pituitary tumor transforming gene (PTTG) expression in pituitary adenomas. J Clin Endocrinol Metab 84(2):761–767
Zhang X et al (1999b) Structure, expression, and function of human pituitary tumor-transforming gene (PTTG). Mol Endocrinol 13(1):156–166
Zhang B et al (2007) Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development 134(17):3191–3201
Zhang N et al (2008a) A handcuff model for the cohesin complex. J Cell Biol 183(6):1019–1031
Zhang J et al (2008b) Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol Cell 31(1):143–151
Zhang N et al (2008c) Overexpression of Separase induces aneuploidy and mammary tumorigenesis. Proc Natl Acad Sci USA 105(35):13033–13038
Zhang B et al (2009) Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. PLoS One 4(5):e5232
Zou H et al (1999) Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science 285(5426):418–422

Chapter 12
Epigenetics and Human Disease

Angeliki Magklara and Stavros Lomvardas

Abstract The completion of the Human Genome Project has advanced our understanding of the biological processes involved in health and disease. The increasing amount of whole-genome sequencing data becoming available from healthy and affected individuals has pinpointed variations in the DNA sequence, like single-nucleotide polymorphisms (SNPs), that may help to explain differences in phenotype, as well as in disease susceptibility and resistance. On the other hand, it is becoming increasingly apparent that the DNA-stored information alone cannot be the sole determinant of human variation and disease. The extreme phenotypic variability that characterizes the >250 different cell types in the human body, where all cells carry the same genetic information, as well as the high monozygotic discordance rates for human diseases clearly indicate so. Nowadays, it is well established that the epigenome exerts an additional layer of regulation on gene expression and can "manipulate" the same genetic code into producing distinct phenotypes. The epigenome shows far greater plasticity than the genome and contributes significantly to development and differentiation by responding to environmental stimuli. Errors in epigenetic programming caused by genetic defects and/or environmental factors have been directly implicated with human disease. In this chapter, we describe known epigenetic mechanisms and discuss the aberrant epigenetic patterns that characterize several human diseases.

Keywords Epigenetics • Chromatin • DNA methylation • HAT • HDAC • HDM • HMT • miRNAs • DNA hypermethylation • Autoimmune diseases • Systemic lupus erythematosus • Rett syndrome • MeCP2 • Acute lymphoblastic leukemia • Acute myeloid leukemia

A. Magklara • S. Lomvardas (✉)
Department of Anatomy, University of California San Francisco,
1550 4th St., Rock Hall #348D, San Francisco, CA 94158, USA
e-mail: Angeliki.Magklara@ucsf.edu; Stavros.Lomvardas@ucsf.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_12,

12.1 Introduction

For many years, chromatin, the nucleoprotein complex of DNA and histones, was considered to be a non-dynamic structure, whose only role was the compaction and confinement of DNA in the nucleus. However, important research breakthroughs over the last two decades have revealed that chromatin is a primary contributing factor in the regulation of gene expression (Berger 2007), and it has been the focus of intense research in the field of epigenetics.

The word epigenetics (from the Greek word *epi* (*επί*) that means over, above and *genetics*) was first used in 1942 by C.H. Waddington (Waddington 1942) to describe how genes might interact with their environment to produce a phenotype. Today, epigenetics is defined as the study of mechanisms affecting gene expression that do not involve changes in the underlying DNA sequence and that can be inherited through cell division. In recent years, major advances in the understanding of epigenetic mechanisms have established them as key players in several cellular processes including cell differentiation (Mohn and Schubeler 2009), aging (Calvanese et al. 2009), DNA replication (Hiratani and Gilbert 2009), and repair (Huertas et al. 2009). The most common epigenetic mechanisms include DNA methylation, post-translational histone modifications, and small noncoding RNAs. All epigenetic factors are in close interplay and are subject to multiple positive and negative feedback mechanisms; the observed outcome (phenotype) is the result of these interactions. As one would expect, deregulation of these mechanisms is associated with the genesis and progression of several grave human diseases, such as cancer (section 12.2), autoimmune diseases (section 12.3), and neurodevelopmental disorders (section 12.4). The growing list of human disorders with an epigenetic link also includes cardiovascular diseases (Ordovas and Smith 2010; Shirodkar and Marsden 2011), myopathies (Hang et al. 2010), and kidney diseases (Liakopoulos et al. 2011).

12.1.1 DNA Methylation

DNA methylation, the most widely studied epigenetic mechanism in humans, is a covalent modification whereby a methyl group is deposited on the carbon 5 of the cytosine ring using S-adenosyl methionine as the donor. This reaction is catalyzed by the family of DNMT (DNA methyltransferases) enzymes, which is comprised of five members: DNMT1, DNMT2, DNMT3a, DNMT3b, and DNMT3L. DNMT1 is known as the "maintenance methyltransferase"; it preferentially binds to hemimethylated DNA (DNA where one strand is already methylated), and it is used by the cell to maintain the DNA methylation status during semiconservative DNA replication. DNMT3a and DNMT3b, known as "de novo" methyltransferases, can introduce cytosine methylation in previously unmethylated sites and are thought to be responsible for establishing the pattern of methylation during embryonic development (Klose and Bird 2006). DNMT2 and DNMT3L do not possess DNA methyltransferase activity (Bourc'his et al. 2001; Hermann et al. 2003). DNMT2's main

function is to methylate the aspartyl-tRNA (Goll et al. 2006), while DNMT3L binds to and regulates the functions of DNMT3a and DNMT3b (Chen et al. 2005).

DNA methylation occurs almost exclusively in the context of CpG dinucleotides, which are scattered throughout the genome in lower than expected frequencies and are usually methylated. However, CpG nucleotides can also be found clustered in GC-rich regions, known as "CpG islands" (Illingworth and Bird 2009) that frequently localize within promoters or other gene-regulatory elements. Approximately 60% of mammalian gene promoters harbor CpG islands, which are unmethylated in the normal cell (Straussman et al. 2009).

In general, DNA methylation is associated with gene silencing. It plays a central role in several physiological phenomena, such as dosage compensation in humans, maintenance of genomic imprinting, and repression of germline- and tissue-specific genes during early development. Dosage compensation is a regulatory mechanism that ensures the equal expression of X-linked genes both in males (XY) and females (XX). In humans, this is achieved by inactivation of one X chromosome (Xi) in females, thus preventing expression of most genes on this chromosome. The Xi is packaged in compact, repressive heterochromatin, rich in DNA methylation (Mohandas et al. 1981). Genomic imprinting is the differential expression of the two alleles of a gene and is dependent on the parent of origin of the allele, where one allele is silenced early in development via DNA methylation. Finally, DNA methylation seems to be the primary silencing mechanism for some germline-specific genes, such as the MAGE and LAGE gene families that are not expressed in any adult tissue (De Smet et al. 1999). Apart from regulating gene expression, DNA methylation is also critical in protecting genome integrity, through the silencing of repetitive elements that could cause chromosomal instability and gene disruption, if reactivated (Konkel and Batzer 2010).

There are two general mechanisms by which DNA methylation can lead to gene silencing. In the first one, cytosine methylation can directly inhibit transcription by blocking transcription activators from binding to target sites (Kuroda et al. 2009; Watt and Molloy 1988). Alternatively, it can promote the recruitment of methyl-binding domain proteins (MBDs), which are present in transcription corepressor complexes along with other members of the epigenetic machinery, such as histone deacetylases (HDACs) and histone methyltransferases (HMTs), resulting in chromatin remodeling and gene silencing (Nan et al. 1998) (Fig. 12.1). Notably, the DNMTs have also been reported to interact with and recruit such repressive factors (reviewed in Klose and Bird 2006).

12.1.2 Post-translational Histone Modifications

The basic unit of chromatin, the nucleosome, contains 146 base pairs of DNA wrapped around an octamer of core histone proteins, H2A, H2B, H3, and H4. The N-terminal regions of the histones are flexible "tails" that protrude outside of the core nucleosome and can undergo multiple post-translational modifications,

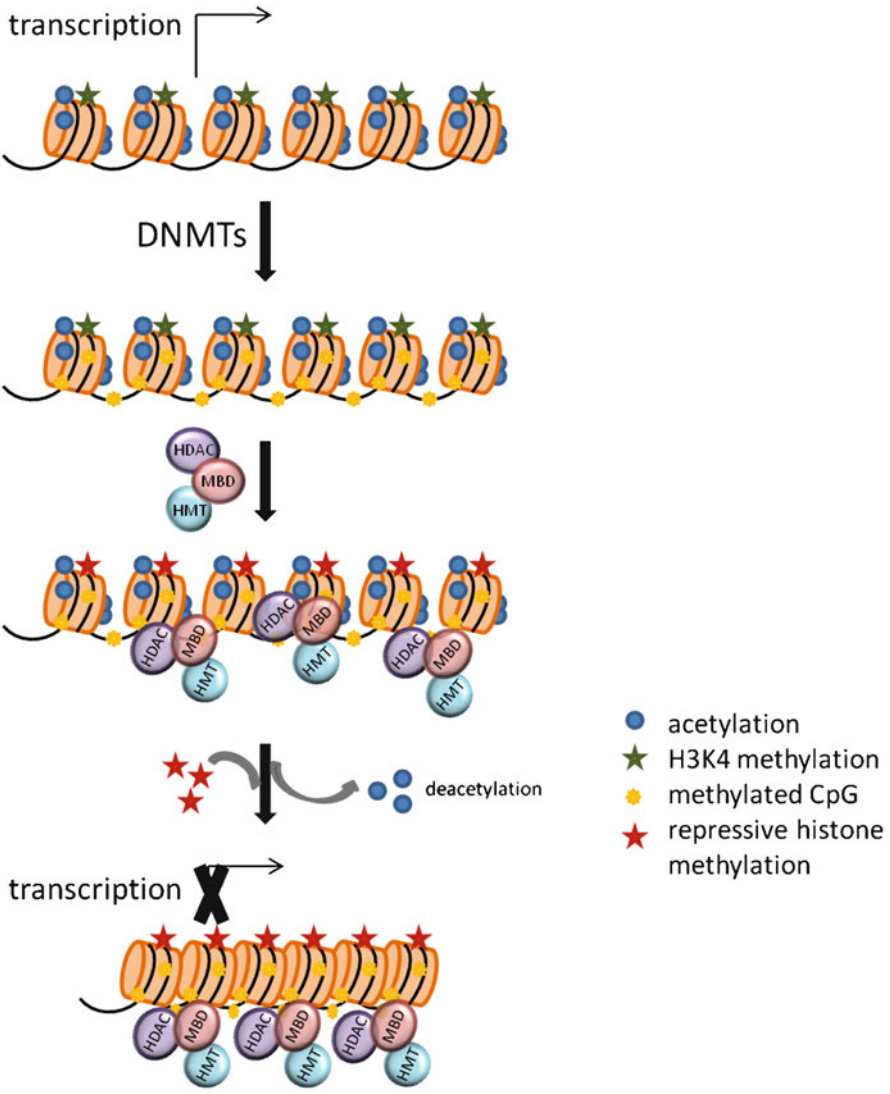

Fig. 12.1 *Chromatin structure.* An active gene has an "open" chromatin structure, where the histones are acetylated and the promoter is enriched in H3K4 methylation. DNMTs' activity leads to DNA methylation and subsequent recruitment of methyl-binding domain proteins (MBDs). The MBDs are found in corepressor complexes along with HDACs and HMTs that are also recruited on the gene, leading to histone deacetylation and methylation with repressive marks (such as H3K27me3 and H3K9me3), respectively. As a consequence, the gene adopts a "closed" chromatin configuration that represses transcription

such as acetylation, methylation, phosphorylation, ubiquitination, sumoylation, deimination, etc. Histones H2A and H2B can also be modified on residues found in their C-terminal tails (Bannister and Kouzarides 2011). Rapid advances in recent years have demonstrated that these modifications provide an important regulatory platform for processes such as gene transcription, DNA replication, and DNA-damage repair (Bannister and Kouzarides 2011). Different sequential or combinatorial patterns of these modifications have been proposed to dictate specific and distinct functional outputs in the genome according to the "histone code" (Strahl and Allis 2000).

Histone modifications are catalyzed by specific sets of specialized enzymes. Acetylation, the most widely studied histone modification, is catalyzed by histone acetyltransferases (HATs), and it occurs at lysine residues, mostly in the tails of histones H3 and H4. Lysine acetylation is associated with transcriptional activity (Verdone et al. 2005), and genome-wide studies show good correlation between hyperacetylation and active promoters and enhancers (Roh et al. 2007). Histone acetylation can regulate gene transcription in two ways. First, the addition of a negatively charged acetyl group destabilizes the interaction between the histone protein and DNA and allows for increased accessibility of transcription factors. Second, it provides a docking site for histone-binding factors that may affect gene expression (Verdone et al. 2005). The levels of histone acetylation depends on the antagonistical function of HATs and histone deacetylases (HDACs) that seem to act in a dynamic fashion both on active and inactive genes (Wang et al. 2009).

Histone methylation is catalyzed by histone methyltransferases (HMTs), while the methyl group can be removed by a recently identified group of enzymes called histone demethylases (HDMs) (Pedersen and Helin 2010). Methylation can occur at several lysine and arginine residues of histones H3 and H4, and unlike acetylation, it does not alter the charge of the histone protein. The fact that lysines can be mono-, di-, or trimethylated and arginines can be mono- or dimethylated (symmetrically or asymmetrically) adds another layer of regulation (Ng et al. 2009). Histone lysine methylation is linked to both transcriptional activation and repression (Martin and Zhang 2005). Genome-wide studies have shown that H3K4me2 and H3K4me3 are strongly enriched at active promoters, while H3K36me3 is elevated in the gene-transcribing regions (Barski et al. 2007); H3K4me1 has been identified as a mark for enhancers (Heintzman et al. 2009). On the other hand, H3K9me2 and H3K27me3 are associated with silenced facultative heterochromatin (Trojer and Reinberg 2007), while H3K9me3 and H4K20me3 are the landmarks of constitutive heterochromatin found mostly on pericentromeric and telomeric repeats (Grewal and Jia 2007). Several protein motifs that are capable of specific interactions with methylated lysine residues have been identified. Proteins that contain these motifs are recruited by specific methylated lysines, and this recruitment step seems to play an important role in the unique biological outcomes that are associated with different methylation events (Martin and Zhang 2005). Thus, histone methylation serves as a molecular mark that signals downstream effects leading to transcriptional activation or repression.

12.1.3 MicroRNAs

MicroRNAs (miRNAs) are 18–24-nucleotide-long noncoding RNA molecules that bind to their target mRNAs, either at a post-transcriptional level leading to their degradation or at a translational level leading to their repression. miRNAs target many genes that play important roles in processes like cell cycle progression, apoptosis, and differentiation (Schickel et al. 2008). A single miRNA can have hundreds of target mRNAs, and each mRNA may be regulated by more than one miRNA, highlighting the implication of this gene regulation system in cellular functions (Lim et al. 2005). The latest release of the miRNA database includes more than 1,400 annotated human miRNAs (http://microrna.sanger.ac.uk; release 17.0). miRNAs are transcribed by RNA polymerase II (Pol II) as long primary transcripts called pri-miRNAs, which, subsequently, are processed by the RNase III enzyme Drosha together with its binding partner DGCR8 into precursor RNAs called pre-miRNAs (70–100 nt in length). Pre-miRNAs are structured as imperfect stem-loops, and they are exported into the cytoplasm by exportin five. The precursor miRNAs are further processed in the cytoplasm by another RNase III called Dicer, along with TRBP, into 18–24-nt-long miRNA duplexes. Finally, these duplexes are loaded into the RNA-induced silencing complex (RISC), where one strand gets degraded, while the other one remains stably associated (mature miRNA) and leads to translational repression of its target mRNAs (reviewed in Bartel 2004). The study of miRNAs has become the subject of intense interest, not only because of their emerging role as master regulators in a diverse and fundamental set of cellular mechanisms, but also because their deregulation has been linked to severe disease states, like cancer (Davalos and Esteller 2010; Schickel et al. 2008) (see 12.2.3).

12.2 Epigenetic Changes in Cancer

In their landmark publication of 2000, Hanahan and Weinberg described the six hallmarks of cancer, providing a foundation for understanding the remarkable diversity of this disease (Hanahan and Weinberg 2000). These include sustaining proliferative signal, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. According to the clonal genetic model of cancer (Nowell 1989), acquisition of these characteristics depends on a succession of genomic alterations that lead to the selective overgrowth of a monoclonal population of tumor cells. However, heritable patterns of disrupted gene expression, for example, inactivation of tumor suppressor genes, can also be acquired through epigenetic mechanisms (Berdasco and Esteller 2010; Jones and Baylin 2007), arguing that cancer is more than a genetic disease. A rapidly growing number of studies in tumor tissues have revealed at least as many epigenetic as genetic alterations for a given gene. These alterations often occur early in tumorigenesis, providing support for the epigenetic progenitor model, which states that “cancer has a fundamentally common basis that is grounded in a

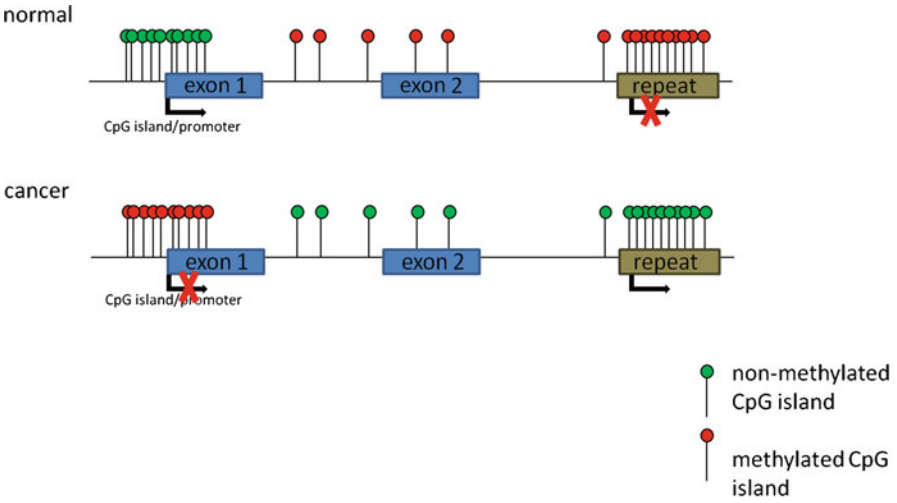

Fig. 12.2 *DNA methylation patterns in normal and cancer cells.* A tumor suppressor gene is given as an example. In a *normal* cell, CpG islands in the promoter of the gene are unmethylated and the gene is expressed. Methylation of CpG islands in the gene body ensures that the gene is not transcribed from other sites, while methylation of repeat elements keeps them repressed. In the *cancerous* state, DNA methylation of the promoter leads to gene silencing, while demethylation of the gene body may lead to aberrant expression. Demethylation of repeat elements allows them to be transcribed, affecting genome stability (for details, see text)

polyclonal epigenetic disruption of stem/progenitor cells" (Feinberg et al. 2006). Whether cancer epigenetic changes have such a profound role in the pathogenesis of the disease or they are just surrogate alterations of mutations remains to be seen. Here, we describe the most abundant epigenetic modifications found in neoplasias and how these may contribute to the tumorigenic potential.

12.2.1 DNA Methylation and Cancer

The cancer epigenome is characterized by site-specific CpG island promoter hypermethylation and genome-wide DNA hypomethylation (Fig. 12.2). Several studies have addressed the question of how alterations in the DNA methylome are triggered in cancer mainly by investigating the expression levels of DNMTs in different malignancies. Overexpression of DNMTs occurs in many cancer types, and it has been associated with hypermethylation of CpG islands, a finding that has not been supported by more recent data (Miremadi et al. 2007). Most studies indicate that there is no significant reduction in the expression levels of DNMTs associated with DNA hypomethylation (Wilson et al. 2007), thus suggesting that disruption of their activity is probably responsible. DNA methylation is a dynamic process that is in close interplay with other genetic and epigenetic factors, such as transcription

factors, chromatin-modifying enzymes, small non-coding RNAs, etc., and one can imagine that any disruption in the activity of such factors may well contribute, directly or indirectly, to targeted or generalized changes in methylation. While the underlying mechanisms that initiate DNA methylation changes are still under investigation, accumulating data indicate that they often occur very early in cancer development and may contribute to cancer initiation.

12.2.1.1 DNA Hypermethylation

The most frequent and most intensely studied epigenetic abnormality in malignant cells is the CpG hypermethylation at the promoters of cancer-associated genes. Table 12.1 records the most commonly hypermethylated genes (for more details, see www.pubmeth.org) in the ten most frequent types of cancer in the US (source National Cancer Institute, USA). These genes include many classical tumor suppressor genes (e.g., APC, PTEN, BRCA1), as well as genes involved in tissue remodeling (e.g., cadherins), DNA repair (e.g., MGMT and MLH1), cell cycle regulation (e.g., CDKN2A and CDKN2B), and apoptosis (e.g., DAPK1 and PYCARD) (for more details on gene function, see Table 12.2). Epigenetic gene silencing by promoter CpG methylation occurs most frequently during the initial stages of tumorigenesis (Esteller 2005), and it is argued that it could predispose cells to the genetic abnormalities that advance the neoplastic process (Feinberg et al. 2006).

A typical example is cyclin-dependent kinase inhibitor 2A (*CDKN2A*), also called p16, an important cell cycle regulator and a known tumor suppressor gene, which is regularly mutated in various types of cancer, but it is also epigenetically regulated and often silenced by promoter DNA methylation (Merlo et al. 1995). p16 is responsible for maintaining the retinoblastoma (Rb) protein in an active and non-phosphorylated state by inhibiting CDK4. It can also bind to Mdm2 p53 binding protein homolog (MDM2), inhibiting its oncogenic action by blocking MDM2-induced degradation of tumor protein p53 (p53) and thus enhancing p53-dependent transactivation and apoptosis. p16 can also induce G2 arrest and apoptosis in a p53-independent manner by preventing the activation of cyclin B1/CDC2 complexes. Loss of p16 expression has been found in preinvasive stages of lung, breast, and colon neoplasia (Baylin and Ohm 2006), and it could allow epithelial cells to escape senescence and start aberrant proliferation resulting in genetic changes that predicate oncogenic evolution. Another example is O-6-methylguanine-DNA methyltransferase (*MGMT*), which encodes a DNA repair protein that removes mutagenic and cytotoxic adducts from O^6-guanine in DNA. Epigenetic silencing of *MGMT* has been documented in a variety of tumor types, where failure to remove the O^6-methylguanine adducts causes G:C to A:T transitions that often affect genes required for genomic stability, such as K-Ras and p53 (Jacinto and Esteller 2007). Similarly, loss of glutathione S-transferase-π1 (GSTP1) expression, an enzyme responsible for detoxifying electrophiles and oxidants, in precancerous prostate lesions and preinvasive prostate tumors may allow for cell and genome damage involved in initiation of carcinogenesis (Nelson et al. 2009).

Table 12.1 The most commonly methylated genes in the ten most frequent types of cancer in the USA (excluding skin cancer). The numbers shown correspond to samples examined in several studies and include both cancer cell lines and tumor samples (data taken from www.pubmeth.org)

Methylation frequency:	0	0-20 %	20-40 %	40-60 %	60-80 %	80-100 %

Cancer / Gene	lung	colorectal	leukemia	breast	prostate	lymphoma	bladder	pancreas	kidney	thyroid
CDKN2A (p16)	4088	4184	1342	811	609	475	865	251	154	162
RASSF1	2610	925	49	593	1050	135	462	140	544	228
MGMT	1475	3184	40	273	443	292	403	154	160	49
CDH1	1015	769	894	745	663	-	477	126	175	0
CDKN2B (p15)	153	393	1837	-	-	438	129	58	-	-
DAPK1	1154	554	731	316	277	145	706	122	268	29
APC	1164	1173	-	750	941	-	205	120	110	-
GSTP1	693	374	0	527	2001	28	424	72	99	-
RARB	1325	91	56	281	933	102	304	98	109	29
MLH1	75	4969	20	218	-	57	160	146	-	-
TIMP3	253	407	-	350	447	28	120	94	152	0
CDH13	1231	159	786	257	280	19	-	33	-	-
ESR1	130	0	259	374	214	-	-	-	-	40
FHIT	1150	40	314	155	101	51	41	-	-	-
RUNX3	336	235	-	106	273	20	181	0	-	-
TP73	-	130	807	0	-	92	31	120	-	-
ESR2	7	-	-	401	61	-	-	-	-	144
PTGS2	7	551	-	258	481	-	105	-	-	-
PYCARD	194	0	587	109	303	21	-	-	-	-
SFRP1	80	105	336	65	41	-	120	75	55	-
BRCA1	98	-	-	850	-	-	-	72	0	-
DLC1	26	0	19	53	27	57	-	-	34	-
PTEN	30	172	587	44	-	-	-	-	-	36
CCND2	80	-	-	397	219	-	18	11	-	-
SCGB3A1	56	0	0	25	21	0	0	17	-	-
RBP1	191	177	53	48	215	90	27	-	-	-
TMEFF2	272	199	-	37	50	20	57	11	-	-
NR0B2	-	-	411	-	-	90	-	-	-	-
IGFBP3	125	56	-	39	-	-	110	-	32	-
CHFR	20	304	41	-	-	21	-	-	-	-
THBS1	-	446	80	148	179	28	-	36	-	-
RPRM	301	0	0	0	-	0	-	140	-	-
SOCS1	40	289	231	148	-	-	105	74	-	-
PGR	7	0	44	148	0	-	-	-	-	-
CADM1	419	0	21	95	-	-	-	91	-	-
HIC1	51	65	37	90	73	-	-	-	-	-
TSHR	-	-	-	-	-	-	-	-	-	120

12.2.1.2 DNA Hypomethylation and Cancer

Aberrant global DNA hypomethylation in human cancer samples was first reported almost 30 years ago (Feinberg and Vogelstein 1983). Since then, this epigenetic alteration has been documented as a frequent event in most malignancies. Interestingly enough, genomic hypomethylation does not associate with overexpression of oncogenes as originally thought, but it is related to the generation of chromosomal instability (see below). DNA hypomethylation appears to be an early event in carcinogenesis; it is, often, evident in the healthy tissue adjacent to the neoplastic, suggesting a role in the initiation of the disease (Wilson et al. 2007).

Table 12.2 Name and function of the genes shown in Table 12.1. Most of the genes that are commonly methylated in cancer are involved in cell cycle regulation, DNA repair, apoptosis, and tissue remodeling

Symbol	Official gene name	Function
APC	Adenomatous polyposis coli	Modulator of Wnt signaling
BRCA1	Breast cancer 1, early onset	DNA repair, tumor suppressor
CADM1	Cell adhesion molecule 1	Cell adhesion
CCND2	Cyclin D2	Cell cycle
CDH1	Cadherin 1, E-cadherin	Cell adhesion
CDH13	Cadherin 13, H-cadherin	Cell adhesion
CDKN2A	Cyclin-dependent kinase inhibitor 2A (p16, p14ARF)	Cell cycle
CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15)	Cell cycle
CHFR	Checkpoint with forkhead and ring finger domain	Cell cycle, candidate tumor suppressor
DAPK1	Death associated protein kinase 1	Apoptosis
DLC1	Deleted in liver cancer 1	Signal transduction, tumor suppressor
ESR1	Estrogen receptor 1	Transcription factor, hormone regulation
ESR2	Estrogen receptor 2	Transcription factor, hormone regulation
FHIT	Fragile histidine triad gene	Purine metabolism, candidate tumor suppressor
GSTP1	Glutathione S-transferase pi 1	Enzyme, DNA repair
HIC-1	Hypermethylated in cancer 1	Transcriptional repressor, apoptosis; candidate tumor suppressor
IGFBP3	Insulin-like growth factor binding protein 3	Binding protein
MGMT	O-6-methylguanine-DNA methyltransferase	DNA repair
MLH1	MutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	DNA repair
NR0B2	Nuclear receptor subfamily 0, group B, member 2	Transcription factor
PGR	Progesterone receptor	Transcription factor, hormone regulation
PTEN	Phosphatase and tensin homolog	Signal transduction, tumor suppressor
PTGS2	Prostaglandin-endoperoxide synthase 2 (Cox-2)	Inflammation
PYCARD	PYD and CARD domain containing	Apoptosis
RARB	Retinoic acid receptor, beta	Transcription factor, inhibition of cell growth
RASSF1A	Ras associated domain family 1	Cell cycle, apoptosis; tumor suppressor
RBP1	Retinol binding protein 1, cellular	Binding protein
RPRM	Reprimo, TP53-dependent G2 arrest mediator candidate	Cell cycle
RUNX3	Runt-related transcription factor 3	Transcription factor, tumor suppressor
SCGB3A1	Secretoglobin, family 3A, member 1	Cytokine, inhibition of cell growth
SFRP1	Secreted frizzled-related protein 1	Modulator of Wnt signaling
SOCS1	suppressor of cytokine signaling 1	Regulator of cytokine signaling
THBS1	Thrombospondin 1	Adhesive glycoprotein, angiogenesis

(continued)

Table 12.2 (continued)

Symbol	Official gene name	Function
TIMP3	TIMP metallopeptidase inhibitor 3	Cell migration/invasion
TSHR	Thyroid stimulating hormone receptor	Binding protein
TMEFF2	Transmembrane protein with EGF-like and two follistatin-like domains 2	Signal transduction
TP73	Tumor protein p73	Transcription factor, candidate tumor suppressor

The degree of global hypomethylation increases through all the tumorigenic steps, from the benign proliferations to the invasive cancers (Fraga et al. 2004).

DNA hypomethylation occurs predominantly at repetitive sequences and, to a lesser extent, at gene bodies and leads to a 20–60% reduction of the 5-methylcytosine content of cancer tissue comparing to its normal counterpart. It is believed to contribute to carcinogenesis, mainly by promoting genomic instability through destabilization of pericentromeric repeats and/or reactivation of transposable elements. Repeat elements include simple repeat sequences, such as DNA satellites that are found in pericentromeric and subtelomeric heterochromatin, and transposable elements (DNA transposons, retrotransposons, and endogenous retroviruses).

The vast majority of repeat elements are silenced in normal somatic cells via dense DNA methylation. Demethylation of pericentromeric repeats can lead to increased chromosomal rearrangements, mitotic recombination, and aneuploidy (Eden et al. 2003; Karpf and Matsui 2005), and it is a frequent finding in a variety of malignancies, including Wilms' tumor (a nephroblastoma that typically occurs in children) and ovarian and breast cancer (Wilson et al. 2007). Demethylation of normally dormant transposons and endogenous retroviruses can potentially lead to reactivation of the strong promoters associated with them, altering global transcription and/or modifying the expression of critical growth-regulatory genes in which these elements reside (Wilson et al. 2007). Moreover, transposon demethylation and their subsequent reactivation and transcription can cause aberrant chromosomal recombination and translocation, thus further disrupting the genome (Esteller 2008; Howard et al. 2008; Schulz et al. 2006). Hypomethylation of Long Interspersed Nuclear Elements (LINEs), a class of retrotransposons, has been observed in several types of cancer, both as an early event (e.g., prostate and colon cancer), as well as in advanced stages (e.g., breast, ovarian, and leukemia), where it correlates with poor prognosis (Wilson et al. 2007).

Gene-specific hypomethylation is less frequent and is usually associated with growth-regulatory genes, enzymes, and developmentally critical and tissue-specific genes, such as germ-cell-specific tumor antigen genes (the MAGE, BAGE, LAGE, and GAGE gene families) (Wilson et al. 2007). Activation of oncogenes due to DNA hypomethylation, such as of R-Ras in gastric cancer, has also been reported

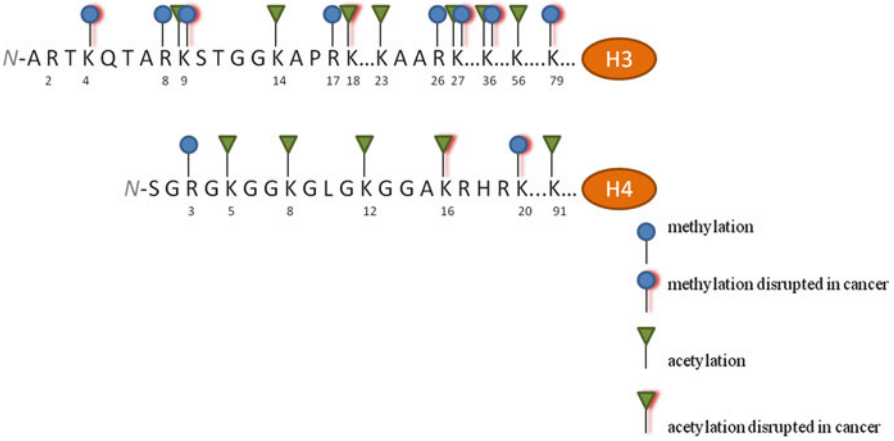

Fig. 12.3 *Methylated and acetylated residues in histones H3 and H4 in normal and cancerous cells.* The NH2-terminal tails of histones *H3 and H4* are depicted and the residues that are commonly known to be acetylated and/or methylated. The modifications that are disrupted in cancer are highlighted

(Nishigaki et al. 2005). Promoter demethylation and subsequent gene activation are often associated with histological grade and/or stage of cancer, for example, cadherin 3, type 1, P-cadherin (*CDH3*) promoter demethylation and P-cadherin expression in invasive breast cancer (Paredes et al. 2005); cyclin D2 activation at advanced stages of gastric cancer (Oshimo et al. 2003); activation of synuclein γ in a range of aggressive, solid tumors (Liu et al. 2005); and elevated maspin expression in high tumor grade colorectal cancer (Bettstetter et al. 2005).

Gene-specific DNA hypomethylation can also lead to aberrant expression of imprinted genes. Loss of imprinting (LOI) has been associated with cancer development in a mouse model (Holm et al. 2005). One of the better studied examples of LOI is that of the insulin-like growth factor 2 (*IGF2*) gene, which was first described in Wilms' tumor (Ogawa et al. 1993; Rainier et al. 1993) but has also been reported in other types of cancer, including colorectal, ovarian, and lung (Feinberg 2004). LOI of *IGF2* results in its pathological biallelic expression that can, potentially, support tumor growth.

12.2.2 Histone Modifications in Cancer

Global loss of monoacetylation at K16 and trimethylation at K20 of histone H4 is a common hallmark of human cancer cells (Fraga et al. 2005). Gene-specific loss of the active mark H3K4me3 and gain of the repressive marks H3K9me3 and H3K27me3 have also been described (Portela and Esteller 2010). Figure 12.3 depicts

the methylated and acetylated residues in histones H3 and H4 in normal cells and the ones that are commonly disrupted in cancer.

In contrast to DNA methylation, where the responsible enzymes (DNMTs) are hardly found mutated in cancer, there is a growing list of alterations in histone-modifying enzymes in specific tumor types (Table 12.3). Mutations in a number of HATs have been observed in solid tumors, while several of them are also involved in chromosomal translocations in hematological malignancies. It appears that these translocations are involved in through aberrant acetylation caused by mistargeting of HATs (reviewed in Miremadi et al. 2007). As Table 12.3 indicates, chromosomal translocations are a common theme in hematological malignancies, whereas solid tumors are more commonly associated with point mutations, deletions, and gene amplification.

The mixed-lineage leukemia 1 gene (*MLL1*), which encodes a well-studied H3K4 HMT, is often implicated in translocations both in acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). It can be found fused to more than 50 distinct partners. The region of the protein that contains the methyltransferase activity is lost in the fusion protein; however, several fusion partners are HMTs themselves (Daser and Rabbitts 2005). MLL1 controls the expression of the *HOX* genes, a group of transcription factors involved in embryonic development and hematopoietic cell differentiation. Several *MLL1* fusions directly recruit DOT1-like, histone H3 methyltransferase (DOT1L), an H3K79 HMT, which activates leukemia-promoting oncogenes, such as homeobox A9 (*Hoxa9*; Chi et al. 2010). Furthermore, a plethora of histone demethylases and effector proteins that "read" specific histone modifications have been reported to have altered expression levels in a variety of cancers (reviewed in Chi et al. 2010), leading to aberrant epigenetic regulation and consequently tumorigenesis.

12.2.3 miRNAs in Cancer

Changes in miRNA expression between normal and tumor specimens can be attributed to a number of reasons: impairment in the miRNA processing machinery; localization in regions of chromosomal instability or nearby chromosomal breakpoint; regulation by tumor suppressor or oncogenic pathways, such as TP53, MYC, and RAS; or changes in their epigenetic regulation (Farazi et al. 2011). The first such study reported that DNA demethylation activated the expression of *mir-127,* a potential tumor suppressor, in bladder cancer cells (Saito et al. 2006). Since then, a plethora of miRNAs have been identified that are aberrantly methylated in several types of cancer leading to deregulation of their target genes (Berdasco and Esteller 2010; Farazi et al. 2011). Consequently, miRNAs can function as oncogenes or tumor suppressors. For example, DNA hypermethylation of *mir-129-2* leads to overexpression of the SRY (sex-determining region Y)-box 4 (*SOX4*) oncogene in endometrial cancer (Huang et al. 2009). In an interesting study, isolation of a subset

Table 12.3 The most common HATs and HMTs that are found altered in different types of cancer

Gene name	Substrate specificity	Deregulation	Tumor type
Histone acetyltransferases (HATs)			
CREBBP (KAT3A)	H2AK5, H2BK12 and K15, H3K14 and K18, H4K5 and K8	Mutations, deletions translocations	Esophageal, lung AML
EP300 (KAT3B)	H2AK5, H2BK12 and K15	Various mutations translocations	Breast, colon, gastric AML
pCAF (KAT2B)	H3K9, K14 and K18	Mutation	epithelial
MYST3 (KAT6A)	H3K14, H4K16	Translocation	AML
MYST4 (KAT6B)	H3K14, H4K16	Translocations	AML, uterine
NCOA1 (KAT13A)	H3 and H4	translocation	rhabdomyosarcoma
NCOA3 (KAT13B)	H3 and H4	Amplification	breast
Histone methyltransferases (HMTs)			
G9a (KMT1C)	H3K9	Overexpression	HCC
MLL1 (KMT2A)	H3K4	Multiple translocations	AML, ALL
MLL2 (KMT2B)	H3K4	Mutation	Kidney
MLL3 (KMT2C)	H3K4	Mutations deletion	Colorectal leukemia
MLL4 (KMT2D)	H3K4	Amplification	Solid tumor cell lines
SETD2 (KMT3A)	H3K36	Mutations underexpression	cRCC breast
NSD1 (KMT3B)	HK36, H4K20	CpG hypermethylation translocation	Neuroblastoma, glioma AML
SMYD2 (KMT3C)	H3K36	Amplification	Esophageal
SMYD3 (KMT3E)	H3K4	Overexpression	Colorectal, HCC breast
DOT1L (KMT4)	H3K79	Translocation	AML
EZH2 (KMT6)	H3K27	Amplification overexpression mutation	Breast, prostate endometrial, colorectal, HCC, melanoma lymphoma
RIZ1 (KMT8)	H3K9	Mutations CpG hypermethylation	Solid tumors breast, liver
PRMT1	H4R3	Underexpression	Breast
PRMT5	H3R8, H4R3	Overexpression	Gastric

AML acute myeloid leukemia, *ALL* acute lymphoblastic leukemia, *HCC* hepatocellular carcinoma, *cRCC* clear cell renal cell carcinoma (for a complete list of histone-modifying enzymes that are aberrantly expressed in cancer, see Berdasco and Esteller (2010), Chi et al. (2010), Miremadi et al. (2007) and references therein)

of highly tumorigenic breast cancer cells showed that they had marked reduction of *let-7* family members and that expression of *let-7* could lead to reduced proliferation, tumor formation, and metastasis (Yu et al. 2007). Currently, miRNAs are under intense investigation for their potential diagnostic, prognostic, and therapeutic use in the field of cancer.

12.3 Epigenetics and Autoimmune Diseases

Self-tolerance is necessary for appropriate immune function; at times, the immune system goes awry and attacks the body itself, resulting into misdirected immune responses that are referred to as autoimmunity and can be demonstrated by the presence of autoantibodies or T lymphocytes reactive with host antigens. Autoimmunity can be the cause of a broad spectrum of human illnesses, known as autoimmune diseases (AID), which are determined by both genetic influences and environmental triggers. Several AID, like rheumatoid arthritis, multiple sclerosis, and type I diabetes, seem to be mediated, at least partly, by environmentally induced epigenetic changes (reviewed in Brooks et al. 2010; Fernandez-Morera et al. 2010). Altered epigenetic patterns can lead to aberrant gene expression in specific cell populations, impairing self-tolerance; such cells might contribute to the development of autoimmunity in genetically predisposed individuals (Fernandez-Morera et al. 2010). Here, we discuss the epigenetic alterations observed in systemic lupus erythematosus (SLE), the most studied paradigm of epigenetic contribution to AID.

12.3.1 Systemic Lupus Erythematosus

SLE is a chronic autoimmune inflammatory disease with numerous clinical and immunological manifestations. It is characterized by the production of antibodies against various nuclear components, which cause inflammation and injury of multiple organs, mainly the skin, joints, kidneys, blood vessel walls, and nervous system. It primarily affects women in their reproductive age and ethnic groups of Asian or African ancestry. SLE is a complex, multifactorial disease, but its precise pathogenesis is unclear. Certain cytokine patterns (like overexpression of the type I interferon pathway) and abnormal signal transduction pathways (e.g., decreased expression of T cell receptor ζ chain and protein kinase C) have been linked to the development of SLE. However, growing evidence points to defects in apoptosis and in the clearance of apoptotic cells as the basis of the pathogenesis of the disease. These defects contribute to the release of, normally intracellular, nuclear components (including nucleosomes, DNA, and histones), triggering an autoimmune response and formation of autoantibodies that cause tissue damage in patients with lupus. This abnormal cellular and humoral response is modulated by genetic, environmental, and hormonal factors. Several susceptibility loci have been identified,

including genes of the major histocompatibility complex (MHC), null alleles that cause deficiency of one of the early complement components (C1q, C2 or C4), a single nucleotide polymorphism within the programmed cell death 1 (*PDCD1*) gene, and several genes on the long arm of chromosome 1q23-24. Among the environmental factors linked to lupus, sunlight is the most prominent one, while many drugs (like procainamide, hydralazine, and quinidine) can cause a variant of lupus, called drug-induced lupus, with manifestations commonly in the skin and joints of patients. Since 90% of patients with SLE are female, an important role for female hormones seems likely; however, it is unclear how sex hormones could promote lupus (D'Cruz et al. 2007; Rahman and Isenberg 2008).

A causal effect between epigenetics and SLE has not yet been established; nevertheless, a significant body of evidence links aberrant epigenetic changes to the onset of the disease. The fact that SLE is characterized by the production of autoantibodies against chromatin (the "carrier" of epigenetic information) adds another layer of interest in the study of the epigenome in affected individuals.

12.3.2 DNA Methylation in Systemic Lupus Erythematosus

Early studies showed that T cells from patients with active lupus exhibited globally hypomethylated DNA (Richardson et al. 1990), and more recent reports have established gene-specific DNA demethylation as a common feature of the disease. Among the first identified genes that were aberrantly overexpressed due to promoter hypomethylation were perforin 1 (*PRF1)*, *CD70* (*TNFSF7*) and integrin, alpha L (*ITGAL* or also called *CD11a*) (reviewed in Ballestar et al. 2006), all implicated in the autoreactivity of SLE T cells (see below). A recent study examining monozygotic twins discordant for the disease identified a new set of 49 differentially methylated genes, most of which were implicated in immune response, cell activation, or response to external stimuli (Javierre et al. 2010). Several of those genes (*IFGNR2, MMP14, LCN2, CSF3R,* and *AIM2*) were hypomethylated and overexpressed in the affected siblings and had been previously associated with SLE.

The molecular basis of DNA hypomethylation is not clearly defined, but several reports indicate that there are multiple mechanisms involved. Different studies have yielded conflicting results regarding the transcript levels of various DNMTs in SLE; one study showed DNMT1 and 3a downregulation in GD4+ T cells (Januchowski et al. 2008), while another failed to confirm such a pattern (Balada et al. 2008). Impaired activity of PKCδ, described in SLE T cells, causes decreased ERK pathway signaling (Gorelik et al. 2007), which has been associated with decreased DNMT1 activity (Deng et al. 2003). The growth arrest and DNA-damage-inducible protein alpha (GADD45a) is involved in DNA demethylation (Barreto et al. 2007). A recent study demonstrated that SLE CD4+ T cells overexpress *GADD45a* and its mRNA levels were inversely proportional to the levels of DNA methylation, while they correlated with CD11a/CD70 mRNA levels (Li et al. 2010). Finally, it was recently reported that DNA hypomethylation in SLE can be mediated by *miR-21*

and *miR-148α* that directly and indirectly target DNMT1 (Pan et al. 2010). Notably, drugs like procainamide and hydralazine that can induce a lupus-like disease have been shown to function as DNA demethylating agents, providing further evidence that DNA methylation changes play an important role in the development of the disease (reviewed in Richardson 2003).

How can DNA hypomethylation cause SLE? Even though the answer is not yet clear, it seems that DNA hypomethylation promotes $CD4^+$ T cell autoreactivity, potentially contributing to the development of the autoimmune disease (reviewed in Ballestar et al. 2006; Richardson 2003). Normally, $CD4^+$ T cells respond to peptides presented by MHC molecules on antigen-presenting cells, but demethylated $CD4^+$ T cells lose this requirement and can respond to these cells without the appropriate antigen. This autoreactivity correlates with increased expression of adhesion molecule lymphocyte function-associated antigen 1 (*ITGAL* or also called *LFA-1*, composed of cluster of differentiation (CD) 11a and CD18 subunits), caused by increased expression of *CD11a* due to promoter hypomethylation, as described above. LFA-1 is an adhesion molecule that surrounds the T cell antigen receptor (TCR) to form the "immunologic synapse," providing both stability to the TCR–MHC interaction and co-stimulatory signals that activate T cells. Increased LFA-1 expression can lead to stabilization of lower affinity interactions between the TCR and MHC molecules bearing inappropriate antigens and increased co-stimulatory signaling, which may be responsible for initiating the T cell autoreactivity. Furthermore, demethylated $CD4^+$ T cells are capable of killing autologous or syngeneic macrophages (Mø) and stimulating B cells and the subsequent release of antigenic apoptotic material could lead to the production of autoantibodies.

12.3.3 Post-translational Histone Modifications in Systemic Lupus Erythematosus

There is little information on the role of histone modifications in SLE both on the global and gene-specific scale. Aberrant patterns of global histone modifications (H3 and H4 hypoacetylation and H3K9 hypomethylation) were observed in $CD4^+$ T cells in SLE patients (Hu et al. 2008). The TNF alpha locus was found to be highly acetylated and more transcriptionally active in SLE monocytes than controls (Sullivan et al. 2007). The histone deacetylase inhibitor trichostatin A (TSA) reverses the aberrant expression of CD40L, IL-10, and IFN-γ in human SLE T cells (Mishra et al. 2001); however, it was not shown if this was a direct or indirect effect. Since these genes play important roles in the immune system, it was suggested that TSA could be a potential candidate for the treatment of SLE (Mishra et al. 2001).

As mentioned above, SLE is characterized by autoantibodies against nucleosomes that are released from apoptotic cells, are not efficiently cleared, and are present in the circulation and tissues. During apoptosis, chromatin can be modified, and several apoptosis-induced modifications have been described, including phosphorylation of serine 14 on histone H2B (Cheung et al. 2003), phosphorylation of threonine

45 (Hurd et al. 2009), and methylation of lysine 27 on histone H3 (Cheng et al. 2009). Interestingly, several apoptosis-associated histone modifications have been identified in SLE, such as specific acetylation of histones H4, H2A, and H2B (Dieker et al. 2007; van Bavel et al. 2009) and methylation of H3K27 (van Bavel et al. 2011), as well as autoantibodies that target them.

12.4 Epigenetics and Neurodevelopmental Disorders

Impairment during the development and growth of the central nervous system can cause a broad range of abnormalities that affect brain functions like learning ability, emotions, and memory. These are collectively known as neurodevelopmental disorders, and they include autism and autism-spectrum disorders (such as Angelman, Prader-Willi, Rett, and Fragile-X syndromes), speech and language disorders, attention-deficit hyperactivity disorder, traumatic brain injuries, and others. The autism-spectrum disorders are characterized by varying degrees of impairment in communication skills and social interactions, as well as restricted, repetitive, and stereotyped patterns of behavior. They have a multifactorial etiology that involves a complex genetic and environmental background (reviewed in Eapen 2011). Current research suggests that epigenetic mechanisms may be involved in the variability in behavior and neurological status of different patients (reviewed in Grafodatskaya et al. 2010). Here, we discuss Rett syndrome (RTT), a well-studied autism-spectrum disorder that is caused by mutations in methyl-CpG binding protein 2 (*MECP2*), a gene located on the X chromosome that encodes for a protein that binds to methylated DNA. Consequently, there are at least two different epigenetic components implicated in the disease. First, the epigenetic process that leads to X chromosome inactivation (XCI) directly regulates MeCP2 expression; random inactivation of the X chromosome that carries the normal allele will lead to the development of RTT in female carriers of MeCP2 mutations. Second, MeCP2 itself is a global epigenetic regulator by binding to a widespread epigenetic mark (methylated DNA) (see 12.4.2).

12.4.1 Rett Syndrome

Rett syndrome (RTT) was first described in 1966 by the homonymous Austrian doctor, but it was not for another 30 years before its genetic and epigenetic basis was discovered. RTT is estimated to affect one in every 10,000–15,000 live female births in all racial and ethnic groups worldwide (source: NINDS/NIH). RTT's clinical manifestations appear progressively in female infants after 6–18 months of age. One of the first clinical features involves deceleration of head growth (microcephaly), which is followed by general growth retardation, weight loss, and muscle hypotonia. Later on, patients lose purposeful hand movements and verbalization skills and exhibit social withdrawal and other autistic features, like expressionless

face and diminished eye contact (reviewed in Chahrour and Zoghbi 2007). In parallel, other physical symptoms develop including apraxia, breathing abnormalities, seizures, and scoliosis and are accompanied by the onset of mental deterioration. Between the ages of 2 and 10, the disease reaches a plateau phase, which can last for years, and for many patients, till the end of their lives. Even though apraxia and motor problems remain grave at this stage, there is an improvement in behavior, with less autistic-like features and increased alertness and social awareness. However, many patients, as they age, progress to a late motor deterioration stage, characterized by severely reduced mobility, often leading to inability to walk, advanced scoliosis, and muscle weakness (source: NINDS/NIH and OMIM #312750).

Nearly all cases of RTT are caused by *de novo* mutations in the X-linked gene that encodes for *MECP2* (see 12.4.2) (Amir et al. 1999). Most of these mutations (~70%) are C-T transitions at eight specific CpG dinucleotides that lead to truncated, partially functional protein or loss of function and are associated with the more severe clinical manifestations of the disease. Small C-terminal deletions occur in about 10% of patients and are associated with a milder phenotype (Smeets et al. 2005).

MECP2 mutations that cause typical RTT in females usually lead to infantile encephalopathy and death in the first year of life in males with normal karyotype. In males, all brain cells will express the mutant *MECP2* X-linked allele, while females with an *MECP2* mutation are typically mosaic, since due to random XCI, half of their cells will express the mutant allele and the other half will express the normal one. Notably, there are exceptions from this rule. Males that carry an extra X chromosome (Klinefelter syndrome) or with somatic mutations of *MECP2* develop a typical RTT phenotype (Clayton-Smith et al. 2000; Maiwald et al. 2002). Rarely, in females, skewed XCI patterns can cause more or less severe phenotypes with wide variability, depending on the direction and the degree of the skewing (Christodoulou and Weaving 2003).

12.4.2 MeCP2(Methyl-CpG Binding Protein 2)

MeCP2 is expressed in a wide variety of tissues, but appears to be most abundant in the brain and primarily in mature post-migratory neurons, where it is speculated to play a role in neuronal activity or plasticity. It is a member of the methyl-CpG binding protein family (Hendrich and Bird 1998) and consists of four functional domains: the methyl-CpG binding domain (MBD), which occupies ~100 amino acids in the N-terminus, the transcriptional repression domain (TRD), a C-terminal domain, and a highly conserved nuclear localization signal (NLS). The MBD shows strong preference for binding to methylated CpG residues in vitro (Nan et al. 1993), and this was confirmed by the binding of the protein to mouse heterochromatic foci in vivo that are known to be heavily methylated (Nan et al. 1996). Chromatin immunoprecipitation (ChIP) assays have demonstrated MeCP2 binding to several methylated promoters as expected; however, MeCP2 binding to non-methylated loci has also been reported (reviewed in Guy et al. 2011). Given the well-established role of DNA methylation in transcriptional repression, it was reasonable to assume

that MeCP2 would mediate gene silencing. It was first demonstrated by in vitro experiments that MeCP2 could function as transcriptional repressor of methylated genes (Nan et al. 1997), and later, it was shown that this was achieved through the recruitment of the transcriptional corepressor Sin3A and HDACs 1 and 2 to the TRD (Nan et al. 1998). Interestingly enough, nowadays, the growing list of interacting partners of MeCP2 includes not only repressors (e.g., N-CoR, c-Ski) but also activators (e.g., CREB), DNA (DNMT1) and histone (Suv39H1) methyltransferases, chromatin remodeling (Brahma), RNA splicing (YB1), and other transcription factors (reviewed in Chahrour and Zoghbi 2007; Guy et al. 2011). Even though the functional implications of the above findings are not clear yet, they suggest that MeCP2 does not have a global transcriptional repressor role, as it was initially thought, but is rather a multifunctional protein involved in diverse nuclear processes. This idea is also corroborated by transcriptional profiling studies. An early expression study in brain tissue from a mouse model of *Mecp2* identified only subtle gene expression changes (Tudor et al. 2002), while more recent work in the mouse hypothalamus and cerebellum found that the majority of the genes affected were downregulated in the absence of the protein and upregulated when it was overexpressed (Chahrour et al. 2008).

Several neuronal-specific genes have been described as targets of MeCP2 including brain-derived neurotrophic factor (*BDNF*), an important signaling molecule in brain development and plasticity, the imprinted genes distal-less homeobox 5 and 6 (*DLX5* and *DLX6*) that encode for neuronal transcription factors, and the paternally brain-imprinted ubiquitin protein ligase E3A (*UBE3A*). However, with the exception of *BDNF*, independent studies have yielded contradictory results as to whether MeCP2 regulates the expression of these genes (Guy et al. 2011). An alternative view on the role of MeCp2 was put forward by a recent study that found that MeCP2 binds wherever DNA methylation occurs, suggesting that it is not a gene-specific regulator, but it may be required to reduce aberrant transcriptional events, thus allowing the transcriptional machinery to function efficiently (Skene et al. 2010).

In summary, the biological function(s) of MeCP2 are still under investigation. The fact that it is expressed mainly in postmitotic neurons along with the postnatal onset of RTT in affected individuals and MeCP2 mouse models supports the idea that MeCP2 plays a key role in the maturation and plasticity of neurons. Recent studies demonstrating that neurological abnormalities resulting from loss of MeCP2 can be reversed upon restoration of endogenous protein production hold great promise for the development of therapies for RTT in the near future (Guy et al. 2007).

12.5 Summary

In the last decade, we have witnessed the emergence of a new biological code, the "epigenetic code," as an equally important determining factor of phenotypic variation in health and disease. With the development and application of new powerful technologies, the field of epigenomics has revealed distinct epigenetic profiles in

different cell types, as well as numerous epigenetic aberrations in a growing number of human disorders. The elucidation of the epigenome and the mechanisms that govern it can help us better understand the interaction between the genome and the environment and how this contributes to the genesis and progression of disease. More importantly, the fact that the epigenetic marks are reversible makes them perfect targets for the development of therapeutic schemes that aim to reestablish the normal epigenetic landscape and opens up new promising possibilities for the fight against these diseases.

Abbreviations

AID	Autoimmune diseases
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
BDNF	Brain-derived neurotrophic factor
CD	Cluster of differentiation
CDH3	Cadherin 3 type 1, P-cadherin
CDKN2A	Cyclin-dependent kinase inhibitor 2A
ChIP	Chromatin immunoprecipitation
DLX5	Distal-less homeobox 5
DLX6	Distal-less homeobox 6
DNMT	DNA methyltransferases
DOT1L	DOT1-like histone H3 methyltransferase
GADD45a	Growth arrest and DNA-damage-inducible protein alpha
GSTP1	Glutathione *S*-transferase-π1
HATs	Histone acetyltransferases
HDAC	Histone deacetylases
HDMs	Histone demethylases
HMT	Histone methyltransferases
Hoxa9	Homeobox A9
IGF2	Insulin-like growth factor 2
ITGAL	Integrin alpha L
LINEs	Long interspersed nuclear elements
LOI	Loss of imprinting
MBD	Methyl-binding domain
MDM2	Mdm2 p53 binding protein homolog
MECP2	Methyl-CpG binding protein 2
MGMT	O-6-methylguanine-DNA methyltransferase
MHC	Major histocompatibility complex
miRNAs	MicroRNAs
MLL1	Mixed-lineage leukemia 1 gene
Mø	Macrophages
NLS	Nuclear localization signal

nt	Nucleotides
p53	Tumor protein p53
PDCD1	Programmed cell death 1
Pol II	RNA polymerase II
PRF1	Perforin 1
RISC	RNA-induced silencing complex
RTT	Rett syndrome
SLE	Systemic lupus erythematosus
SNPs	Single-nucleotide polymorphisms
SOX4	SRY (sex-determining region Y)-box 4
TCR	T cell antigen receptor
TRD	Transcriptional repression domain
TSA	Trichostatin A
UBE3A	Ubiquitin protein ligase E3A
XCI	X chromosome inactivation
Xi	X inactivation

References

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23:185–188

Balada E, Ordi-Ros J, Serrano-Acedo S, Martinez-Lostao L, Rosa-Leyva M, Vilardell-Tarres M (2008) Transcript levels of DNA methyltransferases DNMT1, DNMT3A and DNMT3B in CD4+ T cells from patients with systemic lupus erythematosus. Immunology 124:339–347

Ballestar E, Esteller M, Richardson BC (2006) The epigenetic face of systemic lupus erythematosus. J Immunol 176:7143–7147

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21:381–395

Barreto G, Schafer A, Marhold J, Stach D, Swaminathan SK, Handa V, Doderlein G, Maltry N, Wu W, Lyko F, Niehrs C (2007) Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature 445:671–675

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Baylin SB, Ohm JE (2006) Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6:107–116

Berdasco M, Esteller M (2010) Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell 19:698–711

Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447:407–412

Bettstetter M, Woenckhaus M, Wild PJ, Rummele P, Blaszyk H, Hartmann A, Hofstadter F, Dietmaier W (2005) Elevated nuclear maspin expression is associated with microsatellite instability and high tumour grade in colorectal cancer. J Pathol 205:606–614

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539

Brooks WH, Le Dantec C, Pers JO, Youinou P, Renaudineau Y (2010) Epigenetics and autoimmunity. J Autoimmun 34:J207–J219

Calvanese V, Lara E, Kahn A, Fraga MF (2009) The role of epigenetics in aging and age-related diseases. Ageing Res Rev 8:268–276
Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 320:1224–1229
Chahrour M, Zoghbi HY (2007) The story of Rett syndrome: from clinic to neurobiology. Neuron 56:422–437
Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F (2005) Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem 95:902–917
Cheng MF, Lee CH, Hsia KT, Huang GS, Lee HS (2009) Methylation of histone H3 lysine 27 associated with apoptosis in osteosarcoma cells induced by staurosporine. Histol Histopathol 24:1105–1111
Cheung WL, Ajiro K, Samejima K, Kloc M, Cheung P, Mizzen CA, Beeser A, Etkin LD, Chernoff J, Earnshaw WC, Allis CD (2003) Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. Cell 113:507–517
Chi P, Allis CD, Wang GG (2010) Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 10:457–469
Christodoulou J, Weaving LS (2003) MECP2 and beyond: phenotype-genotype correlations in Rett syndrome. J Child Neurol 18:669–674
Clayton-Smith J, Watson P, Ramsden S, Black GC (2000) Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males. Lancet 356:830–832
D'Cruz DP, Khamashta MA, Hughes GR (2007) Systemic lupus erythematosus. Lancet 369:587–596
Daser A, Rabbitts TH (2005) The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. Semin Cancer Biol 15:175–188
Davalos V, Esteller M (2010) MicroRNAs and cancer epigenetics: a macrorevolution. Curr Opin Oncol 22:35–45
De Smet C, Lurquin C, Lethe B, Martelange V, Boon T (1999) DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol 19:7327–7335
Deng C, Lu Q, Zhang Z, Rao T, Attwood J, Yung R, Richardson B (2003) Hydralazine may induce autoimmunity by inhibiting extracellular signal-regulated kinase pathway signaling. Arthritis Rheum 48:746–756
Dieker JW, Fransen JH, van Bavel CC, Briand JP, Jacobs CW, Muller S, Berden JH, van der Vlag J (2007) Apoptosis-induced acetylation of histones is pathogenic in systemic lupus erythematosus. Arthritis Rheum 56:1921–1933
Eapen V (2011) Genetic basis of autism: is there a way forward? Curr Opin Psychiatry 24:226–236
Eden A, Gaudet F, Waghmare A, Jaenisch R (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300:455
Esteller M (2005) Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol 45:629–656
Esteller M (2008) Epigenetics in cancer. N Engl J Med 358:1148–1159
Farazi TA, Spitzer JI, Morozov P, Tuschl T (2011) miRNAs in human cancer. J Pathol 223: 102–115
Feinberg AP (2004) The epigenetics of cancer etiology. Semin Cancer Biol 14:427–432
Feinberg AP, Ohlsson R, Henikoff S (2006) The epigenetic progenitor origin of human cancer. Nat Rev Genet 7:21–33
Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301:89–92
Fernandez-Morera JL, Calvanese V, Rodriguez-Rodero S, Menendez-Torre E, Fraga MF (2010) Epigenetic regulation of the immune system in health and disease. Tissue Antigens 76:431–439
Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K et al (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet 37:391–400

Fraga MF, Herranz M, Espada J, Ballestar E, Paz MF, Ropero S, Erkek E, Bozdogan O, Peinado H, Niveleau A et al (2004) A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. Cancer Res 64:5527–5534

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311:395–398

Gorelik G, Fang JY, Wu A, Sawalha AH, Richardson B (2007) Impaired T cell protein kinase C delta activation decreases ERK pathway signaling in idiopathic and hydralazine-induced lupus. J Immunol 179:5553–5563

Grafodatskaya D, Chung B, Szatmari P, Weksberg R (2010) Autism spectrum disorders and epigenetics. J Am Acad Child Adolesc Psychiatry 49:794–809

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8:35–46

Guy J, Cheval H, Selfridge J, Bird A (2011) The role of MeCP2 in the brain. Annu Rev Cell Dev Biol 27:631–652

Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. Science 315:1143–1147

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70

Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, Zhou B, Chang CP (2010) Chromatin regulation by Brg1 underlies heart muscle development and disease. Nature 466:62–67

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459:108–112

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547

Hermann A, Schmitt S, Jeltsch A (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. J Biol Chem 278:31717–31721

Hiratani I, Gilbert DM (2009) Replication timing as an epigenetic mark. Epigenetics 4:93–97

Holm TM, Jackson-Grusby L, Brambrink T, Yamada Y, Rideout WM 3rd, Jaenisch R (2005) Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell 8:275–285

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene 27:404–408

Hu N, Qiu X, Luo Y, Yuan J, Li Y, Lei W, Zhang G, Zhou Y, Su Y, Lu Q (2008) Abnormal histone modification patterns in lupus CD4+ T cells. J Rheumatol 35:804–810

Huang YW, Liu JC, Deatherage DE, Luo J, Mutch DG, Goodfellow PJ, Miller DS, Huang TH (2009) Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer. Cancer Res 69:9038–9046

Huertas D, Sendra R, Munoz P (2009) Chromatin dynamics coupled to DNA repair. Epigenetics 4:31–42

Hurd PJ, Bannister AJ, Halls K, Dawson MA, Vermeulen M, Olsen JV, Ismail H, Somers J, Mann M, Owen-Hughes T et al (2009) Phosphorylation of histone H3 Thr-45 is linked to apoptosis. J Biol Chem 284:16575–16583

Illingworth RS, Bird AP (2009) CpG islands–'a rough guide'. FEBS Lett 583:1713–1720

Jacinto FV, Esteller M (2007) MGMT hypermethylation: a prognostic foe, a predictive friend. DNA Repair (Amst) 6:1155–1160

Januchowski R, Wudarski M, Chwalinska-Sadowska H, Jagodzinski PP (2008) Prevalence of ZAP-70, LAT, SLP-76, and DNA methyltransferase 1 expression in CD4+ T cells of patients with systemic lupus erythematosus. Clin Rheumatol 27:21–27

Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG et al (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res 20:170–179

Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128:683–692

Karpf AR, Matsui S (2005) Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer Res 65:8635–8639

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31:89–97
Konkel MK, Batzer MA (2010) A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol 20:211–221
Kuroda A, Rauch TA, Todorov I, Ku HT, Al-Abdullah IH, Kandeel F, Mullen Y, Pfeifer GP, Ferreri K (2009) Insulin gene expression is regulated by DNA methylation. PLoS One 4:e6953
Li Y, Zhao M, Yin H, Gao F, Wu X, Luo Y, Zhao S, Zhang X, Su Y, Hu N et al (2010) Overexpression of the growth arrest and DNA damage-induced 45alpha gene contributes to autoimmunity by promoting DNA demethylation in lupus T cells. Arthritis Rheum 62:1438–1447
Liakopoulos V, Georgianos PI, Eleftheriadis T, Sarafidis PA (2011) Epigenetic mechanisms and kidney diseases. Curr Med Chem 18:1733–1739
Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769–773
Liu H, Liu W, Wu Y, Zhou Y, Xue R, Luo C, Wang L, Zhao W, Jiang JD, Liu J (2005) Loss of epigenetic control of synuclein-gamma gene as a molecular indicator of metastasis in a wide range of human cancers. Cancer Res 65:7635–7643
Maiwald R, Bonte A, Jung H, Bitter P, Storm Z, Laccone F, Herkenrath P (2002) De novo MECP2 mutation in a 46, XX male patient with Rett syndrome. Neurogenetics 4:107–108
Martin C, Zhang Y (2005) The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 6:838–849
Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D (1995) 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med 1:686–692
Miremadi A, Oestergaard MZ, Pharoah PD, Caldas C (2007) Cancer genetics of epigenetic genes. Hum Mol Genet 16(1):R28–R49
Mishra N, Brown DR, Olorenshaw IM, Kammer GM (2001) Trichostatin A reverses skewed expression of CD154, interleukin-10, and interferon-gamma gene and protein expression in lupus T cells. Proc Natl Acad Sci USA 98:2628–2633
Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211:393–396
Mohn F, Schubeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet 25:129–136
Nan X, Campoy FJ, Bird A (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell 88:471–481
Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21:4886–4892
Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386–389
Nan X, Tate P, Li E, Bird A (1996) DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol 16:414–421
Nelson WG, De Marzo AM, Yegnasubramanian S (2009) Epigenetic alterations in human prostate cancers. Endocrinology 150:3991–4002
Ng SS, Yue WW, Oppermann U, Klose RJ (2009) Dynamic protein methylation in chromatin biology. Cell Mol Life Sci 66:407–422
Nishigaki M, Aoyagi K, Danjoh I, Fukaya M, Yanagihara K, Sakamoto H, Yoshida T, Sasaki H (2005) Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. Cancer Res 65:2115–2124
Nowell PC (1989) The clonal nature of neoplasia. Cancer Cells 1:29–30
Ogawa O, Eccles MR, Szeto J, McNoe LA, Yun K, Maw MA, Smith PJ, Reeve AE (1993) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. Nature 362:749–751
Ordovas JM, Smith CE (2010) Epigenetics and cardiovascular disease. Nat Rev Cardiol 7:510–519

Oshimo Y, Nakayama H, Ito R, Kitadai Y, Yoshida K, Chayama K, Yasui W (2003) Promoter methylation of cyclin D2 gene in gastric carcinoma. Int J Oncol 23:1663–1670

Pan W, Zhu S, Yuan M, Cui H, Wang L, Luo X, Li J, Zhou H, Tang Y, Shen N (2010) MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4+ T cells by directly and indirectly targeting DNA methyltransferase 1. J Immunol 184:6773–6781

Paredes J, Albergaria A, Oliveira JT, Jeronimo C, Milanezi F, Schmitt FC (2005) P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. Clin Cancer Res 11:5869–5877

Pedersen MT, Helin K (2010) Histone demethylases in development and disease. Trends Cell Biol 20:662–671

Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28:1057–1068

Rahman A, Isenberg DA (2008) Systemic lupus erythematosus. N Engl J Med 358:929–939

Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP (1993) Relaxation of imprinted genes in human cancer. Nature 362:747–749

Richardson B (2003) DNA methylation and autoimmune disease. Clin Immunol 109:72–79

Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M (1990) Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum 33:1665–1673

Roh TY, Wei G, Farrell CM, Zhao K (2007) Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res 17:74–81

Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, Jones PA (2006) Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell 9:435–443

Schickel R, Boyerinas B, Park SM, Peter ME (2008) MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death. Oncogene 27:5959–5974

Schulz WA, Steinhoff C, Florl AR (2006) Methylation of endogenous human retroelements in health and disease. Curr Top Microbiol Immunol 310:211–250

Shirodkar AV, Marsden PA (2011) Epigenetics in cardiovascular disease. Curr Opin Cardiol 26:209–215

Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell 37:457–468

Smeets E, Terhal P, Casaer P, Peters A, Midro A, Schollen E, van Roozendaal K, Moog U, Matthijs G, Herbergs J et al (2005) Rett syndrome in females with CTS hot spot deletions: a disorder profile. Am J Med Genet A 132A:117–120

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:41–45

Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H (2009) Developmental programming of CpG island methylation profiles in the human genome. Nat Struct Mol Biol 16:564–571

Sullivan KE, Suriano A, Dietzmann K, Lin J, Goldman D, Petri MA (2007) The TNFalpha locus is altered in monocytes from patients with systemic lupus erythematosus. Clin Immunol 123:74–81

Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell 28:1–13

Tudor M, Akbarian S, Chen RZ, Jaenisch R (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc Natl Acad Sci USA 99:15536–15541

van Bavel CC, Dieker J, Muller S, Briand JP, Monestier M, Berden JH, van der Vlag J (2009) Apoptosis-associated acetylation on histone H2B is an epitope for lupus autoantibodies. Mol Immunol 47:511–516

van Bavel CC, Dieker JW, Kroeze Y, Tamboer WP, Voll R, Muller S, Berden JH, van der Vlag J (2011) Apoptosis-induced histone H3 methylation is targeted by autoantibodies in systemic lupus erythematosus. Ann Rheum Dis 70:201–207

Verdone L, Caserta M, Di Mauro E (2005) Role of histone acetylation in the control of gene expression. Biochem Cell Biol 83:344–353

Waddington CH (1942) The Epigenotpye. Endeavour pp. 18–20

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 138:1019–1031

Watt F, Molloy PL (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 2:1136–1143

Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. Biochim Biophys Acta 1775:138–162

Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, Song E (2007) Let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell 131:1109–1123

Index

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8, © Springer Science+Business Media, LLC 2012