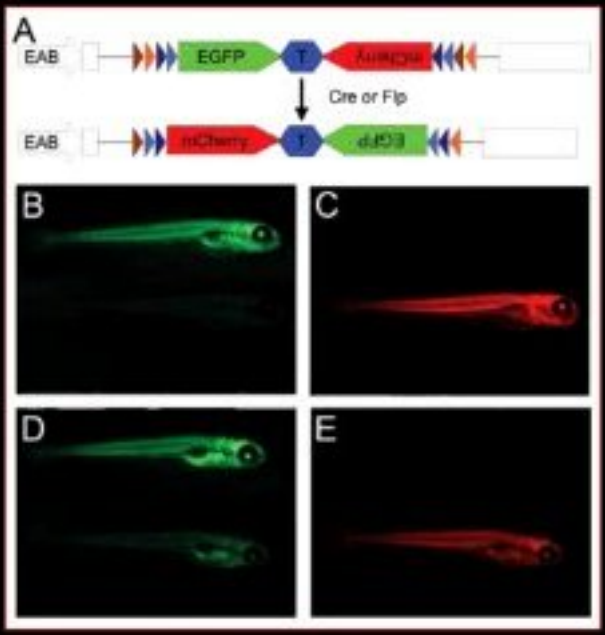
Methods in Cell Biology • Volume 104

The Zebrafish: Genetics, Genomics and Informatics

THIRD EDITION



Edited by

H. William Detrich III, Monte Westerfield and Leonard I. Zon

Methods in Cell Biology

VOLUME 104

The Zebrafish: Genetics, Genomics and Informatics 3rd Edition

Series Editors

Leslie Wilson
Department of Molecular, Cellular and Developmental Biology,
University of California
Santa Barbara, California

Paul Matsudaira
Department of Biological Sciences
National University of Singapore
Singapore

Methods in Cell Biology

VOLUME 104

The Zebrafish: Genetics, Genomics and Informatics 3rd Edition

Edited by

H. William Detrich III

Department of Biology, Northeastern University, Boston, MA, USA

Monte Westerfield

Institute of Neuroscience, University of Oregon, Eugene, OR, USA

Leonard I. Zon

Division of Hematology/Oncology, Children's Hospital of Boson,
Department of Pediatrics and Howard Hughes Medical Institute,
Harvard Medical School, Boston, MA, USA

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
32, Jamestown Road, London NW1 7BY, UK
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

Third edition 2011

ISBN: 978-0-12-374814-0
ISSN: 0091-679X

Printed and bound in USA

11 12 13 14 10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the author's contributions begin.

Gembu Abe (23), Division of Molecular and Developmental Biology, National Institute of Genetics, Mishima, Shizuoka, Japan

Adam Amsterdam (59), Koch Institute for Integrative Cancer Research, Massachusetts Institute for Technology, Cambridge, Massachusetts, USA

Xiaoying Bai (129), Howard Hughes Medical Institute, Boston, Massachusetts, USA; Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, USA; Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA

Teresa Venezia Bowman (341), Stem Cell Program, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA, USA

Ingo Braasch (259), University of Oregon, Institute of Neuroscience, Eugene OR, USA

Yvonne Bradford (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Kim H. Brown (237), Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; Harvard Medical School, Boston, MA, USA

Ewartde Bruijn (121, 221), Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

Shawn Michael Burgess (59, 379), National Human Genome Research Institute, Bethesda, Maryland, USA

Bradley R. Cairns (327), Howard Hughes Medical Institute; Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Julian M. Catchen (259), University of Oregon, Center for Ecology and Evolutionary Biology, Eugene OR, USA

Wenbiao Chen (3), Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

James K. Chen (151), Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

Jose B. Cibelli (209), Comparative Medicine and Integrative Biology program – College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA; Departments of Animal Science and Physiology, Michigan State University, East Lansing, MI, USA; LARCell, Programa Andaluz de Medicina Regenerativa, Seville, Spain

Karl J. Clark (137), Mayo Clinic, Department of Biochemistry and Molecular Biology, Mayo Addiction Research Center, Rochester, Minneapolis, USA

Matthew D. Clark (221), Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB, UK; Present address: The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK

Edwin Cuppen (121, 221), Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

Panos Deloukas (221), Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB, UK

Anthony DiBiase (341), Stem Cell Program, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA, USA

Kimberly P. Dobrinski (237), Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; Harvard Medical School, Boston, MA, USA

Freek van Eeden (121), Department of Biomedical Science, The University of Sheffield, Western Bank, Sheffield, UK

Stephen C. Ekker (137), Mayo Clinic, Department of Biochemistry and Molecular Biology, Mayo Addiction Research Center, 200 1st St SW, 13 Guggenheim, Rochester, MN, USA

David Fashena (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Ken Frazer (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Jennifer L. Freeman (237), School of Health Sciences, Purdue University, West Lafayette, IN, USA

Victor Guryev (221), Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

Saher Sue Hammoud (327), Howard Hughes Medical Institute; Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Douglas G. Howe (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Hong Jiang (129), Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California, USA

David A. Jones (327), Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

J. Keith Joung (51), Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA; Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

Koichi Kawakami (23), Division of Molecular and Developmental Biology, National Institute of Genetics, Mishima, Shizuoka, Japan; Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka, Japan

Ross N.W. Kettleborough (121), Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

Christian Lawrence (431), Aquatic Resources Program, Children's Hospital Boston, Karp Family Research Laboratories, One Blackfan Circle, Boston, MA, USA

Charles Lee (237), Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; Harvard Medical School, Boston, MA, USA

Jin Liang (379), National Human Genome Research Institute, Bethesda MD, USA

Graham J. Lieschke (353), Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria, 3800, Australia; Walter and Eliza Hall Institute, Parkville, Victoria, 3050, Australia

Shuo Lin (129), Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California, USA

James A. Lister, Ph.D. (195), Department of Human and Molecular Genetics and Massey Cancer Center, Virginia Commonwealth University School of Medicine, Box 980033, Richmond, Virginia, USA

Jianjun Lu (3), National Center for Safety Evaluation of Drugs, National Institute for the Control of Pharmaceutical and Biological Products, A8 Hong Da Zhong Lu, Beijing, P. R. China

Lisette A. Maddison (3), Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Jordi Cayuso Mas (401), MRC National Institute for Medical Research, The Ridgeway, London, NW7 1AA, UK

Christian Mosimann (175), Howard Hughes Medical Institute, Boston, MA, USA; Stem Cell Program, Children's Hospital Boston, Boston, MA, USA; Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA, USA

Mary C. Mullins (83), Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

Isaac J. Nijman (221), Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

David A. Nix (327), Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Emily S. Noël (401), MRC National Institute for Medical Research, The Ridgeway, London, NW7 1AA, UK; Present address: Hubrecht Institute and University Medical Centre Utrecht, 3584 CT Utrecht, The Netherlands

Elke A. Ober (401), MRC National Institute for Medical Research, The Ridgeway, London, NW7 1AA, UK

Luke Pase (353), Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria, 3800, Australia; Walter and Eliza Hall Institute, Parkville, Victoria, 3050, Australia

Francisco Pelegri (83), University of Wisconsin – Madison, Laboratory of Genetics, Madison, Wisconsin, USA

Randall T. Peterson (51), Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, USA; Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; Broad Institute, Cambridge, Massachusetts, USA

John H. Postlethwait (221, 259), Institute of Neuroscience, 1254 University of Oregon, Eugene, OR, USA

Magdalena Potok (327), Howard Hughes Medical Institute; Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Sridhar Ramachandran (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Gabriel Renaud (379), National Human Genome Research Institute, Bethesda MD, USA

Barbara J. Ruef (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Leyla Ruzicka (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Jeffry D. Sander (51), Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA; Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

Ilya A. Shestopalov (151), Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

Amy Singer (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Kannika Siripattarapravat (209), Comparative Medicine and Integrative Biology Program – College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA

Kimberly J. Skuster (137), Mayo Clinic, Department of Biochemistry and Molecular Biology, Mayo Addiction Research Center, 200 1st St SW, 13 Guggenheim, Rochester, MN, USA

Ceri Van Slyke (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Derek L. Stemple (121, 221) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB, UK

Maximilliano L. Suster (23), Division of Molecular and Developmental Biology, National Institute of Genetics, Mishima, Shizuoka, Japan; Present address: Sars International Centre for Marine Molecular Biology, Thormøhlensgate 55, Bergen, Norway

Masazumi Tada (221), Department of Cell and Developmental Biology, University College London, Gower Street, London WC, UK

Eirini Trompouki (341), Stem Cell Program, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA, USA

Mark D. Urban (137), Mayo Clinic, Department of Biochemistry and Molecular Biology, Mayo Addiction Research Center, 200 1st St SW, 13 Guggenheim, Rochester, MN, USA

Zoltan M. Varga (453), Zebrafish International Resource Center, Institute of Neuroscience, University of Oregon, Eugene, OR, USA

Gaurav Kumar Varshney (59), National Human Genome Research Institute, Bethesda, Maryland, USA

Monte Westerfield (313), ZFIN, the Zebrafish Model Organism Database, University of Oregon, Eugene, Oregon, USA

Catherine Wilson (221), Institute of Neuroscience, 1254 University of Oregon, Eugene, Oregon, USA

Shan-Fu Wu (327), Howard Hughes Medical Institute; Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Zhongan Yang (129), Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California, USA

J.-R. Joanna Yeh (51) Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, USA; Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

Haiying Zhang (327), Howard Hughes Medical Institute; Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Yi Zhou (287, 341), Stem Cell Program, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA, USA; Division of Hematology/Oncology, Children's Hospital Boston, Boston, MA, USA

Leonard I. Zon (129, 175, 287, 341), Howard Hughes Medical Institute, Boston, Massachusetts, USA; Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, USA; Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA

PREFACE

Building on the foundation of our first (1999) and second (2004) editions of *Methods in Cell Biology: The Zebrafish*, Monte, Len, and I are pleased to continue this Third Edition with *Methods in Cell Biology* Volume 104, *Genetics, Genomics, and Informatics*. In this volume, our contributors present the latest technical advances that have appeared since the second edition. One theme that clearly emerges from these chapters is that the zebrafish is the preeminent vertebrate model for *genetic*, *transgenic*, and *genomic* studies of developmental processes *in vivo.*

Genetics, Genomics, and Informatics covers new technologies under five sections: Forward and Reverse Genetics, Transgenesis, The Zebrafish Genome and Mapping Technologies, Informatics and Comparative Genomics, and Infrastructure. Under Section 1, Forward and Reverse Genetics, new advances in transposon-mediated transgenesis, gene trapping, and enhancer trapping in zebrafish are described, and mutagenic strategies using zinc finger nucleases and retroviral insertion are presented. Reverse-genetic approaches employing target-selected gene inactivation and caged morpholino oligonucleotides are also outlined. Section 2 introduces new transgenic protocols for zebrafish that are based on Cre or phiC31 recombinases or on nuclear transfer. The third section covers new genomic and mapping technologies in zebrafish, including SNP panels for rapid positional cloning of genes, molecular cytogenetics and a BAC probe panel for genome analysis, automated procedures for detection of conserved syntenies among vertebrate genomes, and an update of the Zon laboratory positional cloning manual. Section 4, Informatics and Comparative Genomics, explores ZFIN as a portal for data extraction, transformation, and dissemination, provides protocols for analysis of chromatin structure, DNA methylation, and epigenetic regulation during zebrafish development, describes methods for evaluating the roles of micro RNAs in development, and outlines techniques for sequencing-based transcriptional profiling. Infrastructure, Section 5, reviews zebrafish facilities suitable for small or large laboratories and concludes with a chapter on zebrafish husbandry that emphasizes the importance of standardizing zebrafish lines and genetic backgrounds for cross-comparison of experimental results.

The Third Edition will also introduce *Disease Models and Chemical Screens*, two rapidly emerging and compelling applications of the zebrafish. We trust that the Third Edition will prove valuable both to seasoned zebrafish investigators as well as to those who are newly adopting the zebrafish model as part of their research armamentarium.

We thank the Series Editors, Les Wilson and Paul Matsudaira, and the staff of Elsevier/Academic Press, especially Zoe Kruze, for their enthusiastic support of our Third Edition of *Methods in Cell Biology: The Zebrafish*. Their help, patience, and encouragement are profoundly appreciated.

H. William Detrich III,
Monte Westerfield and
Leonard I. Zon

PART 1

Forward and Reverse Genetics

CHAPTER 1

Generating Conditional Mutations in Zebrafish Using Gene-trap Mutagenesis

Lisette A. Maddison,[*] Jianjun Lu[†] and Wenbiao Chen[*]

[*]Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[†]National Center for Safety Evaluation of Drugs, National Institute for the Control of Pharmaceutical and Biological Products, A8 Hong Da Zhong Lu, Beijing, P. R. China

Abstract

While several mutagenesis methods have been successfully applied in zebrafish, these mutations do not allow tissue- or temporal-specific functional analysis. We have developed a strategy that will allow tissue- or temporal-specific disruption of genes in zebrafish. This strategy combines gene-trap mutagenesis and FlEx modules containing target sites for site-specific recombinases. The gene-trap cassette is highly mutagenic in one orientation and nonmutagenic in the opposite orientation, with different fluorescent proteins as indicators of the orientation. The inclusion of the FlEx modules

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00001-X

allows two rounds of stable inversion mediated by the Cre and Flp recombinases. This gene-trap cassette can be easily delivered via transposons. Through large-scale community-wide efforts, broad genome coverage can be obtained. This should allow investigation of cell/tissue-specific gene function of a wide range of genes.

I. Introduction

Zebrafish has proven to be a powerful model for forward genetic analysis of vertebrate development. It can be argued that the advantages of zebrafish, including high fecundity, low cost maintenance, and feasibility of live imaging, also make it a good model for reverse genetic analysis. Several strategies have been developed to generate and identify mutations in a gene of interest after chemical mutagenesis or insertional mutagenesis (Jao *et al*., 2008; Moens *et al*., 2008; Sood *et al*., 2006; Wienholds *et al*., 2003). Furthermore, targeted mutagenesis using preselected zinc finger nucleases is also feasible (Doyon *et al*., 2008; Foley *et al*., 2009; Meng *et al*., 2008). Despite their proven power, one limitation of these approaches is that the genes are mutated in all cells and at all times. This often complicates determining the function of the mutated genes in a specific tissue or a specific process. Conditional mutations are more desirable because they allow inactivation of genes in somatic cells in a temporal- or tissue-specific manner.

In the mouse, conditional mutations generally refer to alleles containing two target sites of a site-specific recombinase configured to delete either an essential part of or the entire gene. These mutations are constructed by modifying a cloned fragment of the gene of interest, which is then used to replace the endogenous locus via homologous recombination in ES cells (Feil, 2007; Gu *et al*., 1993). Recombination between two target sites mediated by the site-specific recombinase leads to the deletion of the intervening sequence. The Cre recombinase and its target loxP sites are most frequently used in mouse, although the Flp recombinase and its target FRT are also used, but to a lesser extent. Both of these recombinases recognize a 34 bp target site that consists of two 13 bp inverted repeats and an 8 bp spacer (Branda and Dymecki, 2004). Each repeat provides a binding site for one recombinase molecule and the spacer dictates the configuration of the heteroduplex intermediate and the outcome of recombination. If the targets are in the same orientation, as used in conditional alleles in mouse, recombination results in the deletion of the intervening sequence. If the targets are in opposite orientation, recombination will result in the inversion of the intervening sequence. With tissue-specific expression of Cre and ligand-dependent Cre (e.g., CreER), conditional alleles allow functional determination of genes in different tissues/organs and at different stages during development as well as in adult animals.

II. Rationale for Conditional Mutations

Most genes have a dynamic expression patterns and likely function in multiple cells types and multiple stages. Traditional mutations abolish gene function in all

cells from the beginning of embryogenesis, often revealing only the earliest or most conspicuous functions and concealing later or less pronounced functions. This is particularly true for mutations that lead to embryonic lethality, a common feature for most of the available zebrafish mutations. Conditional mutations of these essential genes can reveal additional functional insights. A good example comes from mutation of the insulin receptor in mouse. Global mutation of insulin receptor leads to growth retardation and perinatal death due to diabetic ketoacidosis. But tissue-specific inactivation has revealed that insulin signaling may play distinct roles in different tissues (Kitamura *et al*., 2003). In skeletal muscle, insulin resistance caused by the lack of insulin receptor plays a key role in the development of metabolic syndrome. In the liver, insulin resistance is involved in the development of diabetes, whereas in fat, insulin signaling plays a role in the regulation of life span (Bluher *et al*., 2003; Fisher and Kahn, 2003). Insulin resistance in the CNS regulates both appetite and reproduction (Kahn, 2003). By specifically inhibiting the insulin signaling in these individual tissues, a better overall understanding of the role insulin signaling in multiple physiological processes has been gained. Undoubtedly, an approach for tissue-specific gene inactivation in zebrafish may reveal a more precise functional understanding of genes, particularly the essential genes. The Ekker lab has developed a "gene-breaking" transposon which contains a gene-trap cassette that can be deleted (Petzold *et al*., 2009). However, it only allows temporal- or tissue-specific rescue but not temporal- or tissue-specific inactivation of the mutated genes.

III. Rationale for Gene-trap Mutagenesis to Generate Conditional Mutations

While gene targeting is not available for zebrafish because of the lack of ES cells, gene trapping using the *Tol2* transposon system (Kawakami, 2004) is very efficient. Because the mutagenicity of an intronic gene-trap insertion is orientation dependent, it is possible to make such alleles conditional if the orientation of the gene trap can be stably switched (Schnutgen *et al*., 2005; Xin *et al*., 2005). The orientation switch utilizes site-specific recombinases and a configuration of heterodimers of heterotypic (incompatible) recombinase target sites for Flip and Excision (FlEx) (Floss and Schnutgen, 2008; Schnutgen *et al*., 2003). With a FlEx approach incorporating target sites for both Cre and Flp, the intervening sequence can be inverted in two successive rounds. Coupling the FlEx strategy to the efficient *Tol2* transgenesis in zebrafish will allow conditional gene inactivation without the need for ES cells.

IV. Rationale for Vector Design

There are three keys to generating an ideal gene-trap vector. First, for a conditional gene trap to be successful the cassette needs to be highly mutagenic in one orientation and nonmutagenic in the other orientation. The mutagenicity is

determined by how efficiently the gene trap is recognized as a terminal exon; intercepting and prematurely terminating the transcription of the trapped genes. Because recognition of a terminal exon requires both a splice acceptor and a polyadenylation signal (Levitt *et al*., 1989; Niwa *et al*., 1992), inclusion of a strong splice acceptor and a strong polyadenylation signal in the cassette would ensure a highly mutagenic gene trap that is efficiently recognized as a terminal exon. For the splice acceptor, we used the consensus splice acceptor from zebrafish (Yeo *et al*., 2004) along with intronic and exonic splice enhancers (Fig. 1). We also used the tandem repeats of the well-characterized bovine growth hormone gene polyadenylation and transcription termination sequence as the polyadenylation signal. Tandem repeat of polyadenylation signal is known to stop transcription more effectively (Maxwell *et al*., 1989; Soriano, 1999). Conversely, a weak splice acceptor coupled with a weak polyadenylation site will likely produce a nonmutagenic gene trap. For the nonmutagenic gene trap, we chose a truncated zebrafish splice acceptor consensus without intronic or exonic splicing enhancers and a truncated polyadenylation signal of the SV40 genome. Second, a mechanism is needed to allow two rounds of stable inversion of the gene trap. Two rounds of stable inversion are necessary to achieve tissue-specific gene inactivation for insertions in an initially mutagenic orientation. For this we included FlEx modules for both Cre and Flp to allow two rounds of stable inversion. Third, the vector used to deliver the gene trap needs to be nonmutagenic. A mutagenic vector will negate the conditionality of the gene trap. Although retrovirus is the most efficient insertional mutagen, it is not a suitable for delivering the conditional gene trap since the virus itself is highly mutagenic (Jao *et al*., 2008). Transposons are suitable for generating conditional gene-trap insertions because they require only minimal inverted terminal repeats for transposition and are easy to work with. Furthermore, multiple transposons such as *Tol2*, piggyBac, and Ac/Ds are available (Emelyanov *et al*., 2006; Ivics *et al*., 2009), each potentially with different integration site preferences. The combination of these transposons should make the entire genome more accessible for mutagenesis. Fourth, although not essential, fluorescent reporter genes are useful to indicate the expression pattern of the mutated genes, as well as the orientation of the insertion. The small size of zebrafish makes it difficult to isolate a specific tissue for verification of the mutational status of the trapped gene using DNA-based methods. A fluorescent marker

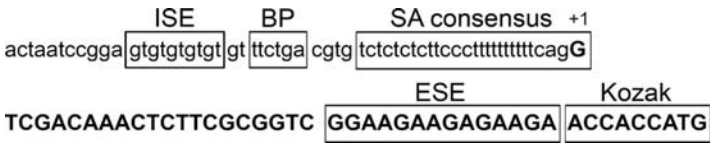


Fig. 1 Features of the synthetic splice acceptor. Capital letters represent exonic sequence and lowercase letters represent intronic sequence. BP, branchpoint; ESE, exonic splice enhancer; ISE, intronic splice enhancer; Kozak, Kozak consensus; SA, splice acceptor.

for each orientation would allow visualization of the genotype of individual cells. However, only a fraction of insertions are in frame, and not all will result in levels high enough for detection.

V. Methods

We have constructed a FlEx module that allows two rounds of inversion, incorporated it in a reporter gene for Cre and Flp, and demonstrated its function in zebrafish (Boniface *et al*., 2009). The FlEx module contains two heterotypic Cre recognition sites, loxP and lox5171, and two heterotypic Flp sites, FRT and F3. Orientation and organization of the module is critical to allow FlEx (Fig. 2). For simplicity, Fig. 3 outlines the flip and excision process using recognition sites for a single recombinase. Following the initial flipping of the cassette between one set of homotypic recognition sites, the other set of homotypic recognition sites is in an orientation that will result in excision. However, the excision event does not impact the cassette and the end result is an inverted cassette with heterotypic recognition sites on either side of the cassette, which is resistant to further recombination by the same recombinase. However, the recognition sites may be used for recombination-

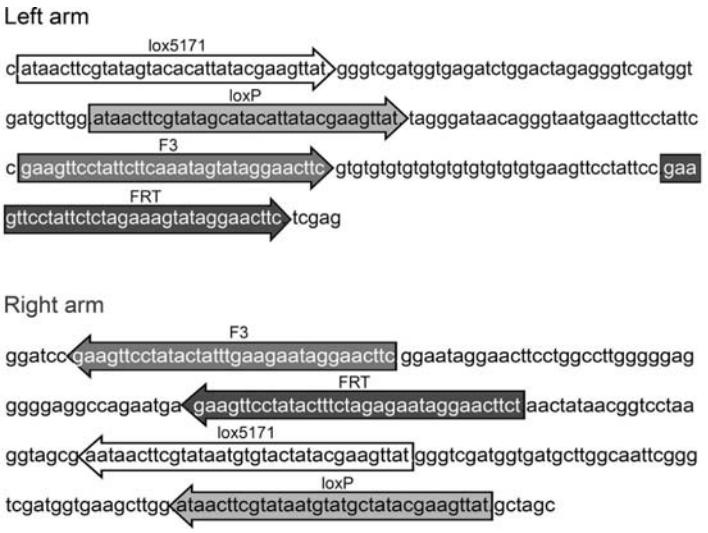


Fig. 2 Organization of FlEx modules. The organization and orientation of the recombinase target sites in each arm are critical for correct and successive FlEx. The location, sequence, and orientation of each recombinase target site are indicated for each arm in 5′ to 3′ configuration. The target sites are lox5171, open arrow with black text, loxP, light gray arrow with black text; F3, dark gray arrow with white text; FRT, black arrow with white text.

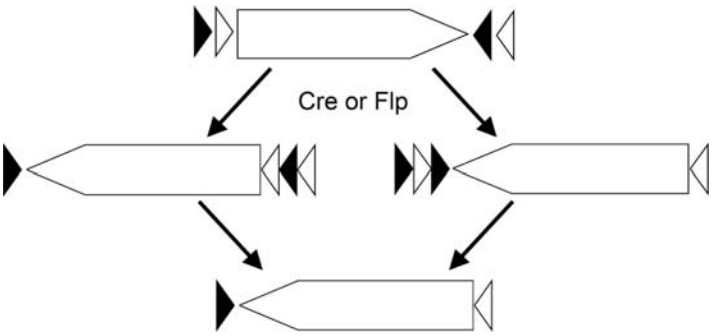


Fig. 3 Schematic of FlEx. DNA between two opposing dimers of noncompatible heterotypic targets can be inverted by a recombinase. The inversion also places two homotypic targets in a configuration for recombinase-mediated deletion, leaving one heterotypic target at each side. The product is resistant to additional inversion.

mediated cassette exchange. Fig. 4 demonstrates inversion induced by *Tg(hsp70I:Cre)* or *Tg(hsp70I:Flp)* as indicated by switching from EGFP expression to mCherry expression in the reporter. Furthermore, in fish with the reporter stably inverted by Cre recombinase, *Tg(hsp70I:Flp)* can reinvert the reporter to EGFP following heat shock.

We have been testing the feasibility of FlEx-based gene-trap mutagenesis in zebrafish and a manuscript detailing this project is forthcoming. For the purposes here, we outline the approach to generate conditional alleles with full details of a pilot screen to be presented elsewhere.

A. Construction of Gene-trap Cassettes

The mutagenic gene trap consists of a strong splice acceptor, mCherry as the reporter gene, and a strong polyadenylation signal. The strong splice acceptor was generated by annealing and repairing primers Splice Acceptor Forward (SAF) and Splice Acceptor Reverse (SAR) (Table I). The strong polyadenylation signal is 5× tandem repeats of the bgh polyadenylation site that was produced by conventional molecular cloning. The nonmutagenic gene trap was generated by introducing a truncated zebrafish splice acceptor consensus to the 5′ end of mCitrine by PCR and removing sequence downstream of the HpaI site of the SV40 polyadenylation signal in pmCitrine-N1. The two gene traps were combined tail to tail and inserted between the two FlEx arms in a Tol2 vector and injected into one-cell stage embryos. An in-frame gene trap resulted in detectable expression of one of the fluorescent reporter genes dependant on the orientation. The mCherry expressing gene trap is much more mutagenic than the mCitrine expressing gene trap using RT-PCR analysis of the trapped gene (data not shown).

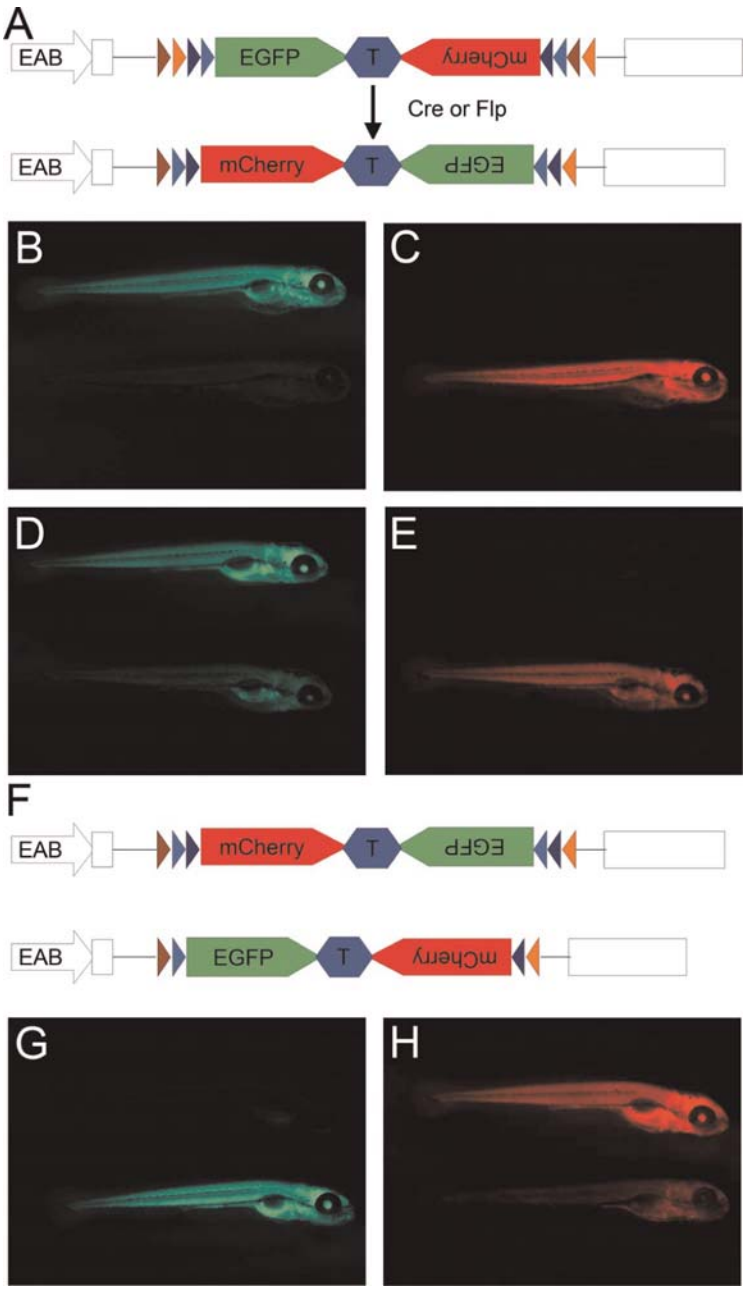


Fig. 4 demonstration of FlEx using a dual reporter fish. Panel A: schematic of the initial configuration of the transgenic reporter. Following FlEx mediated by Cre or Flp, expression will switch from EGFP to mCherry. Panels B–E: inversion of the reporter in embryos. Fluorescent protein expression in progeny of transgenic reporter fish crossed to *Tg(hsp70I:Cre)* (B, C) or *Tg(hsp70I:Flp)* (D, E) transgenic fish. Embryos

B. Generating Insertions

1. *Tol2* Transgenesis

I. DNA and RNA preparation

1. Prepare the *Tol2* gene-trap plasmid DNA using standard plasmid preparation protocols.

   i. Purify the gene-trap plasmid using GeneClean (Bio101, Solon, OH) according to manufacturer's instructions.

2. Prepare transposase RNA using mMESSAGE mMACHINE (Ambion, Austin, TX) according to manufacturer's instructions.

   i. Dilute the RNA to 300 ng/μL and store in 2 μL aliquots at −80 °C until use.

II. Injection

1. Prepare injection solution

   For 5 μL injection solution
   150 ng purified gene-trap DNA
   150 ng transposase RNA
   0.5 μL 0.5% phenol red
   RNAse, DNAse free water to 5 μL

2. Inject 1 nl solution into the cell of one-cell stage wild-type embryos avoiding injection into the yolk.
3. Place injected embryos in 0.3× Danieau solution with 1% penicillin and streptomycin stock solution (Invitrogen, Carlsbad, CA) and incubate at 28 °C.

III. Recovery

1. 4–6 h after injection remove unfertilized and dead embryos.

   i. Keep not more than 70 embryos per dish.

2. 24 h after injection remove any dead embryos.
3. Once the embryos hatch at 3 days after injection, change to fresh 0.3× Danieau solution without antibiotics, removing the chorions and any abnormal embryos.
4. 5 days after injection begin feeding the larvae and raise to maturity.

Fig. 4 (Cont.) were subjected to a 1 h heat shock at 24 hpf and images obtained at 5 dpf. Images were obtained for both EGFP (B, D) and mCherry (C, E). For each panel the upper embryo is Cre or Flp negative and the lower embryo is Cre or Flp positive. Only the recombinase positive embryos show mCherry expression. Panel F: schematic of the transgenic reporter line with stable inversion by Cre recombinase. Following FlEx mediated by Flp, expression will switch from mCherry to EGFP. Panels G, H: re-inversion of the reporter in embryos. Fluorescent protein expression in progeny of inverted transgenic reporter fish crossed to *Tg(hsp70I:Flp)* transgenic fish. Embryos were subjected to a 1 h heat shock at 24 hpf and images obtained at 5 dpf. Images were obtained for both EGFP (G) and mCherry (H). For each panel the upper embryo is Flp negative and the lower embryo is Flp positive. Only Flp positive larvae show EGFP expression. (See color plate)

Table I
Primers used for constructing synthetic splice acceptor

Primer name	Primer sequence 5′ – 3′
SAF	CTCGAGGACTATCCGGAGTGTGTGTGTTTCTGACGTGTCTCTCTCTTCCCTTTTTTTTTTCAGGTCGACAAACTCTTCG
SAR	CCATGGTGGTTCTTCTCTTCTTCCGACCGCGAAGAGTTTGTCGACCTG
LM-PCR long	CTAGGATTAGCTGCTGGAGTACACGATCGCTTAATAGAGGCACGTGGAACGCGGGC
LM-PCR short CG	PCGGCCCGCGTTCCACGTGZZOG
LM-PCR long CATG	CTAGGATTAGCTGCTGGAGTACACGATCGCTTAATAGAGGCACGTGGAACGCGGGCCGCATG
LM-PCR UP1	CTAGGATTAGCTGCTGGAGTACACG
LM-PCR NUP2	ATCGCTTAATAGAGGCACGTGGAAC
5′ TIR1	CCAAAGGACCAATGAACATGTCTGAC
5′ TIR2	AACTGGGCATCAGCGCAATTCA
5′ TIR3	TTGTACTCAAGTAAAGTAAAAATCCC
3′ TIR1	TCAGCCCCAAAAGAGCTAGGCTTG
3′ TIR2	GCGTGTACTGGCATTAGATTGTCTGTC
3′ TIR3	TCAAGTAAGATTCTAGCCAGATAC
GT1	CTTCCGACCGCGAAGAGTTTGTC
GT2	CTTCTAACTATAACGGTCCTAAGGTAGCG

2. Screening for Gene-trap Events

I. Embryo production

1. Cross individual gene-trap injected fish to a wild-type fish.
2. Collect eggs from successful matings.
   i. Hold the gene-trap fish in a separate small tank.

II. Screening for fluorescence

1. At 24 hpf, 48 hpf, and 5 dpf.
   i. Check for fluorescence on both red and yellow channels.
   ii. Keep embryos with fluorescent protein expression and raise them to maturity. Genomic DNA can be isolated from tailfin biopsies at 6 weeks of age for insert identification.
   iii. If more than 10 embryos with identical expression pattern are available from a single clutch, 1–3 embryos can be used for insert identification and the rest can be raised to maturity.

3. Insert Identification via Linker-Mediated PCR

All oligonucleotide sequences are given in Table I

I. DNA isolation

1. From tailfin
   i. Anesthetize the fish in 0.05% MESAB.
   ii. Cut the tail with a scalpel.
   iii. Transfer the tail into a tube containing 250 μL of 20 mM NaOH.
   iv. Place fish in individual plastic cups with 350 mL of system water.
2. From embryos
   i. Place individual embryos in a tube containing 100 μL of 20 mM NaOH.
3. Incubate the samples at 95 °C for 20 min.
4. Add 1/5 volume of 1 M Tris.HCl to neutralize the samples.
   i. 50 μL for tailfin, 20 μL for embryos

II. Template preparation

1. Genomic DNA digestion
   i. In a 50 μL reaction, digest 20 μL of tailfin or embryo lysate (about 1 μg DNA) with one of the following for more than 2 h.

a. 1 μL TaqI at 65 °C
b. 1 μL NlaIII at 37 °C

ii. Heat inactivate the enzyme at 80 °C for 20 min.

2. Linker preparation

i. Determine appropriate linker pair.

a. For Taq1 digestion – CG linker (LM-PCR long + LM-PCR Short CG)
b. For NlaIII digestion – CATG linker (LM-PCR long CATG + LM-PCR Short CG)

ii. Anneal linkers

4.5 μL	20 mM long linker
4.5 μL	20 mM short linker
1.0 μL	10× NEB3 buffer
95 °C	2 min
45 °C	10 min
4 °C	5 min

3. Ligate linkers to genomic DNA fragments at 16°C for more than 2 h.

50.0 μL	Digested genomic DNA
6.0 μL	10× T4 DNA ligase buffer
2.0 μL	10 uM annealed linkers
1.0 μL	T4 DNA ligase

4. Purify DNA using Qiagen PCR purification columns and elute DNA in 40 μL 10 mM Tris pH 8.0.

III. Amplification of 5′ side flanking sequence

1. First round PCR

i. Mix the following.

5.0 μL	5× GoTaq buffer (Promega)
3.0 μL	25 mM $MgCl_2$
5.0 μL	DNA with linker
0.5 μL	10 mM dNTPs
0.5 μL	10 μM UP1
0.5 μL	10 μM 5′TIR1
10 μL	H20
0.2 μL	Taq DNA polymerase (5 U/μL)

ii. Perform PCR using the following thermal profile.

1 cycle	
94 °C	2 min
25 cycles	
94 °C	20 s
65 °C	3 min
1 cycle	
72 °C	5 min

iii. Dilute the PCR 1:25.

2. Second round PCR

i. Mix the following.

10.0 μL	5× GoTaq buffer (Promega)
6.0 μL	25 mM $MgCl_2$
1.0 μL	Diluted first round product
1.0 μL	10 mM dNTPs
1.0 μL	10 μM NUP2
1.0 μL	10 μM 5′TIR2
31.0 μL	H20
0.2 μL	Taq DNA polymerase (5U/μl)

ii. Perform PCR using the following thermal profile.

1 cycle	
94 °C	2 min
35 cycles	
94 °C	20 s
65 °C	3 min
1 cycle	
72 °C	5 min

3. Run the products on a 1.0% agarose gel.
4. Isolate bands, purify fragments, and directly sequence using 5′ TIR3.

IV. Amplification of 3′ side flanking sequence

1. First round PCR

i. Mix the following.

5.0 μL	5× GoTaq buffer (Promega)
3.0 μL	25 mM $MgCl_2$
5.0 μL	DNA with linker
0.5 μL	10 mM dNTPs

0.5 μL	10 μM UP1
0.5 μL	10 μM 3′TIR1
10 μL	H20
0.2 μL	Taq DNA polymerase (5 U/μL)

ii. Perform PCR using the following thermal profile.

1 cycle	
94 °C	2 min
25 cycles	
94 °C	20 s
65 °C	3 min
1 cycle	
72 °C	5 min

iii. Dilute the PCR 1:25.

2. Second round PCR

i. Mix the following.

10.0 μL	5× GoTaq buffer (Promega)
6.0 μL	25 mM $MgCl_2$
1.0 μL	Diluted first round product
1.0 μL	10 mM dNTPs
1.0 μL	10 μM NUP2
1.0 μL	10 μM 5′TIR2
31.0 μL	H20
0.2 μL	Taq DNA polymerase (5U/μl)

ii. Perform PCR using the following thermal profile.

1 cycle	
94 °C	2 min
35 cycles	
94 °C	20 s
65 °C	3 min
1 cycle	
72 °C	5 min

3. Run the products on a 1.0% agarose gel.
4. Isolate bands, purify fragments, and directly sequence using 3′ TIR3.

C. Genotyping

A 3-primer PCR genotyping protocol is used to determine the genotype of individual embryos. Two of the three primers target the genomic DNA flanking the

insertion and will produce an amplicon from the wild-type allele. The third primer is specific to the insertion and will produce an amplicon with one of the other two primers when the insertion is present. We have used the TIR primers in LM-PCR protocol (5′ TIR1, 5′TIR2, 3′ TIR1, and 3′TIR2) as the gene-trap primers. However, these TIR-specific primers are not ideal to distinguish the orientation of the gene trap. To determine the orientation of the gene trap, we used either one of the two primers that are internal of the FlEx modules (GT1 and GT2, Table I). When designing the gene-specific primers, it is important to make sure that the size of the amplicon from the wild-type allele is sufficiently different from the amplicon from the mutant allele to easily distinguish them in a conventional agarose gel.

I. DNA isolation

1. For tailfin from adult fish

   i. Anesthetize the fish in 0.05% MESAB.
   ii. Cut the tail with a scalpel.
   iii. Transfer the tail into a tube containing 250 μL of 20 mM NaOH.
   iv. Place fish in individual plastic cups with 350 mL of system water.

2. For embryos

   i. Place individual embryos in a tube containing 100 μL of 20 mM NaOH.

3. Incubate the samples at 95 °C for 20 min.
4. Add 1/5 volume of 1 M Tris.HCl to neutralize the samples.

   i. 50 μL for tailfin, 20 μL for embryos

II. PCR

1. Prepare a PCR reaction cocktail sufficient for the number of samples.

   i. For a single reaction

5× GoTaq buffer (Promega)	5.0 μL
Sterile water	15.0 μL
25 mM MgCl2	2.5 μL
10 μM gene-specific primer 1	0.5 μL
10 μM gene-specific primer 2	0.5 μL
10 μM transposon-specific primer	0.5 μL
Taq DNA polymerase (5 U/μL)	0.05 μL

2. Add 24 μL of PCR reaction cocktail to each PCR tube.
3. Add 1 μL of the DNA sample to the corresponding PCR tube.

4. Perform PCR using the following thermal profile.

1 cycle	
94 °C	2 min
35 cycles	
94 °C	15 s
65 °C	15 s
72 °C	1 min/kb amplicon
1 cycle	
72 °C	5 min

5. Run 15 μL of each sample on a 1–2% agarose gel depending on amplicon size.

D. Work Flow for Using Conditional Mutations

Once the genomic location and orientation of the insertions have been determined, one can manipulate the orientation of the gene trap either using Cre or Flp transgenic lines, or by RNA injection. Before utilizing a conditional mutation, one should determine whether the gene of interest is expressed in the target tissue as well as other tissues. One should also determine whether global inactivation of the gene of interest causes a discernable defect outside of the target tissues. Due to the greater availability of transgenic Cre lines in zebrafish, Flp recombinase is ideal for germline inversion. Although transgenic Flp lines can be used, we prefer to globally invert the gene trap by injecting Flp RNA into the 1-cell embryos, which can result in 100% germline inversion. By injecting the Flp RNA into progeny of a gene-trap carrier and a Cre transgenic fish, one accomplishes both germline inversion and generation of Cre-expressing carriers of conditional alleles in one generation. We prefer to use a male Cre carrier in these cases. In our hands we have found that even when Cre expression is driven by a previously characterized tissue-specific promoter with no maternal activity as indicated by a fluorescent reporter, germline inversion may still occur. This is potentially because the amount of Cre protein required for recombination is much less than the amount of fluorescent protein needed for visual detection. Therefore, using a male Cre carrier avoids the complication of germline inversion. The following outlines how conditional mutations can be manipulated depending on the initial orientation of the insertion.

E. Starting with a Nonmutagenic Insertion

I. Global gene inactivation

a. Cross the insertion carriers with wild-type fish.

b. Inject Flp RNA in one-cell stage progeny and raise to maturity.
c. Identify injected fish with high levels of germline inversion.

i. Cross the injected fish with wildtype.
ii. Determine insert orientation in 48 embryos to determine frequency of inversion.

d. Incross fish with high levels (close to 100%) of germline inversion and observe progeny for phenotypes.

II. Tissue-specific gene inactivation.

a. Cross the insertion carriers with Cre transgenic fish.
b. Identify double carrier progeny.
c. Cross a double carrier to a Cre-negative, insertion carrier and observe progeny for a phenotype in the target tissue.

III. Tissue-specific rescue

a. Cross the insertion carriers with Cre transgenic fish.

i. Inject Flp RNA in one-cell stage embryos and raise to maturity

b. Identify injected fish with high levels of germline inversion.

i. Cross the injected fish with wildtype.
ii. Determine insert orientation in 48 embryos to determine frequency of inversion.

c. Genotype the fish with high levels of germline inversion for the Cre transgene.
d. Cross a germline inverted, Cre-positive fish to a germline inverted Cre-negative fish and observe progeny for rescue of phenotype presented in global inactivation mutants.

F. Starting with a Mutagenic Insertion

I. Global gene inactivation

a. Cross two insertion carriers and observe embryos for phenotypes.

II. Tissue-specific gene inactivation.

a. Cross the insertion carriers with Cre transgenic fish.

i. Inject Flp RNA in one-cell stage embryos and raise to maturity.

b. Identify injected fish with high levels of germline inversion.

i. Cross the injected fish with wildtype.
ii. Determine insert orientation in 48 embryos to determine frequency of inversion.

c. Cross a double germline inverted, Cre-positive fish to a Cre-negative, germline inverted fish and observe progeny for a phenotype in the target tissue.

III. Tissue-specific rescue

a. Cross the insertion carriers with Cre transgenic fish.
b. Identify double carrier progeny.
c. Cross a double carrier to a Cre-negative, insertion carrier and observe progeny for the rescue in the target tissue.

G. Injection of Cre or Flp RNA

I. Prepare RNA using mMESSAGE mMACHINE (Ambion, Austin, TX) according to manufacturer's instructions

a. Dilute the RNA to 300 ng/μL and store in 2 μL aliquots at −80 °C until use.

II. Injection

a. Dilute RNA with RNAse/DNAse free water to 100 ng/μL.
b. Inject 1 nL (100 pg RNA) into the cell of one-cell stage embryos.
c. Keep injected embryos (not more than 70 per dish) in 0.3× Danieau solution plus antibiotics.

III. Recovery

a. 4–6 h after injection remove unfertilized and dead embryos.
b. 24 h after injection remove any dead embryos.
c. 3 days after injection once the embryos hatch, change the solution to antibiotic-free 0.3× Danieau solution, removing the chorions and any abnormal embryos.
d. 5 days after injection begin feeding the larvae and raise to maturity.

VI. Discussion

This approach to generate and use conditional mutations should allow gene functions to be investigated not only in the entire organism but also in a tissue-specific manner. Although this approach is not as efficient as tilling or retroviral mutagenesis and is not targeted as ZFN mutagenesis, these types of alleles will have a greater utility and are worth the effort. An ambitious goal is to achieve a scale large enough to approach genome saturation. With the availability of multiple transposon systems and the ease of generating insertions using them, this could be achieved. An attractive perspective is for many laboratories to perform conditional gene-trap mutagenesis using different transposon vectors. This will insure the most efficient production of conditional mutations with minimal redundancy.

An essential component to achieve spatial and temporal control of gene inactivation is the availability of suitable Cre- and Flp-expressing lines. In mice, a large collection of such lines has been generated by many labs over the past 15 years [Cre transgenic database, http://www.mshri.on.ca/nagy/Cre-pub.html]. Comparatively, very few lines suitable for conditional gene inactivation exist in zebrafish. The majority of lines used in zebrafish have been based on the hsp70 promoter, although several labs have recently used specific promoters to express the Cre recombinase in a cell/tissue-specific manner, for example, Cre lines specific for cardiomyocytes (Boniface *et al*., 2009; Jopling *et al*., 2010; Kikuchi *et al*., 2010), oocytes (Liu *et al*., 2008), pancreatic beta-cells (Hesselson *et al*., 2009) among others. The predominantly used Cre transgenic line has been Tg(hsp70:Cre) which in theory allows spatial and temporal regulation of Cre expression (Halloran *et al*., 2000). However, the zebrafish hsp70l promoter has basal activity in the absence of heat shock (Blechinger *et al*., 2002; Le *et al*., 2007; Scott *et al*., 2007) and can be induced even at standard laboratory conditions (Feng *et al*., 2007) including in germ cells. Another way to achieve precise control of recombinase activity is to use a conditional variant of Cre. The most commonly used conditional recombinase is Cre fused to mutant versions of the ligand-binding domain of estrogen receptors that are insensitive to endogenous hormone but remain sensitive to tamoxifen, a synthetic anti-estrogen (Casanova *et al*., 2002; Feil *et al*., 1997; Hayashi and McMahon, 2002; Indra *et al*., 1999). The combination of a tissue-specific promoter and a tamoxifen-inducible Cre could provide a powerful means for gene functional analysis. However, for robust analysis of tissue-specific inhibition, the majority of cells in the tissue would need to be affected. There is some question of the ligand-induced efficiency of the Cre-ER and Cre-ERT2 fusion proteins and the ability to affect all cells.

VII. Summary

We have presented here an approach to generating conditional mutations in zebrafish. The core of this approach rests with an asymmetrically mutagenic gene-trap vector and FlEx modules which allow two rounds of stable inversion. This allows the insertion to be flipped from a nonmutagenic to a mutagenic orientation in a tissue- or temporal-specific manner using an appropriate Cre or Flp transgenic line. Through large-scale, community-wide efforts incorporating different transposon systems, extensive genome coverage can be achieved. This will likely allow conditional mutagenesis of most genes of interest and further extend the usefulness of the zebrafish.

Acknowledgments

We thank members in the Chen laboratory for discussions. The work is supported by a grant from National Eye Institute at NIH (EY016092).

References

Blechinger, S. R., Evans, T. G., Tang, P. T., Kuwada, J. Y., Warren Jr., J. T., and Krone, P. H. (2002). The heat-inducible zebrafish hsp70 gene is expressed during normal lens development under non-stress conditions. *Mech. Dev.* **112**, 213–215.

Bluher, M., Kahn, B. B., and Kahn, C. R. (2003). Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574.

Boniface, E. J., Lu, J., Victoroff, T., Zhu, M., and Chen, W. (2009). FlEx-based transgenic reporter lines for visualization of Cre and Flp activity in live zebrafish. *Genesis* **47**, 484–491.

Branda, C. S., and Dymecki, S. M. (2004). Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell.* **6**, 7–28.

Casanova, E., Fehsenfeld, S., Lemberger, T., Shimshek, D. R., Sprengel, R., and Mantamadiotis, T. (2002). ER-based double iCre fusion protein allows partial recombination in forebrain. *Genesis* **34**, 208–214.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702–708.

Emelyanov, A., Gao, Y., Naqvi, N. I., and Parinov, S. (2006). Trans-kingdom transposition of the maize dissociation element. *Genetics* **174**, 1095–1104.

Feil, R. (2007). Conditional somatic mutagenesis in the mouse using site-specific recombinases. *Handb. Exp. Pharmacol.* **178**, 3–28.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757.

Feng, H., Langenau, D. M., Madge, J. A., Quinkertz, A., Gutierrez, A., Neuberg, D. S., Kanki, J. P., and Look, A. T. (2007). Heat-shock induction of T-cell lymphoma/leukaemia in conditional Cre/lox-regulated transgenic zebrafish. *Br. J. Haematol.* **138**, 169–175.

Fisher, S. J., and Kahn, C. R. (2003). Insulin signaling is required for insulin's direct and indirect action on hepatic glucose production. *J. Clin. Invest.* **111**, 463–468.

Floss, T., and Schnutgen, F. (2008). Conditional gene trapping using the FLEx system. *Methods Mol. Biol.* **435**, 127–138.

Foley, J. E., Maeder, M. L., Pearlberg, J., Joung, J. K., Peterson, R. T., and Yeh, J. R. (2009). Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat. Protoc.* **4**, 1855–1867.

Gu, H., Zou, Y. R., and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164.

Halloran, M. C., Sato-Maeda, M., Warren, J. T., Su, F., Lele, Z., Krone, P. H., Kuwada, J. Y., and Shoji, W. (2000). Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* **127**, 1953–1960.

Hayashi, S., and McMahon, A. P. (2002). Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* **244**, 305–318.

Hesselson, D., Anderson, R. M., Beinat, M., and Stainier, D. Y. (2009). Distinct populations of quiescent and proliferative pancreatic beta-cells identified by HOTcre mediated labeling. *Proc. Natl Acad. Sci. U. S. A.* **106**, 14896–14901.

Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res.* **27**, 4324–4327.

Ivics, Z., Li, M. A., Mates, L., Boeke, J. D., Nagy, A., Bradley, A., and Izsvak, Z. (2009). Transposon-mediated genome manipulation in vertebrates. *Nat. Methods* **6**, 415–422.

Jao, L. E., Maddison, L., Chen, W., and Burgess, S. M. (2008). Using retroviruses as a mutagenesis tool to explore the zebrafish genome. *Brief Funct. Genomic. Proteomic.* **7**, 427–443.

Jopling, C., Sleep, E., Raya, M., Marti, M., Raya, A., and Belmonte, J. C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606–609.

Kahn, C. R. (2003). Knockout mice challenge our concepts of glucose homeostasis and the pathogenesis of diabetes. *Exp. Diabesity Res.* **4**, 169–182.

Kawakami, K. (2004). Transgenesis and gene trap methods in zebrafish by using the *Tol2* transposable element. *Methods Cell Biol.* **77**, 201–222.

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., Evans, T., Macrae, C. A., Stainier, D. Y., and Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. *Nature* **464**, 601–605.

Kitamura, T., Kahn, C. R., and Accili, D. (2003). Insulin receptor knockout mice. *Annu. Rev. Physiol.* **65**, 313–332.

Le, X., Langenau, D. M., Keefe, M. D., Kutok, J. L., Neuberg, D. S., and Zon, L. I. (2007). Heat shock-inducible Cre/Lox approaches to induce diverse types of tumors and hyperplasia in transgenic zebrafish. *Proc. Natl Acad. Sci. U. S. A.* **104**, 9410–9415.

Levitt, N., Briggs, D., Gil, A., and Proudfoot, N. J. (1989). Definition of an efficient synthetic poly(A) site. *Genes Dev.* **3**, 1019–1025.

Liu, X., Li, Z., Emelyanov, A., Parinov, S., and Gong, Z. (2008). Generation of oocyte-specifically expressed cre transgenic zebrafish for female germline excision of loxP-flanked transgene. *Dev. Dyn.* **237**, 2955–2962.

Maxwell, I. H., Harrison, G. S., Wood, W. M., and Maxwell, F. (1989). A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. *Biotechniques* **7**, 276–280.

Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D., and Wolfe, S. A. (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701.

Moens, C. B., Donn, T. M., Wolf-Saxon, E. R., and Ma, T. P. (2008). Reverse genetics in zebrafish by TILLING. *Brief Funct. Genomic. Proteomic.* **7**, 454–459.

Niwa, M., MacDonald, C. C., and Berget, S. M. (1992). Are vertebrate exons scanned during splice-site selection? *Nature* **360**, 277–280.

Petzold, A. M., Balciunas, D., Sivasubbu, S., Clark, K. J., Bedell, V. M., Westcot, S. E., Myers, S. R., Moulder, G. L., Thomas, M. J., and Ekker, S. C. (2009). Nicotine response genetics in the zebrafish. *Proc. Natl Acad. Sci. U. S. A.* **106**, 18662–18667.

Schnutgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N. B., Ruiz, P., Chambon, P., Wurst, W., and von Melchner, H. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7221–7226.

Schnutgen, F., Doerflinger, N., Calleja, C., Wendling, O., Chambon, P., and Ghyselinck, N. B. (2003). A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat Biotechnol.* **21**, 562–565.

Scott, E. K., Mason, L., Arrenberg, A. B., Ziv, L., Gosse, N. J., Xiao, T., Chi, N. C., Asakawa, K., Kawakami, K., and Baier, H. (2007). Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat. Methods* **4**, 323–326.

Sood, R., English, M. A., Jones, M., Mullikin, J., Wang, D. M., Anderson, M., Wu, D., Chandrasekharappa, S. C., Yu, J., Zhang, J., and Paul Liu, P. (2006). Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods* **39**, 220–227.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**, 70–71.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R. H., and Cuppen, E. (2003). Efficient target-selected mutagenesis in zebrafish. *Genome Res.* **13**, 2700–2707.

Xin, H. B., Deng, K. Y., Shui, B., Qu, S., Sun, Q., Lee, J., Greene, K. S., Wilson, J., Yu, Y., Feldman, M., and Kotlikoff, M. I. (2005). Gene trap and gene inversion methods for conditional gene inactivation in the mouse. *Nucleic Acids Res.* **33**, e14.

Yeo, G., Hoon, S., Venkatesh, B., and Burge, C. B. (2004). Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. U. S. A.* **101**, 15700–15705.

CHAPTER 2

Tol2-mediated Transgenesis, Gene Trapping, Enhancer Trapping, and the Gal4-UAS System

Gembu Abe,[*] Maximilliano L. Suster[‡] *and* Koichi Kawakami[*,†]

[*]Division of Molecular and Developmental Biology, National Institute of Genetics, Mishima, Shizuoka, Japan

[†]Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka, Japan

[‡]Sars International Centre for Marine Molecular Biology, Thormøhlensgate 55, Bergen, Norway

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00002-1

Abstract

The *Tol2* transposable element was originally found in the genome of the Japanese medaka fish (*Oryzias latipes*). *Tol2* contains a gene encoding an active transposase that can catalyze DNA transposition in vertebrate cells. In zebrafish, *Tol2* generates genomic integrations in the germ cells very efficiently. By using the *Tol2* transposition system, we have developed important genetic methods including transgenesis, gene trapping, enhancer trapping, and the Gal4-UAS system in zebrafish. In this chapter, we describe how these methods can be performed.

I. Introduction

The *Tol2* transposable element, which was found in the genome of the Japanese medaka fish (*Oryzias latipes*), belongs to the hAT family of transposons that includes *hobo* of *Drosophila*, *Ac* of maize, and *Tam3* of snapdragon (Koga *et al*., 1996). The *Tol2* element is an autonomous transposon that contains a gene encoding an active transposase (Kawakami *et al*., 1998; Kawakami and Shima, 1999) (Fig. 1A). By using the *Tol2* element, transgenesis methods have been developed in various vertebrate cells (Kawakami, 2007). In zebrafish, a plasmid DNA containing a non-autonomous *Tol2* construct (a *Tol2*-donor plasmid) and mRNA synthesized *in vitro* encoding the transposase are injected into fertilized eggs. The *Tol2* construct transposes from the plasmid to the genome in the germ lineage during embryonic development very efficiently, and more than 50% of the injected fish can transmit *Tol2* insertions to the next generation (Kawakami *et al*., 2000, 2004). Therefore the *Tol2* system has become an essential tool for transgenesis in zebrafish.

The *Tol2* element contains DNA sequences that are recognized by the transposase and are essential for transposition. The minimal *cis*-sequences necessary for transposition are short, that is, 200-bp and 150-bp of DNA from the left and right ends of the original *Tol2* element (Fig. 1A) (Urasaki *et al*., 2006). This enabled construction of compact *Tol2* vectors in our lab and also in other labs (Balciunas *et al*., 2006; Urasaki *et al*., 2006).

Any DNA fragment can be cloned between these *cis*-sequences. Unlike retroviral vectors and other transposon vectors belonging to the *Tc1*/*mariner* family, the *Tol2* vector has fairly large cargo capacity. Namely, ~10-kb DNA can be cloned without reducing the transpositional activity (Urasaki *et al*., 2006). Moreover, we recently showed that DNA fragments of much larger sizes, such as BAC plasmids, can be cloned between the *cis*-sequences, and the entire construct can be integrated through transposition (Suster *et al*., 2009b). This *Tol2*-mediated BAC transgenesis method will be useful for analysis of *cis*-regulatory elements scattered in the genome, for rescue of mutant phenotypes, and for targeted expression of desired genes.

From single germline transmitting founder fish, six to seven insertions are typically transmitted to the offspring (Kawakami *et al*., 2004; Urasaki *et al*., 2006). This feature, together with high germline transmission rates, enables us to generate thousands of

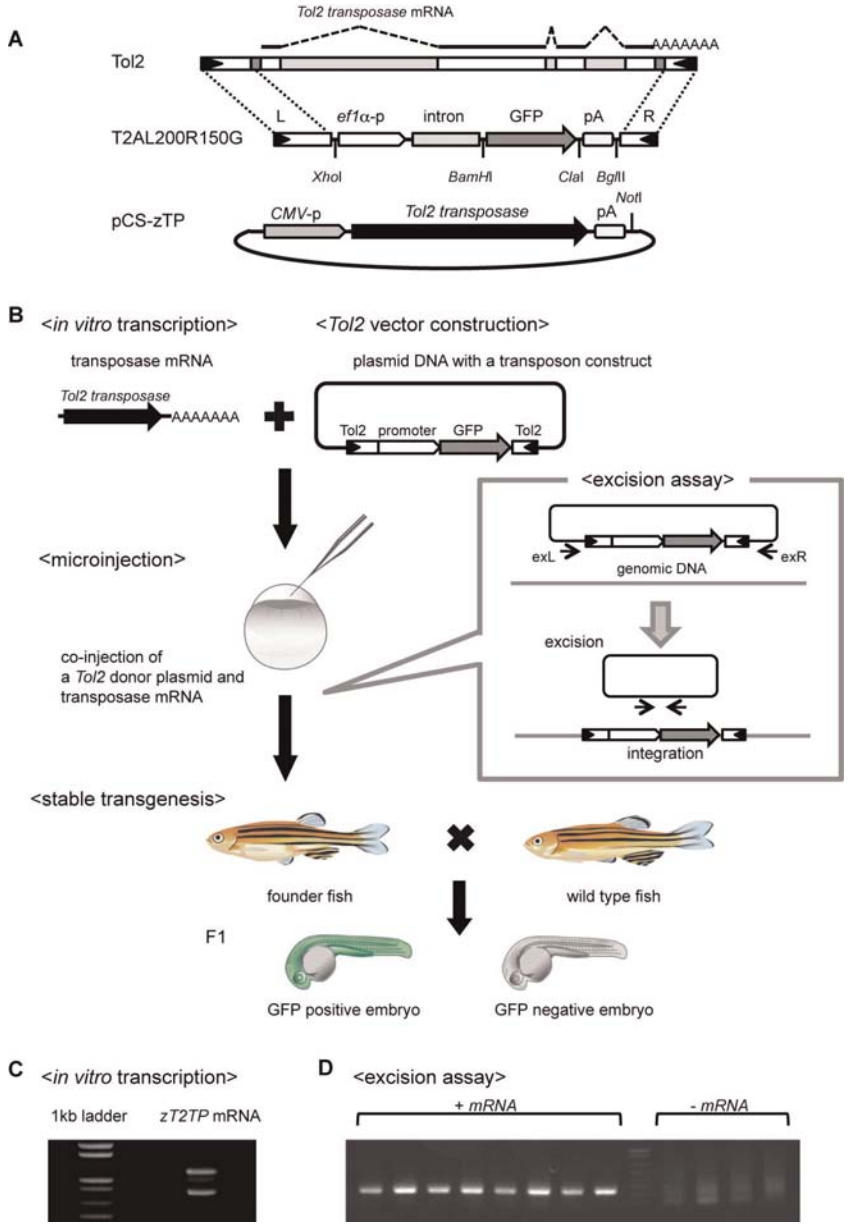


Fig. 1 Transgenesis by using the *Tol2* transposable element. **(A)** Structures of the *Tol2* element and of the plasmid to synthesize the transposase mRNA. Full-length *Tol2* consists of 4682 bp and encodes a gene for transposase (dotted lines indicate introns). T2AL200R150G contains 200-bp and 150-bp DNA from the left (L) and the right (R) termini of *Tol2* (open boxes with triangles), the *Xenopus* EF1α promoter (*ef1α*-p), the rabbit β-*globin* intron, the GFP gene, and the SV40 polyA signal (pA). Unique restriction

Tol2 insertions rather easily thus facilitating development of important genetic methods. First, we created gene trap and enhancer trap methods, and a number of fish expressing GFP in temporally and spatially restricted patterns have been produced (Kawakami *et al*., 2004; Nagayoshi *et al*., 2008; Parinov *et al*., 2004). Second, we combined the Gal4-UAS system with gene trapping and enhancer trapping to generate fish expressing the Gal4 transcription activator in specific cells, tissues, and organs (Asakawa *et al*., 2008; Davison *et al*., 2007; Scott *et al*., 2007). These transgenic zebrafish lines should be extremely useful tools because they can be used to express specific genes in desired locations and at desired developmental stages.

In this chapter, we describe how these *Tol2*-based genetic methods can be implemented.

II. Transgenesis by Using the *Tol2* Transposable Element

A. Rationale

A *Tol2*-donor plasmid DNA and the transposase mRNA (Fig. 1A) are introduced into zebrafish fertilized eggs by microinjection. In the injected embryos, the transposase protein is synthesized and catalyzes excision of the *Tol2* construct from the donor plasmid. The excised *Tol2* construct integrates into the genome during embryonic development, and some cells harboring *Tol2* insertions in the genome differentiate into germ cells. When such injected fish are used for mating, transgenic fish will be obtained in the progeny. In the example shown here, ubiquitous expression of GFP is driven by the EF1α promoter (Fig. 1B).

B. Methods

1. Synthesis of Transposase mRNA *In vitro*

1. Linearize pCS-zTP (Fig. 1A) by digestion with *Not*I and use it as a template to synthesize mRNA according to the protocol in the mMESSAGE mMACHINE SP6 Kit (Ambion Inc., Austin, TX).

Fig. 1 (Cont.) enzyme sites are indicated. pCS-zTP carries a codon-optimized *Tol2* transposase cDNA. **(B)** A scheme for transgenesis in zebrafish. The transposase mRNA synthesized *in vitro* and a plasmid DNA containing the *Tol2* construct are co-injected into fertilized eggs. The transposase protein synthesized from the mRNA catalyzes excision of the *Tol2* construct from the plasmid and integration of the excised *Tol2* into the genome. The injected embryos are raised and crossed with wild-type fish. The integrated *Tol2* construct is transmitted to the F1 generation. **(C)** Electrophoresis of the synthesized transposase mRNA on a standard agarose/TAE gel. Two bands are detected presumably due to its higher-order structure. **(D)** Electrophoresis of PCR products generated by excision assay. After excision of the *Tol2* construct, the DNA double strand break on the donor plasmid is repaired and re-ligated. Therefore, PCR using the exL and exR primers that are located at both sides of the *Tol2* construct generates short PCR products from the backbone plasmid. The PCR products are detected in embryos injected with both the transposase mRNA and the donor plasmid, but not in embryos injected only with the donor plasmid. (See color plate.)

2. Purifiy the transposase mRNA by using "Quick Spin Columns for radiolabeled RNA purification" (Roche, Indianapolis, IN), then precipitate the mRNA and resuspend it in nuclease-free water at 250 ng/μL.
3. Analyze the product by gel electrophoresis. For electrophoresis of RNA, a denaturing gel is preferable, but, alternatively, a standard agarose/TAE gel can be used (Fig. 1C).

2. Preparation of a *Tol2*-donor Plasmid

1. Clone the desired DNA fragment into an appropriate *Tol2* vector, for instance either by using the *Xho*I and *Bgl*II sites on T2AL200R150G (Urasaki *et al*., 2006) (Fig. 1A) or by using the *Tol2* vectors with the Gateway system (Kwan *et al*., 2007; Villefranc *et al*., 2007).
2. Prepare the transposon-donor plasmid DNA using QIAfilter Plasmid Maxi Kit (QIAGEN), purify the recombinant plasmid once by phenol/chloroform extraction, precipitate it with ethanol, and suspend the plasmid in nuclease-free water at 250 ng/μL.

3. Microinjection

1. Put male and female adult zebrafish in a mating box in the evening and collect fertilized eggs in the next morning (Fig. 2A). Microinjection should be carried out at the one-cell stage, within 30 min post fertilization.
2. Make an injection ramp by using 1% agarose, a glass plate, and a 6-cm plastic dish (Fig. 2B). Create fine needles for microinjection by using a glass capillary (GC-1, Narishige, Japan) and a puller (PC-10, Narishige, Japan). Cut the tip with a surgical blade (No. 11, Akiyama MEDICAL MFG. CO., Japan).
3. Prepare DNA/RNA solution by mixing the following components: 10 μL of 0.4 M KCl, 2 μL of Phenol Red solution (Sigma-Aldrich, St. Louis, MO), 2 μL of 250 ng/μL transposase mRNA, 2 μL of 250 ng/μL *Tol2*-donor plasmid DNA, and 4 μL of nuclease-free water (final volume 20 μL). Before injection, centrifuge the mixture at the maximum speed for 1 min to precipitate and remove debris that may clog the injection needle. Transfer the upper 18 μL to a new tube.
4. Fill the DNA/RNA solution into the glass capillary from the backside by using a Microloader tip (Eppendorf, Germany). Attach the filled capillary to a holder (No. 11520145, Leica, Germany) and connect the holder to a 10 mL syringe via a Teflon tube (inner diameter: 0.56 mm, Chukoh Chemical Industries, Japan) (Fig. 2C).
5. Inject ~1 nL of the DNA/RNA solution (the approximate volume can be measured by observing the diameter of the injected bolus by eye) into the cytoplasm of fertilized eggs (Fig. 2D). Incubate the injected embryos in a plastic dish at 28 °C.

4. Excision Assay

To confirm that transposition reaction has occurred, the excision assay should be performed in a subsample of embryos (Kawakami and Shima, 1999) (Fig. 1B).

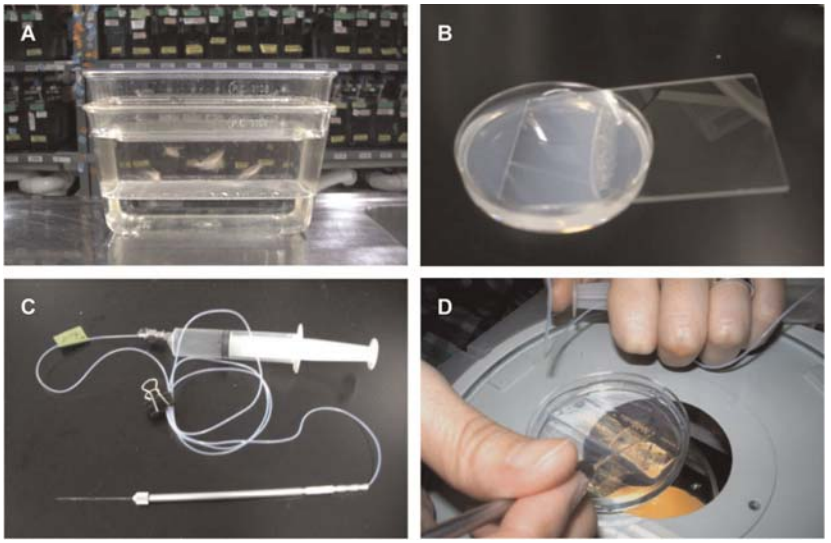


Fig. 2 Apparatus and tools for mating and microinjection. **(A)** A zebrafish mating box (Aquaschwarz, Germany). Male/female pairs of adult zebrafish are placed in a plastic tank with a sieve insert through which fertilized eggs settle, thus separating them from the parents. **(B)** An agarose ramp for microinjection. Melted 1% agarose is poured in a 60 mm Petri dish and then a glass plate is placed to create a slant. **(C)** Microinjection apparatus. A glass capillary is attached to a holder and connected to a syringe through a Teflon tube. The DNA/RNA mixture is backloaded into the capillary prior to attachment to the holder. **(D)** Microinjection is performed under a stereoscope. The left hand provides air pressure to the capillary. (See color plate.)

1. About 10 h after microinjection, transfer several of the injected embryos one by one to 0.2-mL strip tubes (eight tubes per strip). Remove water and add 50 μL of lysis buffer (10 mM Tris-HCl pH8.0, 10; mM EDTA, 200 μg/mL proteinase K). Incubate the sample at 50 °C for 2 h to overnight.
2. Inactivate the proteinase K by heating at 95 °C for 5 min. Prepare PCR reaction mixture containing 1 μM primers (exL and exR), buffer, Hi-Fi taq (Roche), and 1 μL of the sample. Perform PCR; 35cycles of 94 °C for 30 s; 55 °C for 30 s; 72 °C for 30 s, and analyze the PCR product on 1.5% gel electrophoresis. When the *Tol2* portion is excised from the donor plasmid, the backbone plasmid is re-ligated and DNA without *Tol2* will be amplified (Fig. 1D).

   exL: 5′-ACC CTC ACT AAA GGG AAC AAA AG-3′
   exR: 5′-CAA GGC GAT TAA GTT GGG TAA C-3′

5. Identification of Transgenic Fish

1. Raise the remaining injected embryos to the sexual maturity. This usually takes about 3 months.

2. Cross the injected fish with wild type fish and analyze the offspring. When the transgene includes sequence for an in-frame fluorescent marker, such as GFP, monitor expression in embryos using a stereomicroscope equipped for epifluorescence. Pick GFP-positive embryos and raise them.
3. Alternatively, collect a subsample (∼50) of the day1-embryos for PCR analysis. Place embryos into a microtube, add 250 μL DNA extraction buffer (10 mM Tris-HCl pH8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 μg/mL proteinase K) and incubate them at 50 °C overnight. Purify embryonic DNA by phenol/chloroform extraction, precipitate with ethanol, and resuspend in 50 μL TE. Use 1 μL of the DNA sample for PCR (35 cycles of 94 °C for 30 s; 55 °C for 30 s; 72 °C for 30 s) using transgene specific primers. When a PCR-positive F1 pool is found, raise their siblings and analyze them individually at the adult stage for the presence of the transgene by PCR of caudal fin clips.
4. We highly recommend analyzing the F1 fish by Southern blot hybridization to identify fish with single *Tol2* insertions. F1 fish often carry multiple insertions, and this may complicate further analyses. When all of the F1 fish analyzed carry multiple insertions, cross fish with the smallest number of insertions to wild type fish, raise F2 offspring, and analyze F2 fish again by Southern blot hybridization (see below).

C. Materials

pCS-zTP: A codon-optimized version of the transposase cDNA was cloned into pCS+ (Rupp *et al*., 1994; Turner and Weintraub, 1994), resulting in pCS-zTP (Fig. 1A).

Tol2 vectors: Our standard vector pT2AL200R150G contains 200-bp and 150-bp DNA from the left and right ends of *Tol2*, respectively (Fig. 1A) (Urasaki *et al*., 2006). A foreign DNA can be cloned between these sequences using unique *BglII* and *XhoI* sites. Recently, to simplify the cloning process, Kwan *et al*. and Villefranc *et al*. constructed *Tol2* vectors containing sequences for the Gateway technology (Kwan *et al*., 2007; Villefranc *et al*., 2007). By using these *Tol2*-Gateway plasmids, any DNA fragment cloned in an entry vector, for instance pCR8/GW/TOPO (Invitrogen, Carlsbad, CA), can be transferred into those *Tol2* vectors through recombination between *att*L and *att*R sites catalyzed by LR clonase.

D. Discussion

Transgenesis using the *Tol2* transposon system is highly efficient. 50–70% of fish injected with the *Tol2* system at the one-cell stage and grown up to the adulthood are germline-transmitting founder fish that can transmit the transgene to their offspring. In addition, *Tol2*-mediated transgenesis has the following merits. First, transgenic fish carrying a single copy transgene integration can easily be created, whereas transgenic fish constructed by the plasmid DNA injection method often carry transgene concatemers at a single locus whose expression is silenced (Stuart *et al*., 1988). Second, end-to-end integration of the transgene is guaranteed.

Third, the transposon insertion does not cause gross rearrangement of the integration locus. Since the integration site is clean, it can be analyzed by PCR-based methods such as inverse PCR (Kawakami *et al*., 2000, 2004), and adaptor-ligation PCR (Kotani *et al*., 2006; Urasaki *et al*., 2006).

III. *Tol2*-mediated BAC Transgenesis

A. Rationale

Large-insert bacterial artificial chromosomes (BACs) are useful tools for studies of the regulation and function of genes. BAC DNA has been introduced into the zebrafish genome by microinjection of the BAC clones into fertilized eggs, but the frequency of recovery of transgenic offspring is low (Yang *et al*., 2006). Recently, we demonstrated that the *Tol2* transposon system can be applied to BAC transgenesis in zebrafish and mouse (Suster *et al*., 2009b). To this end, we developed the *iTol2* cassette, which enables introduction of the *Tol2 cis*-sequences essential for transposition into BAC plasmids (Suster *et al*., 2009b) (Fig. 3A). The *iTol2* cassette is incorporated into a BAC clone by recombineering (homologous recombination) in *E. coli* (Warming *et al*., 2005), and the resulting *Tol2*-BAC plasmid is injected into fertilized eggs with transposase mRNA (Fig. 3). The *Tol2*-BAC construct integrates into the zebrafish genome in the germ lineage efficiently and is transmitted to the progeny from 5–20% of the injected fish.

B. Methods

1. Preparation of a BAC Plasmid

1. Identify a BAC clone containing your gene of interest by using resources available on the ZFIN database or by querying to the *Ensembl* genome browser. BAC clones from the zebrafish genomic BAC library CHORI-211, which were used for the zebrafish genome sequencing project at Sanger Institute, can be purchased from imaGenes (http://www.imagenes-bio.de/) or from the BACPAC Resource Center (http://bacpac.chori.org/home.htm).
2. Cultivate *E. coli* transformed with the BAC clone of interest in 5 mL LB with 12.5 μg/mL chloramphenicol (in a 15 mL Falcon tube) overnight.
3. Collect the bacteria by centrifugation at 4000 rpm for 5 min. Remove the supernatant and dissolve the pellet in 250 μL buffer P1 (QIAprep Spin Miniprep Kit, Qiagen).
4. Add 250 μL of the kit's P2 buffer to the suspension. Mix gently by inverting and incubate for 5 min at room temperature.
5. Add 250 μL of N3 buffer to the suspension. Mix gently by inverting and incubate for 5 min on ice.
6. Centrifuge the suspension twice at 15,000 rpm for 5 min; be sure to transfer the first supernatant to a new tube. Recover the supernatant.

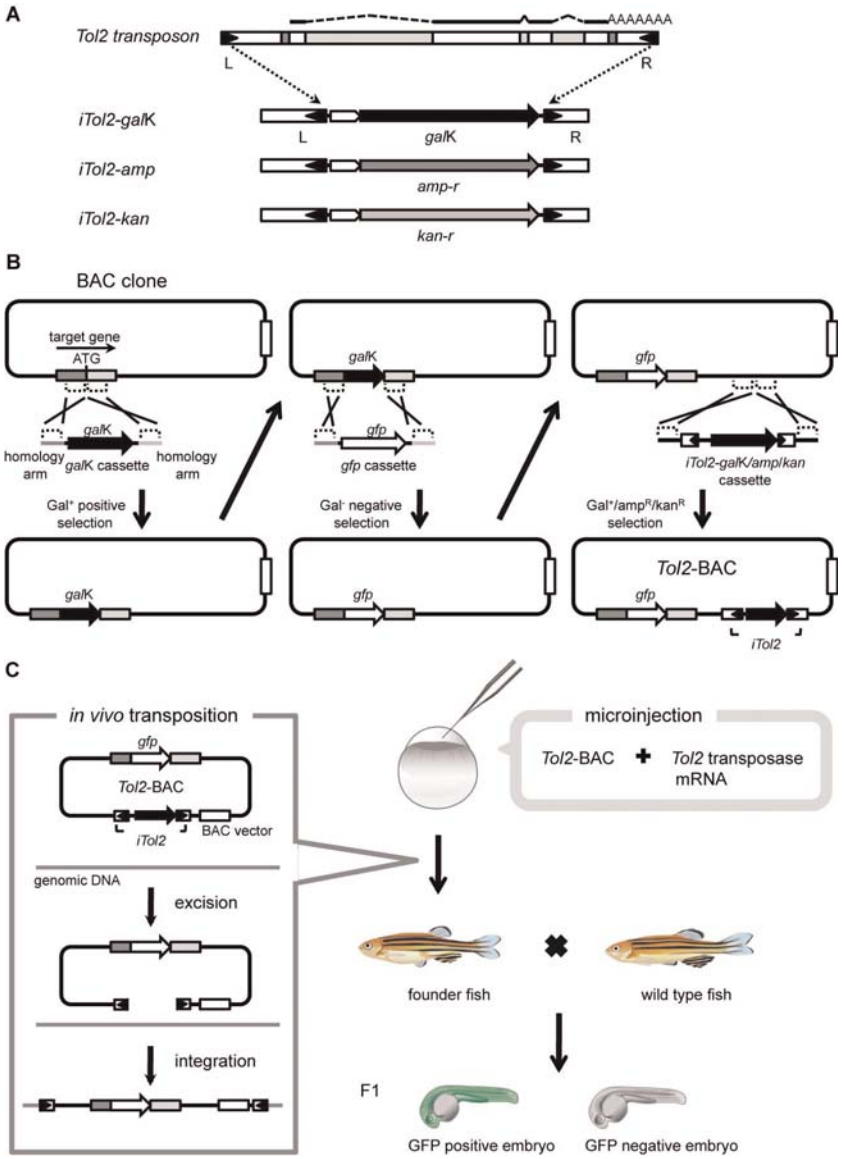


Fig. 3 *Tol2*-mediated BAC transgenesis. **(A)** Structures of the *iTol2* cassettes. The *cis*-sequences required for transposition are shown by open boxes with black triangles (black triangles indicate terminal inverted repeats and their orientations). Note that the *cis*-sequences are inverted in the *iTol2* cassette. *iTol2-gal*K, *iTol2-amp*, and *iTol2-kan* contain the *gal*K, ampicillin resistance, and kanamycin-resistance genes, respectively. **(B)** Schema for BAC plasmid recombineering for zebrafish transgenesis. Host cells are *E. coli* SW102. *Left*: integration of the *gal*K sequence into a target site on BAC by homologous recombination. *Center*: replacement of the *gal*K cassette with the *gfp* cassette by homologous

7. Add 750 μL isopropanol to the supernatant to precipitate BAC DNA. Mix gently, incubate on ice for 10 min, and centrifuge for 10 min at 15,000 rpm. Wash the pellet in 70% ethanol and then allow it to air-dry. Approximately 1–1.5 μg of DNA will be obtained.
8. Dissolve the pellet in 50 μL TE. Use the sample for restriction enzyme analysis, for sequencing, and for transformation of *E. coli* SW102 cells.

2. Transformation of SW102 Cells with BAC DNA

1. Incubate SW102 bacteria in 5 mL LB overnight with tetracycline (20 μg/mL) at 32 °C.
2. Transfer 500 μL of the culture into 25 mL LB with tetracycline (20 μg/mL). Grow bacteria at 32 °C to $OD_{600} = 0.6$ (∼ 4 h).
3. Cool the SW102 culture in an ice water bath for 5 min and transfer aliquots of 10 mL into pre-chilled 15 mL Falcon tubes. Pellet cells by centrifugation at 0 °C for 5 min. Place the tubes on ice, remove the supernatant, and re-suspend the cells in 2 mL ice-cold H_2O by gently pipetting with a 10 mL pipette. Add 8 mL ice-cold H_2O and collect the cells by centrifugation (0 °C, 5 min). Repeat this step once more.
4. After the final centrifugation, remove the supernatant from each tube. The pellets in the tubes should be ∼50 μL. Transfer 25 μL of each pellet to a pre-chilled microtube. Add 1–5 μL (∼25–100 ng) BAC DNA and incubate the cell–DNA mixture on ice for 5–10 min.
5. Transfer the cells to a pre-chilled electroporation cuvette (BioRad, CA, USA) and transform by using a MicroPulser (BioRad) pre-set to condition Ec1.
6. Transfer the bacteria to a tube with 1 mL LB medium, and incubate at 32 °C for 1 h. Spread the transformed bacteria on LB agar plates with 12.5 μg/mL chloramphenicol (BAC clones from the CHORI-211 library carry the chloramphenicol-resistance gene). Incubate at 32 °C overnight.

3. Introduction of the GFP Gene (or Any Gene of Interest) into BAC by Recombineering

1. Amplify the *galK* cassette (Fig. 3B) by PCR using 1–2 ng of the p*galK* plasmid and primers which contain 50-nt extensions that are homologous with target sequences on the BAC plasmid.
2. Digest the PCR product with *Dpn*I to remove the template plasmid DNA. Then electrophorese the DNA on an agarose gel, excise the gel slice containing the

Fig. 3 (Cont.) recombination and *galK* negative selection. *Right*: integration of the *iTol2* cassette into a second target site in the *gfp* BAC by homologous recombination. **(C)** Schema for *Tol2*-BAC transgenesis in zebrafish. The *Tol2*-BAC plasmid containing a recombineered gene of interest (*gfp* in this example) is co-injected with transposase mRNA into the cytoplasm of fertilized eggs (one-cell stage). Transposase excises the *Tol2*-BAC fragment (bearing the gene of interest), which integrates as a single copy insertion into a single genomic locus. (See color plate.)

galK DNA fragment, and recover the DNA using the QIA quick gel extraction kit (QIAGEN).

3. Add 500 μL of an overnight culture of SW102 cells harboring the BAC of interest to 25 mL LB plus chloramphenicol (12.5 μg/mL). Grow cells at 32 °C to OD_{600} = 0.6 (∼4 h).
4. Heat-shock the cells at 42 °C for 15 min.
5. Wash the cells twice as described in step III.B.*b*-3.
6. Transform the cells with the *galK* cassette DNA (10–30 ng) by electroporation as described in step III.B.*b*-4 and step III.B.*b*-5.
7. Transfer the bacteria to a tube with 1 mL LB medium and incubate at 32 °C for 1 h.
8. Wash the bacteria twice with 1× M9 medium (1 mL) by centrifugation and resuspension. Resuspend the pellet in 1 mL 1× M9 medium, and plate cells onto M63 minimal medium plates with galactose, leucine, biotin, and chloramphenicol (12.5 μg/mL). Incubate at 32 °C for 3 days.
9. Pick colonies and streak on MacConkey/galactose indicator plates containing chloramphenicol (12.5 μg/mL). Incubate the plates at 32 °C overnight. Pick single bright red (Gal+) colored colonies and cultivate in 5 mL LB medium containing chloramphenicol (12.5 μg/mL) at 32 °C overnight. These colonies should harbor recombinant BAC clones with the *gal*K cassette and its homology arm extensions.
10. Amplify the *gfp* cassette (Fig. 3B) by PCR using any plasmid containing the *gfp* gene and primers which contain 50-nt extensions that are homologous with the target sequences flanking the *galK* sequence on the BAC plasmid. Digest the PCR product with *Dpn*I and purify by electrophoresis and gel extraction as described in step III.B.*c*-2.
11. Repeat steps III.B.*c*-3 through -6 to introduce the extended *gfp* cassette into SW102 cells harboring the *gal*K-modified BAC.
12. Transfer cells to 10 mL LB medium and incubate the culture at 32 °C for 4.5 h.
13. Wash the bacteria twice with 1 × M9 medium (1 mL) by centrifugation and resuspension. Resuspend the pellet in 1 mL 1 × M9 medium and plate cells onto M63 minimal medium plates with glycerol, leucine, biotin, 2-deoxygalactose (DOG), and chloramphenicol (12.5 μg/mL). Incubate at 32 °C for 3 days.
14. Pick 10–12 colonies, prepare BAC plasmid DNA as described above (III.B.*a*1-8), and analyze the inserts by restriction enzyme digestion. Compare the digestion patterns with those of the parental BAC plasmid and then confirm successful homologous recombination by DNA sequencing. In the example shown in Fig. 3B, the expected outcome is the replacement of the *gal*K cassette by the *gfp* cassette with maintenance of the flanking homology arms.

4. Introduction of Tol2 Sequences into BAC Clones by Use of the *iTol2* Cassette and Recombineering

1. Amplify an appropriate *iTol2* cassette (Fig. 3A) by PCR using 1–2 ng of plasmid DNA (*piTol2-amp, piTol2-kan*, or *piTol2-galK*) and primers which contain 50-nt

extensions that are homologous with a second pair of target sequences on the BAC plasmid.

2. Introduce the extended *iTol2* cassette into SW102 cells harboring the previously recombineered BAC (Section III.B.*c*) by electroporation. Pick positive recombinant clones by selection using antibiotics or by *galK* selection (steps III.B.*c*-8 and -9). Positive colonies should harbor a BAC clone that contains two inserts: (1) *gfp* or another gene of interest and (2) the *iTol2* cassette at a second locus. We refer to this two-insert construct as *Tol2*-BAC.

5. Preparation of *Tol2*-BAC DNA and Microinjection

1. Grow SW102 cells containing the *Tol2*-BAC clone and isolate *Tol2*-BAC plasmid DNA by using NucleoBond BAC 100 (MACHEREY-NAGEL). Purify the plasmid DNA by performing three phenol extractions, one phenol/chloroform extraction, and one ethanol precipitation. Resuspend the DNA in nuclease-free water at 250 ng/μL.
2. *Tol2*-BAC plasmid and *Tol2* transposase mRNA are mixed at final concentrations of 50 ng/μL and 25 ng/μL, respectively, in 0.2 M KCl. Inject approximately 1 nL of the DNA/RNA mixture into the cytoplasm of one-cell stage embryos. Raise the injected fish to sexual maturity (Fig. 3C).
3. Cross each injected fish with wild type fish. Analyze more than 100 embryos per mating. If the *gfp* gene was inserted into the BAC, Pick GFP-positive embryos and raise them.
4. For other genes of interest, assay for transgenesis by PCR using genomic DNA extracted from a pool of F1 embryos and gene-specific primers (see II.B.*e*-3).

C. Materials

SW102 bacteria: SW102 (*mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7649 *gal*U *rsp*L *nup*G *Φ*80d*lacZ*ΔM15 [λ*cI*857 (*cro-bio*A)<>*tet*] Δ*gal*K) harbors a defective λ prophage ([λ*cI*857 (*cro-bio*A)<>*tet*]) that encodes the heat-inducible recombinase, and has a deletion in the galactokinase gene (Δ*galK*) (Warming *et al*., 2005). Galactokinase is required for SW102 to grow on minimal medium containing galactose as a sole sugar source. Galactokinase also catalyzes the phosphorylation of a galactose analog, DOG, which leads to accumulation of the toxic metabolite, 2-deoxy-galactose-1-phosphate, in the cell (Alper and Ames, 1975). Thus, the *galK* system can be used for both positive and negative selection. Consequently, BAC DNA can be modified without introduction of an unwanted selectable marker at the modification site.

p*galK* plasmid and primers: p*galK* has the *galK* ORF and the prokaryotic *em7* promoter (Warming *et al*., 2005). To introduce the *galK* gene into a BAC by homologous recombination, select two target sequences (homology arms; 50 bp each) from the BAC clone of interest. Synthesize *galK* primers that incorporate

the homology arm sequences at their 5′ ends. The target sequences on the BAC may be adjacent or separated. When the target sequences are separated, the DNA between them will be deleted after homologous recombination.

Forward primer: 5′- "50-nt target sequence"-CCT GTT GAC AAT TAA TCAT CGG CA-3′
Reverse primer: 5′- "50-nt target sequence"-TCA GCA CTG TCC TGC TCC TT-3′

gfp plasmid and primers: pT2AL200R150G (Fig. 1A) is used to amplify the *gfp* gene. For homologous replacement of the *galK* gene by the *gfp* gene (or other gene of interest), the same BAC target sequences (homology arms) employed to insert the former are incorporated at the 5′ ends of the primers used to amplify the *gfp* sequence.

Forward Primer: 5′- "identical 50-nt homology arm"-GTG AGC AAG GGC GAG GAG CTG-3′
Reverse primer: 5′- "identical 50-nt homology arm"-CTT TAA AAA ACC TCC CAC ACC TCC C-3′

iTol2-galK, *iTol2-amp*, and *iTol2-kan* plasmids and primers: These plasmids contain the *galK* gene, ampicillin-resistance gene, and kanamycine-resistance gene, respectively, between inverted *Tol2* left and right sequences, respectively (Fig. 3A). These plasmids are used to synthesize *iTol2* cassettes. Select two target sequences (50-nt each) on a BAC plasmid that differ from the homology arms used for insertion of *gfp* or other gene of interest. Incorporate these sequences at the 5′ ends of the primers used to amplify the *iTol2* cassettes.

Forward primer of *iTol2* cassette: 5′- "50-nt homology arm"-CCC TCT CGA GCC GGG CCC AAG TG-3′
Reverse primer of *iTol2* cassette: 5′- "50-nt homology arm"-ATT ATG ATC CTC TAG ATC AGA TCT-3′

D. Discussion

We have successfully created several transgenic zebrafish lines using the BAC recombineering methods described above. To demonstrate proof of principle, we tested the ability of the *Tol2* transposon to mobilize a relatively small BAC clone, *Fugu* BAC 240G7 (containing ∼59 kb of DNA from the pufferfish *evx1* locus and obtained from Geneservice Ltd, UK) (Suster *et al*., 2009b). First, we introduced by recombineering the *Gal4FF* sequence in frame at the start codon of the *evx1* gene. Second, we introduced the *iTol2galK* cassette ∼10-kb downstream of the *evx1* gene. The resulting BAC plasmid, *Tol2-Frevx1*:*Gal4*-BAC (∼65 kb), was injected into zebrafish embryos with transposase mRNA. The injected fish were raised to adulthood and used for mating. Transgenic progeny were obtained from 5% (2/40) of the injected fish. The germline transmission frequencies in BAC transgenesis with microinjection of naked DNA have been 1–3% (Yang *et al*., 2006) or lower.

Thus, *Tol2*-mediated BAC transgenesis increases the germline transmission frequency. Moreover, *Tol2-Frevx1*:*Gal4*-BAC integrated as single copies in the two fish, and the integrations sites (on chromosomes 8 and 18, respectively) bore the landmarks of transposase-mediated insertion. When these *Tol2-Frevx1*:*Gal4*-BAC transgenic fish were mated to homozygous UAS:GFP reporter fish, Gal4FF expression recapitulated that of the *evx1* gene (Suster *et al*., 2009b) as shown by detecting GFP fluorescence in the nervous system, including spinal neurons. The successful outcome of this test motivated attempts to transpose BACs of larger size.

Recently we integrated the *iTol2* cassette into a BAC clone that contains ~174-kb DNA of the zebrafish *fgf24* locus and the *gfp* gene. The resulting *Tol2*-BAC plasmid was co-injected into zebrafish one-cell embryos with mRNA encoding a codon-optimized transposase. The germline transmission frequency was 20%, which significantly exceeded frequencies obtained in our prior transgenesis trials. The construct integrated as a single copy and GFP expression in transgenic offspring was identical to that observed for the wild-type *fgf24* gene (G. Abe, unpublished). We conclude that *Tol2*-mediated BAC transgenesis facilitates the high frequency, germline insertion of BACs of large size into the zebrafish genome.

IV. Gene Trapping and Enhancer Trapping with the *Tol2* Transposon System

A. Rationale

Using *Tol2*-mediated transgenesis, one can generate hundreds to thousands of transposon insertions in the zebrafish genome quickly and easily. We have taken advantage of this transposition system to develop gene trap and enhancer trap methods(Kawakami *et al*., 2004; Nagayoshi *et al*., 2008; Parinov *et al*., 2004). *Gene trap* constructs commonly contain an upstream 3′-splice acceptor, coding sequence for the GFP gene (or other reporter or selectable marker), and a downstream polyadenylation signal sequence. After transposon-mediated integration into an intron, transcription of the "trapped" gene from its endogenous promoter, splicing and processing of the pre-RNA, and translation of the mRNA lead to production of GFP (or other reporter). Thus, gene trap constructs simultaneously inactivate a trapped gene while reporting its expression pattern and providing a tag for its isolation and identification. *Enhancer trap* constructs contain a minimal promoter (e.g., the zebrafish *hsp70* promoter) and the GFP gene (or other reporter). When an enhancer trap construct integrates into the genome and its minimal promoter is activated by a nearby enhancer, GFP is expressed in a spatial pattern dictated by the trapped enhancer (Fig. 4A and 4B). Because gene trap and enhancer trap zebrafish often carry multiple insertions, GFP-positive F1s should be analyzed by Southern blot hybridization to identify insertions that may be responsible for the expression patterns. The integration sites of *Tol2* gene trap and enhancer trap constructs can be readily determined by inverse PCR and adapter ligation PCR.

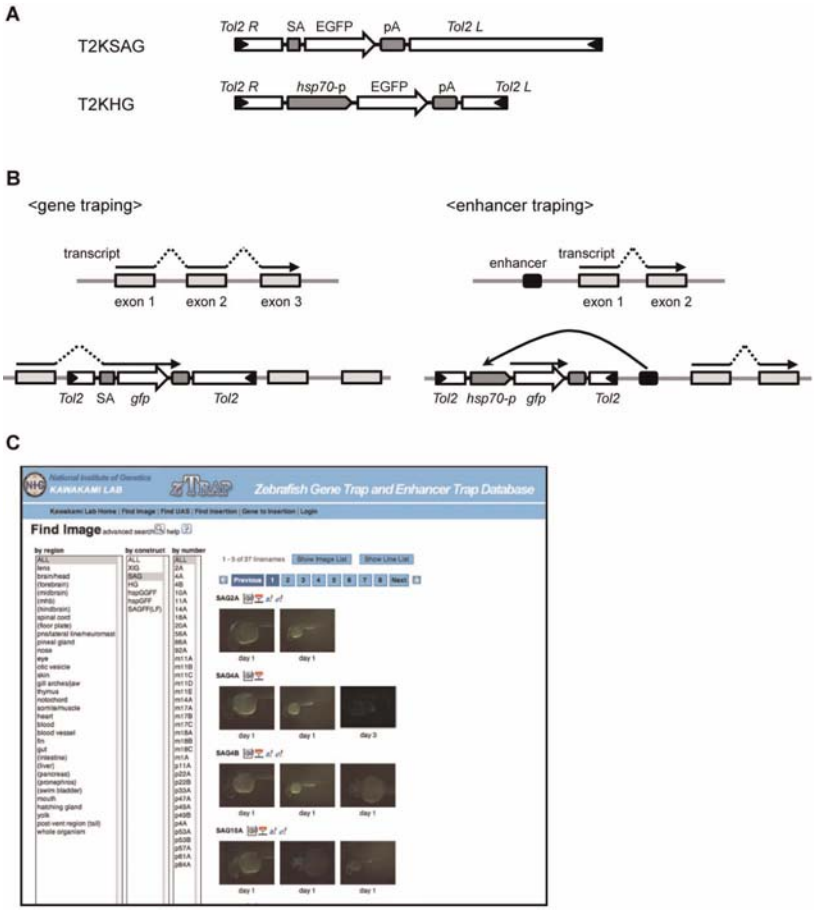


Fig. 4 *Tol2*-mediated gene trapping and enhancer trapping. **(A)** The structures of the gene trap construct (T2KSAG) and enhancer trap construct (T2KHG). The *cis*-sequences are shown as open boxes with black triangles. SA, pA, and *hsp70*-p indicate the splice acceptor of the rabbit β-globin gene, the SV40 polyA signal, and the zebrafish *hsp70* promoter respectively. **(B)** Schematic integration of the gene trap and the enhancer trap constructs in the genome. <**gene trapping**>**:** Integration of the gene trap construct within a putative gene and trapping of its transcript give rise to GFP expression. <**enhancer trapping**>**:** Integration of the enhancer trap construct in a locus which is under the influence of a putative genomic enhancer gives rise to GFP expression. **(C)** Search for useful gene trap and enhancer trap fish through the *z*Trap database (http://kawakami.lab.nig.ac.jp/ztrap/). Transgenic lines that express GFP or Gal4FF in regions of interest are shown by clicking a region name in the left column. (See color plate.)

B. Methods

1. Gene Trap and Enhancer Trap Screens

1. Co-inject fertilized eggs with a plasmid containing either the T2KSAG gene trap construct or the T2KHG enhancer trap and the transposase mRNA. Raise the injected fish.
2. Cross the injected fish each other or cross them with wild type fish.
3. Analyze the offspring. Observe GFP expression under a fluorescent microscope at different stages (e.g., 24 hpf, 48 hpf, 72 hpf, and 5 dpf). Pick GFP-positive embryos and raise them to adulthood.
4. To identify a gene or an enhancer trapped by the insertion, analyze genomic DNA surrounding the insertion by Southern blot hybridization and adaptor-ligation PCR or inverse PCR.
5. In some cases, a transposon insertion disrupts the function of an endogenous gene (Kotani and Kawakami, 2008; Nagayoshi *et al*., 2008). To identify the insertional mutation, mate heterozygous male and female fish that carry it. When the insertion causes a recessive mutation, a mutant phenotype may be identified in a quarter of the offspring.

2. Analysis of *Tol2* Insertions by Southern Blot Hybridization

1. Clip caudal fins of the F1 fish and lyse the tissue in 200 μL of DNA extraction buffer (10 mM Tris-HCl pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 μg/mL proteinase K) at 50 °C for 3 h to overnight. Purify DNA by phenol/chloroform extraction, precipitate with ethanol, and suspend in 50 μL TE. Approximately 20–30 μg DNA will be obtained.
2. Digest 5 μg of the genomic DNA with *Bgl*II, which cuts most of our transposon constructs once. Perform electrophoresis by using 1% TAE–agarose gel. Confirm the samples are completely digested.
3. Soak the gel in 0.1 N HCl for 15 min, rinse it with deionized water, soak it in 0.5 N NaOH for 30 min, rinse it with water, and transfer it in 10X SSC.
4. Place the gel in a vacuum transfer apparatus (BS-31, BIO CRAFT, Japan) with Hybribond-N+ (GE Healthcare, England) presoaked in 10X SSC. Perform transfer according to the manufacturer's instructions. After transfer, rinse the membrane in 2X SSC and dry completely at 50 °C for 1 h to overnight.
5. Amplify the GFP DNA fragment by PCR using primers GFP-f1 and GFP-r1 and T2KHG or T2KSAG as a template. When Southern blot hybridization is performed for transgenic fish carrying the Gal4FF sequence (see Section V), amplify the Gal4FF DNA fragment by PCR using primers Gal4FF-f2 and Gal4FF-r2 and T2KhspGFF or T2KSAGFF as a template. Purify the DNA fragment an agarose gel with QIAGEN. Label DNA with ^{32}P-dCTP using Prime-it II random labeling kit (Agilent Technologies, Palo Alto, CA). Purify the labeled DNA with mini-spin column (Roche).

6. Soak the membrane in 25 mL preheated hybridization buffer (0.25 M $Na_2H_3PO_4$ (adjust pH to 7.4 with H_3PO_4), 1 mM EDTA, 10 g/L bovine serum albumin (BSA), 7% SDS) on a tray and together with buffer transfer into a hybridization tube. Incubate in a hybridization oven at 65 °C for 1 h.
7. Remove hybridization buffer from the tube, then add 10 mL preheated hybridization buffer containing the labeled DNA probe. Perform hybridization at 65 °C overnight.
8. Wash the membrane with ∼100 mL wash solution (0.1X SSC, 0.1% SDS) at 65 °C twice for 30 min each. Transfer the membrane on a tray and rinse with 0.1X SSC. Wrap the membrane in plastic wrap and place on an imaging plate (or X-ray film) in an X-ray cassette. After exposure overnight, the image can be analyzed by using BAS2500 (Fuji Photo Film) (Fig. 5A).

3. Identification of *Tol2* Integration Sites by Inverse PCR

1. Digest 1 μg of genomic DNA with *Mbo*I in 10 μL of reaction buffer at 37 °C for 1 h. Incubate the sample at 70 °C for 15 min. Add 430 μL H_2O to the sample, incubate at 70 °C for 10 min, and cool to 16 °C.
2. Add 50 μL 10X T4 DNA ligation buffer (TAKARA, Japan) and 2 μL T4 DNA ligase, then incubate the sample at 16 °C for 3 h to overnight.
3. Add 50 μL of 3 M sodium acetate and 1 mL ethanol to the sample. Chill the sample at −20 °C for 30 min. Centrifuge the sample at 15,000 rpm for 20 min at 4 °C. Rinse once with 70% ethanol and suspend in 20 μL H_2O.
4. Using 10 μL of the ligation sample, perform the first PCR (30 cycles of 94 °C for 30 s; 57 °C for 30 s; 72 °C for 1 min) using *Tol2*-5′inv-f1 and *Tol2*-5′inv-r1 primers for the 5′ junction or *Tol2*-3′inv-f1 and *Tol2*-3′inv-r1 primers for the 3′ junction.
5. Using 2 μl of the first PCR product, perform the second PCR (30 cycles of 94 °C for 30 s; 57 °C for 30 s; 72 °C for 1 min) using *Tol2*-5′inv-f2 and *Tol2*-5′inv-r2 primers for the 5′ junction or *Tol2*-3′inv-f2 and *Tol2*-3′inv-r2 primers for the 3′ junction.
6. Analyze the PCR products on a 1.5% TAE–agarose gel, then extract from the gel and perform sequencing (Fig. 5B).

4. Identification of Tol2 Integration Sites by Adaptor-ligation PCR

1. Digest 1 μg of genomic DNA with *Mbo*I in 10 μL of reaction buffer at 37 °C for 1 h. Incubate the enzyme at 70 °C for 15 min. Ligate the DNA sample with the GATC adaptor using T4 DNA ligase at 16 °C for 1–3 h.
2. Heat the sample at 70 °C for 15 min to inactivate the ligase and dilute 10-fold with H_2O. Use 1 μL of the diluted sample for 30–35 cycles of PCR (94 °C for 30 s; 57 °C for 30 s; 72 °C for 2 min) using primers Ap1 and L175-out for the 5′ junction, and Ap1 and R175-out for the 3′ junction.
3. Dilute the PCR product 10-fold with H_2O and use 1 μL for second PCR. Thirty cycles of PCR are carried out using primers Ap2 and L150-out for the 5′ junction, and Ap2 and R150-out for the 3′ junction.

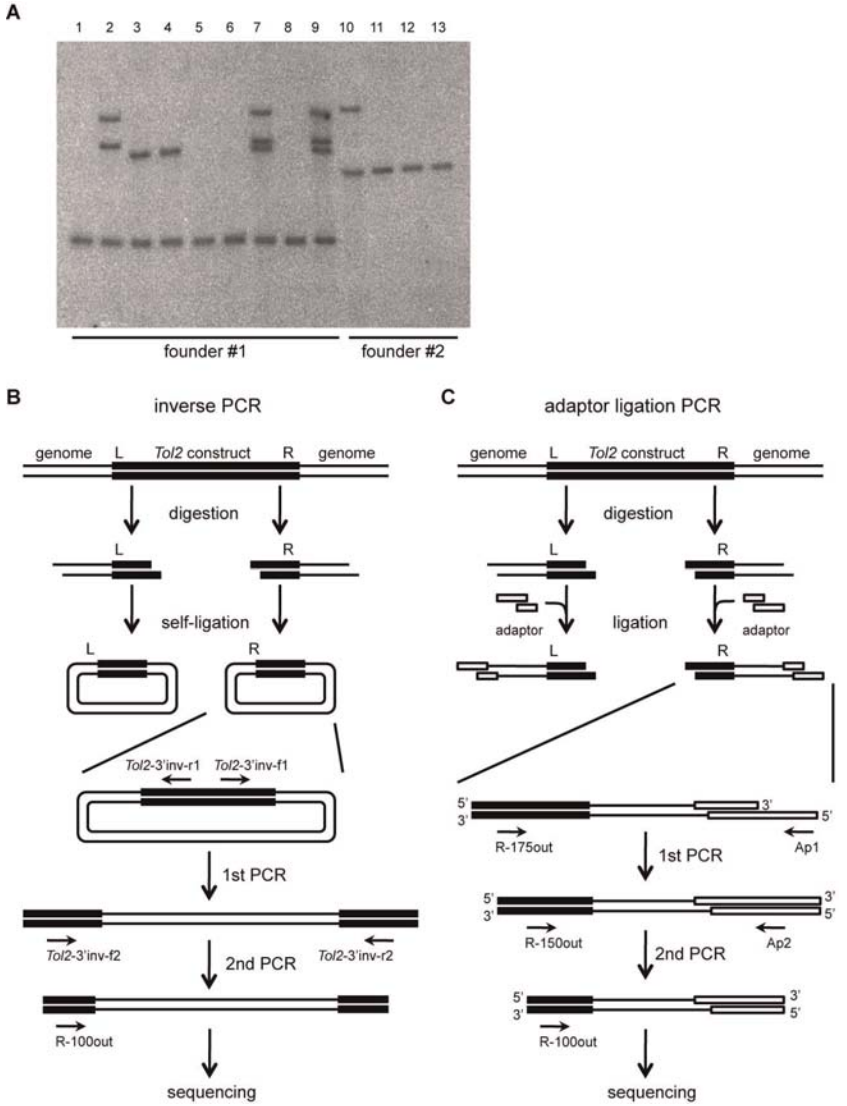


Fig. 5 Molecular characterization of *Tol2* insertions. **(A)** Southern blot hybridization analysis. *Lanes 1–9* and *10–13* represent F1 fish from two different founder fish. Fish carrying a single *Tol2* insertion are identified. **(B)** A scheme for inverse PCR. Genomic DNA is digested with an appropriate restriction enzyme. DNA fragments that contain the *Tol2* sequence (thick lines) and flanking genomic sequences (thin lines) are self-ligated to form circular DNAs. Then, DNA containing *Tol2*/genomic DNA junctions is amplified by two-rounds of PCR using primers directed away from the central *Tol2* sequence. Arrows show positions and directions of the primers. **(C)** A scheme for adaptor-ligation PCR. Genomic DNA is digested with an appropriate restriction enzyme and ligated to adaptors (open boxes). Then, the junction fragments containing *Tol2* and genomic DNA are amplified by two-rounds of PCR using primers in the *Tol2* sequence and the adaptor sequence.

4. Analyze the PCR product on a 1.5% TAE–agarose gel, then purify and sequence the DNA using primers L100-out for the 5′ junction, and R100-out for the 3′ junction (Fig. 5C).

5. Search for Useful Gene Trap and Enhancer Trap Fish on the zTrap Database

1. Useful gene trap and enhancer trap fish are searchable through the *z*Trap database that we described recently (Kawakami *et al*., 2010). Go to http://kawakami.lab.nig.ac.jp/ztrap/ and click the anatomical region of interest (Fig. 4C). Transgenic fish that express GFP (or Gal4FF, see below) in specific cells, tissues, and organs of interests will be shown. Information about the integration sites of the *Tol2* constructs can also be seen.

C. Materials

Gene trap and enhancer trap constructs: For gene trapping, a plasmid containing the T2KSAG construct is used. T2KSAG is composed of a splice acceptor from rabbit β-globin intron, the EGFP gene, and the SV40 polyA signal (Kawakami *et al*., 2004) (Fig. 4A). For enhancer trapping, a plasmid containing the T2KHG construct is used. T2KHG is composed of the zebrafish *hsp70* promoter, the EGFP gene, and the SV40 polyA signal (Nagayoshi *et al*., 2008) (Fig. 4A). Other enhancer trap constructs may also be used (Parinov *et al*., 2004).

Primers used to synthesize the GFP and Gal4FF probes:

GFP-f1: 5′- ATG GTG AGC AAG GGC GAG GA-3′
GFP-r1: 5′-ACT TGT ACA GCT CGT CCA TGC CG-3′
Gal4FF-f2: 5′-ATG AAG CTA CTG TCT TCT-3′
Gal4FF-r2: 5′-TCT AGA TTA GTT ACC CGG-3′

Primers used for inverse PCR:

Tol2-5′inv-f1: 5′- GTC ATG TCA CAT CTA TTA CCA C-3′
Tol2-5′inv-r1 : 5′- CTC AAG TAA AGT AAA AAT CC-3′
Tol2-3′inv-f1: 5′- AGT ACA ATT TTA ATG GAG TAC T-3′
Tol2-3′inv-r1 : 5′- TGA GTA TTA AGG AAG TAA AAG T-3′
Tol2-5′inv-f2 : 5′- AAT GCA CAG CAC CTT GAC CTG G-3′
Tol2-5′inv-r2 : 5′- CAG TAA TCA AGT AAA ATT ACT C-3′
Tol2-3′inv-f2 : 5′- -TTT ACT CAA GTA AGA TTC TAG-3′
Tol2-3′inv-r2 : 5′- AAA GCA AGA AAG AAA ACT AGA G-3′

Adaptor and primers used for adapter ligation PCR:

GATC adaptor: AL (5′-CTA ATA CGA CTC ACT ATA GGG CTC GAG CGG CCG CCG CGG GGG CAG GT-3′) is hybridized with GATC-AS (5′-GAT CAC CTG CCC CCG CTT-3′).
Ap1: 5′-GGA TCC TAA TAC GAC TCA CTA TAG GG-3′

L175-out: 5′-CCC GAA TTC GGC TCG AGC CGG GCC CTT TTT GAC TGT AAA TAA AAT TG-3′
R175-out: 5′-CCC GAA TTC GGA GAT CTT TCT TGC TTT TAC TTT TAC TTC C-3′
Ap2: 5′-CAC TAT AGG GCT CGA GCG G-3′
L150-out: 5′-CCC GAA TTC GGC TCG AGC CGG GCC CGA GTA AAA AGT ACT TTT TTT TCT-3′
R150-out: 5′-CCC GAA TTC GGA GAT CTA ATA CTC AAG TAC AAT TTT A-3′
L100-out: 5′-CCC GAA TTC GGC TCG AGC CGG GCC CAG TAT TGA TTT TTA ATT GTA-3′
R100-out: 5′-CCC GAA TTC GGA GAT CTA GAT TCT AGC CAG ATA CT-3′

D. Discussion

Gene trap and enhancer trap methods are important and powerful for genetic studies in zebrafish. First, transgenic fish expressing GFP in specific tissues and cells are useful for studying morphogenesis and organogenesis. Second, analysis of the genomic DNA surrounding transposon insertions identifies rapidly genes expressed in specific patterns. Third, genomic analysis reveals the enhancers/promoters controlling such specific expression patterns. Finally, these insertions may disrupt important developmental genes, and the mutant phenotypes can be studied.

We previously performed large scale screens by gene trapping and enhancer trapping (Kawakami *et al*., 2004; Kotani *et al*., 2006; Nagayoshi *et al*., 2008). One hundred fifty six fish injected with pT2KSAG and 77 fish injected with pT2KHG were crossed with wild-type fish or inbred. From them, we identified 36 and 125 unique GFP expression patterns, respectively. Further, we identified one and two recessive mutants among the transgenic fish, respectively (Kotani and Kawakami, 2008; Nagayoshi *et al*., 2008).

To clarify the relationship between an insertion and an expression pattern, it is important to identify fish with single insertions. Such fish may be identified in F1s by Southern blot analysis. When all of the F1 fish analyzed carry multiple insertions, we recommend outcrossing the fish with the smallest number of insertions to wild-type fish to obtain fish with single insertions in the next generation.

V. Gene Trapping and Enhancer Trapping with the Gal4FF-UAS System

A. Rationale

The yeast transcriptional activator Gal4 binds a specific DNA sequence called UAS (upstream activating sequence) and activates transcription of a gene placed downstream of the UAS. We constructed a gene trap construct (T2KSAGFF) that

contains a splice acceptor and the *gal4ff* gene and an enhancer trap construct (T2KhspGFF) that contains the zebrafish *hsp70* promoter and the *gal4ff* gene (Fig. 6A) (Asakawa *et al*., 2008). Gal4FF is a modified version of the yeast transcription activator Gal4, which has the Gal4 DNA-binding domain fused to two short transcription activator segments from the herpes simplex viral protein VP16 (Baron *et al*., 1997). Plasmids carrying these constructs are injected with the transposase mRNA into fertilized eggs. By crossing the injected fish with the UAS:GFP reporter fish, Gal4FF transgene insertion and expression are identified as GFP expression in the offspring (Fig. 6B). Thus, a large number of transgenic fish expressing Gal4FF in specific cells, tissues, and organs have been created (Fig. 6C).

The merit of the Gal4FF-UAS system is that one can express desired genes in the Gal4FF-expressing cells. For instance, when the UAS:TeTxLC:CFP effector fish, which carries a fusion of the tetanus toxin light-chain (TeTxLC) gene and CFP gene downstream of UAS, is crossed with a transgenic fish that expresses Gal4FF in subpopulations of neurons (Fig. 7B), the activity of the Gal4FF-expressing neurons is reduced or abolished and behavioral abnormalities are observed (Asakawa *et al*., 2008) (Fig. 7C).

B. Methods

1. Gene Trap and Enhancer Trap Screens for Specific Gal4FF Expression

1. Perform microinjection using a plasmid harboring T2KSAGFF or T2KhspGFF. Raise the injected fish to sexual maturity.
2. Cross the injected fish (founder fish) with UAS:GFP reporter fish. GFP is expressed where Gal4FF is expressed. Collect GFP-positive F1 embryos under a fluorescent microscope at different stages (e.g., 24 hpf, 48 hpf, 72 hpf, and 5 dpf). Raise the embryos to adulthood.
3. To identify an insertion responsible for the expression pattern, analyze genomic DNA by Southern blot hybridization and inverse PCR or adaptor-ligation PCR (see above).

2. Targeted Gene Expression with the Gal4–UAS System and Its Application to the Study of Neural Circuits

1. Cross transgenic fish lines that express Gal4FF in specific subpopulations of neurons with UAS:TeTxLC:CFP effector fish (Fig. 7). The Gal4FF-expressing fish are usually maintained as double transgenic for Gal4FF and UAS:GFP. Use a Gal4FF transgenic male and a UAS:TeTxLC:CFP transgenic female to avoid maternal expression of Gal4FF and GFP.
2. Pick embryos that show CFP but not GFP fluorescence at 24 hpf. These embryos are doubly transgenic for Gal4FF and UAS:TeTxLC:CFP.

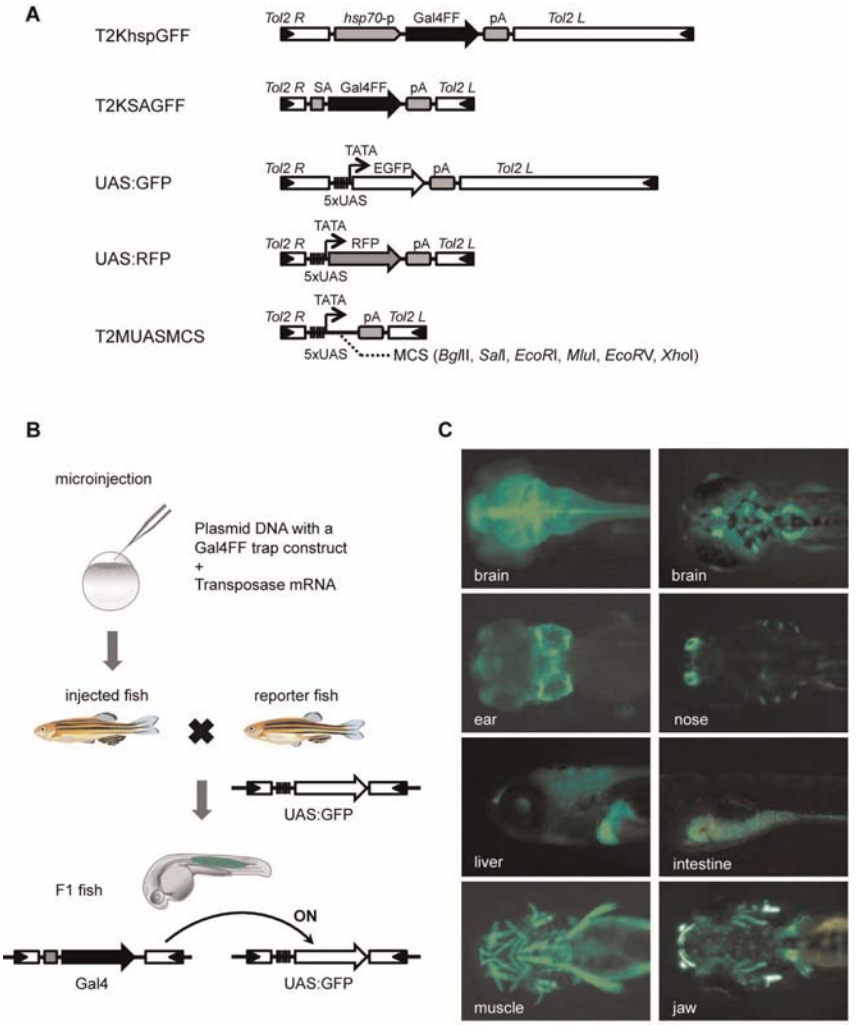


Fig. 6 *Tol2*-mediated gene trap and enhancer trap screens adapted to the Gal4FF-UAS system. **(A)** The structures of the *Tol2* constructs for the Gal4FF-UAS system. The T2KhspGFF enhancer trap construct contains the *hsp70* promoter and the *gal4ff* gene. The T2KSAGFF gene trap construct contains the splice acceptor (SA) from the rabbit β-globin gene and the *gal4ff* gene. UAS:GFP and UAS:RFP contain five tandem repeats of the Gal4-recognition sequence (5x UAS) followed by a minimal TATA sequence, the GFP and the RFP gene respectively, and SV40 polyA signal (pA). T2MUASMCS is a cloning vector that contains 5x UAS, followed by a minimal TATA sequence and a multiple cloning site. **(B)** A scheme for gene trapping and enhancer trapping. The trap constructs containing Gal4FF are injected into fertilized eggs with the transposase mRNA. Injected fish are raised and mated with homozygous UAS:GFP reporter fish. Doubly transgenic F1 embryos express GFP in regions where Gal4FF is expressed. **(C)** Examples of specific GFP expression patterns that are caused by specific Gal4FF expression. (See color plate.)

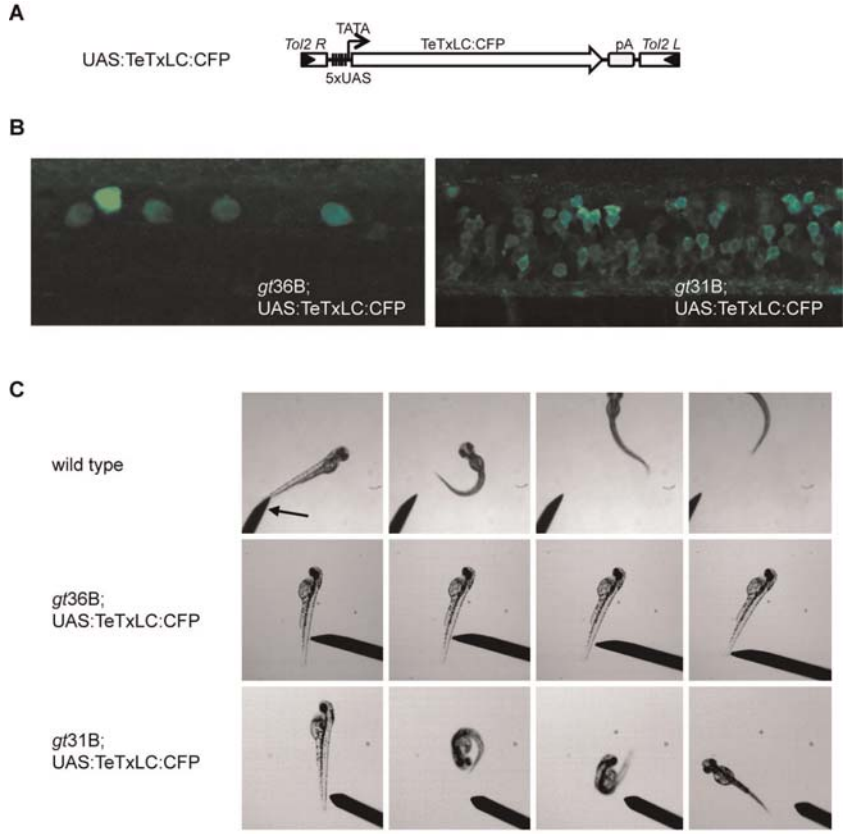


Fig. 7 Application of the Gal4FF-UAS system to the genetic study of neural circuits. **(A)** The structure of UAS:TeTxLC:CFP. UAS:TeTxLC:CFP contains the tetanus toxin light-chain gene (TeTxLC) fused to the CFP gene downstream of 5x UAS. **(B)** Expression of the TeTxLC:CFP fusion protein in Gal4FF;UAS:TeTxLC:CFP double transgenic embryos at 48 hpf as revealed by immunostaining using anti-GFP antibody. The *gt*36B and *gt*31B gene trap insertions give rise to Gal4FF expression in different neuronal types in the spinal cord. **(C)** Behavioral analysis of Gal4FF;UAS:TeTxLC:CFP double transgenic embryos at 2 dpf by touch response assay. Escape behaviors of 2-day-old wild type and double transgenic embryos in response to gentle touch to the tail are analyzed by high-speed digital recording. A wild-type embryo rapidly escapes from a gentle touch to the tail (touch response). The *gt*36B;UAS:TeTxLC:CFP transgenic fish, which expresses TeTxLC:CFP in sensory neurons, does not respond to the touch. The *gt*31B;UAS:TeTxLC:CFP transgenic fish, which expresses TeTxLC:CFP in subsets of interneurons and motor neurons, responds to touch but shows abnormal escape swimming. (See color plate.)

3. Place 48 hpf Gal4FF;UAS:TeTxLC:CFP double transgenic embryos in a plastic dish.
4. Touch the tail gently three times with a needle. Take images by using a high-speed digital video camera (FASTCAM-512PC1, Photoron, Japan). At this stage, wild-type embryos respond to the gentle touch to the tail and swim rapidly away from

the stimulus. However, Gal4FF;UAS:TeTxLC:CFP double transgenic larvae, in which activities of distinct spinal neural circuits are inhibited, display behavioral abnormalities (Fig. 7C).

C. Materials

Gene trap and enhancer trap constructs for the Gal4FF-UAS system: The pT2KSAGFF gene trap construct contains the *gal4ff* gene downstream of a rabbit β-globin splice acceptor (Fig. 6A). The pT2KhspGFF enhancer trap construct contains the *gal4ff* gene downstream of the zebrafish *hsp70* promoter (Fig. 6A).

UAS reporter and effector lines: UAS:GFP fish contain the GFP gene downstream of 5xUAS and UAS:RFP fish contains the mRFP gene downstream of 5xUAS (Fig. 6A). UAS:TeTxLC:CFP effecter fish carry a fusion of the tetanus toxin light-chain (TeTxLC) gene and the CFP gene downstream of 5x UAS (Fig. 7A). TeTxLC cleaves a vesicle membrane protein Synaptobrevin and thereby blocks neurotransmitter release from synaptic vesicles (Schiavo *et al*., 1992).

T2MUASMCS: The T2MUASMCS vector contains a multicloning site (MCS) between 5xUAS and the polyA signal. This vector is used to create UAS reporter and effector constructs (Suster *et al*., 2009a).

D. Discussion

We employed Gal4FF to develop the Gal4-UAS system in zebrafish. In previous studies, full-length Gal4 and Gal4-VP16 were used in zebrafish (Koster and Fraser, 2001; Scheer and Campos-Ortega, 1999). Full-length Gal4 turned out to be a weak transcription activator. Gal4-VP16, which contains a strong transcription activator domain, gave greater activity than full-length Gal4 but also caused developmental abnormalities (Koster and Fraser, 2001; Scott *et al*., 2007). In contrast, Gal4FF is sufficiently strong to induce transcription from UAS without developmental toxicity (Asakawa *et al*., 2008).

We have observed different levels of reporter and effector gene expression when the same Gal4FF fish were crossed with fish carrying the identical 5×UAS transgenes integrated at different chromosomal loci, which indicates that expression from the 5×UAS transgene is sensitive to position effects. Therefore, it is important to create many lines of 5×UAS transgenic fish so that one may select a line that that gives optimal levels of reporter expression when crossed with a Gal4FF tester line. In fact, the UAS:TeTxLC:CFP fish was established by testing 75 different insertions for their abilities to inhibit neural functions. It is noteworthy that UAS-reporter and UAS-effector insertions thus established are often located within regions that are thought to be transcriptionally active.

The Gal4FF-UAS system can be applied to manipulate the function of specific cell types. We have demonstrated that neuronal functions can be inhibited by the

Table I
UAS reporter and effecter fish lines

UAS:gene name	Description	Reference
UAS:*myc-notch1a-intra*	An activated form of the Notch receptor	Scheer and Campos-Ortega, 1999
UAS:*fezl*	Forebrain embryonic zinc-finger like transcription factor	Jeong *et al*., 2007
UAS:*nfsB-mCherry*	Nitroreductase tagged with mCherry	Davison *et al*., 2007
UAS:iGluR6(L439C)	Light-gated iGluR6 receptor	Szobota *et al*., 2007
UAS:GFP	EGFP	Asakawa *et al*., 2008
UAS:RFP	monomeric RFP	Asakawa *et al*., 2008
UAS:TeTxLC	Tetanus toxin light chain	Asakawa *et al*., 2008
UAS:TeTxLC:CFP	Tetanus toxin light chain tagged with CFP	Asakawa *et al*., 2008
UAS:GCaMP1.6	A high-affinity Ca^{2+} probe composed of a single GFP	Sumbre *et al*., 2008
UAS:NpHR-mCherry	Halorhodopsin (NpHR) tagged with mCherry	Arrenberg *et al*., 2009
UAS:rvcyclin	A retrovirus homologue gene of cellular cyclin D1	Zhan *et al*., 2010
UAS:Gtuba2	Tubulin tagged with GFP	Asakawa *et al*., 2010

UAS:TeTxLC:CFP effector fish (Asakawa *et al*., 2008). Other UAS effector fish including UAS:myc-notch1a:intra (Scheer and Campos-Ortega, 1999), UAS:*fezl* (Jeong *et al*., 2007), UAS:nfsB-mCherry (Davison *et al*., 2007), UAS:iGluR6 (L439C) (Szobota *et al*., 2007), UAS:NpHR- eYFP (Arrenberg *et al*., 2009), UAS: rvcyclin (Zhan *et al*., 2010), UAS:GCaMP1.6 (Sumbre *et al*., 2008), and UAS: Gtuba2 (Asakawa and Kawakami, 2010) have also been created (Table 1). Further development of various UAS effector lines should increase the utility of the Gal4FF-UAS strategy for tissue- and organ-specific transgene expression in zebrafish.

Acknowledgments

We express our gratitude for support from the Sasakawa scientific research grant of the Japan Science Society (to GA), the National BioResource Project, and grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Alper, M. D., and Ames, B. N. (1975). Positive selection of mutants with deletions of the *gal-chl* region of the *Salmonella* chromosome as a screening procedure for mutagens that cause deletions. *J. Bacteriol.* **121**, 259–266.

Arrenberg, A. B., Del Bene, F., and Baier, H. (2009). Optical control of zebrafish behavior with halorhodopsin. *Proc. Natl Dev. Sci. U. S. A.* **106**, 17968–17973.

Asakawa, K., and Kawakami, K. (2010). A transgenic zebrafish for monitoring *in vivo* microtubule structures. *Dev. Dyn.* **239**, 2695–2699.

Asakawa, K., Suster, M. L., Mizusawa, K., Nagayoshi, S., Kotani, T., Urasaki, A., Kishimoto, Y., Hibi, M., and Kawakami, K. (2008). Genetic dissection of neural circuits by *Tol2* transposon-mediated Gal4 gene and enhancer trapping in zebrafish. *Proc. Natl Acad. Sci. U. S. A.* **105**, 1255–1260.

Balciunas, D., Wangensteen, K. J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P. B., Largaespada, D. A., McIvor, R. S., and Ekker, S. C. (2006). Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet.* **2**, e169.

Baron, U., Gossen, M., and Bujard, H. (1997). Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. *Nucleic Acids Res.* **25**, 2723–2729.

Davison, J. M., Akitake, C. M., Goll, M. G., Rhee, J. M., Gosse, N., Baier, H., Halpern, M. E., Leach, S. D., and Parsons, M. J. (2007). Transactivation from Gal4-VP16 transgenic insertions for tissue-specific cell labeling and ablation in zebrafish. *Dev. Biol.* **304**, 811–824.

Jeong, J. Y., Einhorn, Z., Mathur, P., Chen, L., Lee, S., Kawakami, K., and Guo, S. (2007). Patterning the zebrafish diencephalon by the conserved zinc-finger protein. *Fezl. Dev.* **134**, 127–136.

Kawakami, K. (2007). *Tol2*: a versatile gene transfer vector in vertebrates. *Genome Biol.* **8**(Suppl 1), S7.

Kawakami, K., Koga, A., Hori, H., and Shima, A. (1998). Excision of the *Tol2* transposable element of the medaka fish, *Oryzias latipes*, in zebrafish, *Danio rerio*. *Gene* **225**, 17–22.

Kawakami, K., and Shima, A. (1999). Identification of the *Tol2* transposase of the medaka fish *Oryzias latipes* that catalyzes excision of a nonautonomous *Tol2* element in zebrafish *Danio rerio*. *Gene* **240**, 239–244.

Kawakami, K., Shima, A., and Kawakami, N. (2000). Identification of a functional transposase of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. U. S. A.* **97**, 11403–11408.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144.

Kawakami, K., Abe, G., Asada, T., Asakawa, K., Fukuda, R., Ito, A., Lal, P., Mouri, N., Muto, A., Suster, M. L., Takakubo, H., Urasaki, A., Wada, H., and Yoshida, M. (2010). *z*Trap: zebrafish gene trap and enhancer trap database. *BMC Dev. Biol.* **10**(1), 105.

Koga, A., Suzuki, M., Inagaki, H., Bessho, Y., and Hori, H. (1996). Transposable element in fish. *Nature* **383**, 30.

Koster, R. W., and Fraser, S. E. (2001). Tracing transgene expression in living zebrafish embryos. *Dev. Biol.* **233**, 329–346.

Kotani, T., Nagayoshi, S., Urasaki, A., and Kawakami, K. (2006). Transposon-mediated gene trapping in zebrafish. *Methods* **39**, 199–206.

Kotani, T., and Kawakami, K. (2008). *Misty somites*, a maternal effect gene identified by transposon-mediated insertional mutagenesis in zebrafish that is essential for the somite boundary maintenance. *Dev. Biol.* **316**, 383–396.

Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P., and Chien, C. B. (2007). The Tol2kit: a multisite gateway-based construction kit for *Tol2* transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099.

Nagayoshi, S., Hayashi, E., Abe, G., Osato, N., Asakawa, K., Urasaki, A., Horikawa, K., Ikeo, K., Takeda, H., and Kawakami, K. (2008). Insertional mutagenesis by the *Tol2* transposon-mediated enhancer trap approach generated mutations in two developmental genes: *tcf7* and *synembryn-like*. *Development* **135**, 159–169.

Parinov, S., Kondrichin, I., Korzh, V., and Emelyanov, A. (2004). *Tol2* transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes *in vivo*. *Dev. Dyn.* **231**, 449–459.

Rupp, R. A., Snider, L., and Weintraub, H. (1994). *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311–1323.

Scheer, N., and Campos-Ortega, J. A. (1999). Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. *Mech. Dev.* **80**, 153–158.

Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B. R., and Montecucco, C. (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835.

Scott, E. K., Mason, L., Arrenberg, A. B., Ziv, L., Gosse, N. J., Xiao, T., Chi, N. C., Asakawa, K., Kawakami, K., and Baier, H. (2007). Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat. Methods* **4**, 323–326.

Stuart, G. W., McMurray, J. V., and Westerfield, M. (1988). Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**, 403–412.

Sumbre, G., Muto, A., Baier, H., and Poo, M. M. (2008). Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval. *Nature* **456**, 102–106.

Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K., and Kawakami, K. (2009a). Transgenesis in zebrafish with the *Tol2* transposon system. *Methods Mol. Biol.* **561**, 41–63.

Suster, M. L., Sumiyama, K., and Kawakami, K. (2009b). Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* **10**, 477.

Szobota, S., Gorostiza, P., Del Bene, F., Wyart, C., Fortin, D. L., Kolstad, K. D., Tulyathan, O., Volgraf, M., Numano, R., Aaron, H. L., Scott, E. K., Kramer, R. H., Flannery, J., Baier, H., Trauner, D., and Isacoff, E. Y. (2007). Remote control of neuronal activity with a light-gated glutamate receptor. *Neuron* **54**, 535–545.

Turner, D. L., and Weintraub, H. (1994). Expression of *achaete-scute homolog 3* in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8**, 1434–1447.

Urasaki, A., Morvan, G., and Kawakami, K. (2006). Functional dissection of the *Tol2* transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649.

Villefranc, J. A., Amigo, J., and Lawson, N. D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev. Dyn.* **236**, 3077–3087.

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic Acids Res.* **33**, e36.

Yang, Z., Jiang, H., Chachainasakul, T., Gong, S., Yang, X. W., Heintz, N., and Lin, S. (2006). Modified bacterial artificial chromosomes for zebrafish transgenesis. *Methods* **39**, 183–188.

Zhan, H., Spitsbergen, J. M., Qing, W., Wu, Y. L., Paul, T. A., Casey, J. W., Her, G. M., and Gong, Z. (2010). Transgenic expression of Walleye dermal sarcoma virus *rv-cyclin* gene in zebrafish and its suppressive effect on liver tumor development after carcinogen treatment. *Mar. Biotechnol. (NY)* **12**, 640–649.

CHAPTER 3

Engineering Zinc Finger Nucleases for Targeted Mutagenesis of Zebrafish

Jeffry D. Sander,[*,†] J.-R. Joanna Yeh,[‡,§] Randall T. Peterson[‡,§,¶] and J. Keith Joung[*,†]

[*]Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA

[†]Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

[‡]Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, USA

[§]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[¶]Broad Institute, Cambridge, Massachusetts, USA

Abstract

Zinc finger nucleases provide an important platform for performing reverse genetics in zebrafish. Here we review various methods and resources that have been used to create customized zinc finger nucleases for use in zebrafish. We also provide a framework for choosing among the various publicly available platforms available to engineer ZFNs.

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00003-3

I. Introduction

Engineered zinc finger nucleases (ZFNs) enable the rapid introduction of targeted mutations in zebrafish. ZFNs consist of an engineered zinc finger array, typically composed of three or four fingers, fused to the non-specific cleavage domain from the Type IIS restriction enzyme *Fok*I (Fig. 1A). ZFNs function as dimers, with each monomer binding to "half-sites" separated by a spacer sequence that is cleaved by two *Fok*I domains (Fig. 1B). Work from the laboratories of Amacher, Wolfe, and Lawson and from scientists at Sangamo BioSciences first demonstrated that repair of ZFN-induced double-strand breaks (DSBs) by non-homologous end-joining can lead to the introduction of insertion or deletion (indel) mutations at the site of the break (Doyon *et al*., 2008; Meng *et al*., 2008).

For the zebrafish community, the most challenging step of this technology has been engineering the zinc finger arrays required to construct customized ZFNs. To date, five different methods have been used to engineer ZFNs for mutagenesis of zebrafish: modular assembly (Cifuentes *et al*., 2010; Siekmann *et al*., 2009), Oligomerized Pool Engineering (OPEN) (Foley *et al*., 2009), bacterial one-hybrid selection (Meng *et al*., 2008), Context-Dependent Assembly (CoDA) (Sander *et al*., 2011), and the proprietary method of Sangamo BioSciences (Doyon *et al*., 2008). In this chapter, we briefly review the various methods and associated resources used to engineer zinc finger arrays and then propose a framework for considering and utilizing different methods based on our collective experience to date.

II. Methods and Resources for Engineering Zinc Finger Nucleases

A. Modular Assembly

The modular assembly method for engineering zinc finger arrays involves the linking together of individual fingers pre-selected to bind to specific DNA

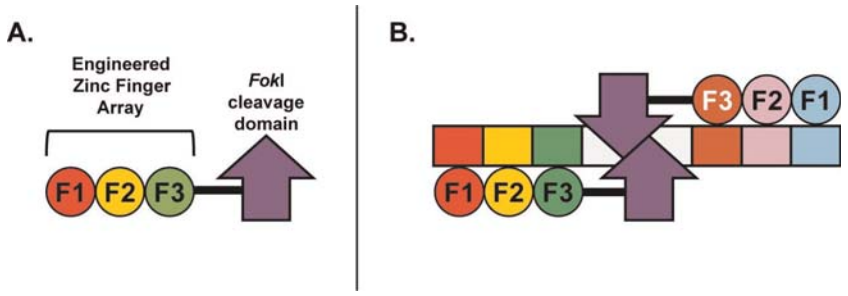


Fig. 1 Schematics illustrating (A) a zinc finger nuclease and (B) a dimer of zinc finger nucleases cleaving DNA. (See color plate.)

sequences. The Barbas lab, Sangamo BioSciences, and Toolgen, Inc. have each described archives of finger modules that can be used for modular assembly. A major appeal of modular assembly is its simplicity – practicing the method requires only standard molecular biology techniques. However, modular assembly also has been shown to have a very high (>94%) failure rate for engineering ZFN pairs (Joung *et al*., 2010; Kim *et al*., 2009; Ramirez *et al*., 2008), most likely due to its failure to account for the well-established context-dependent activities of zinc fingers in an array (Elrod-Erickson *et al*., 1996; Isalan *et al*., 1997, 1998; Wolfe *et al*., 1999, 2001). Web-based software and/or published protocols exist for using the Barbas modules (http://www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php) (Mandell and Barbas, 2006), the Barbas and Sangamo modules (Carroll *et al*., 2006), or all three module sets (http://zifit.partners.org) (Sander *et al*., 2007; Wright *et al*., 2006). Reagents for practicing modular assembly are available through the non-profit plasmid distribution service Addgene (http://www.addgene.org/pgvec1?f=c&cmd=showcol&colid=235&page=6).

B. Oligomerized Pool Engineering (OPEN)

OPEN is a robust method in which combinatorial libraries of zinc finger arrays are interrogated using a bacterial two-hybrid (B2H) selection method to identify candidates that bind efficiently to target DNA sites of interest (Maeder *et al*., 2008, 2009). OPEN explicitly accounts for context-dependent effects among zinc fingers in an array and yields high quality ZFNs that have been successfully used to modify multiple endogenous genes in zebrafish (Foley *et al*., 2009). However, the need to construct a combinatorial library for each target site and to practice the B2H selection method make OPEN more challenging to practice than modular assembly. Potential OPEN target sites can be identified in genes of interest using the web-based ZiFiT (http://zifit.partners.org) (Sander *et al*., 2010) or ZFNGenome (http://bindr.gdcb.iastate.edu:88/) (Reyon *et al*., 2011) programs. A detailed protocol for practicing OPEN has been described (Maeder *et al*., 2009) and is freely available online (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2858690/). The OPEN pools are available by request from the Joung lab (http://www.jounglab.org/contact.htm) and other reagents required to practice OPEN are available from Addgene (http://www.addgene.org/pgvec1?f=c&cmd=showcol&colid=235&page=5).

C. Bacterial One-Hybrid (B1H) Selection Method

Wolfe and colleagues have described a selection-based method similar to OPEN and other previously described selection-based strategies (Hurt *et al*., 2003). This method also builds combinatorial libraries to account for context-dependent effects among zinc fingers. However, although the approach uses a somewhat simplified B1H selection system (Meng *et al*., 2008) based on the original B2H system (Dove *et al*., 1997; Joung *et al*., 2000), it uses a more limited finger randomization strategy

than that employed by OPEN. In addition, this method is more labor-intensive than OPEN because it requires the construction of four combinatorial libraries and the performance of four B1H selections to identify zinc finger arrays for a single target site. Web-based software for practicing this method are available (http://pgfe.umassmed.edu/ZFPsearch.html) and reagents can be obtained from Addgene (http://www.addgene.org/Scot_Wolfe).

D. Context-Dependent Assembly (CoDA)

We recently described a new method for engineering multi-finger arrays that accounts for context-dependent effects between neighboring fingers but avoids the need to build combinatorial libraries or perform labor-intensive selections. This method assembles three-finger arrays using an archive of pre-selected finger units known to work well when positioned adjacent to one another (Sander *et al.*, 2011). Using CoDA, researchers can construct zinc finger arrays using either standard cloning techniques or commercial DNA synthesis. A limitation of CoDA is its current targeting range of only 1 in 500 bps of random DNA sequence. The web-based ZiFiT program can be used to identify potential CoDA sites in genes of interest and also provides DNA sequences encoding zinc finger arrays for commercial DNA synthesis (http://zifit.partners.org/ZiFiT/).

E. Proprietary Method of Sangamo BioSciences

Sangamo BioSciences has developed a proprietary platform for engineering ZFNs (Doyon *et al.*, 2008). Researchers interested in acquiring ZFNs made by this method can purchase these proteins from Sigma-Aldrich through their CompoZr product line (http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology.html).

F. ZFN Expression Vectors

Zinc finger arrays engineered using OPEN, CoDA, or modular assembly can be cloned into various ZFN expression vectors available from Addgene (http://www.addgene.org/pgvec1?f=c&cmd=showcol&colid=235&page=7). These vectors are designed to express ZFNs encoding either homodimeric or heterodimeric (Miller *et al.*, 2007) *Fok*I nuclease domains with linkers between the zinc finger array and the nuclease domain required to target sites with spacer sequences of 5, 6, or 7 base pairs.

III. Choice of Method to Engineer Zinc Finger Nucleases

In this final section, we outline a "decision tree" for using the methods described above to engineer ZFNs for a zebrafish gene of interest. Our overall framework is

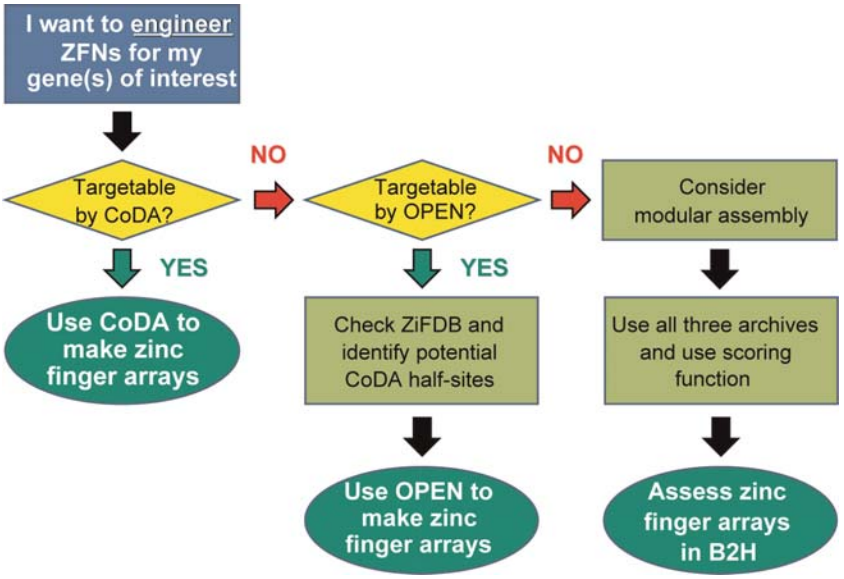


Fig. 2 "Decision tree" used to prioritize various zinc finger engineering methods when targeting endogenous genes in zebrafish. (See color plate.)

illustrated in Fig. 2 and presumes that the researcher is interested in and deems it appropriate to engineer ZFNs themselves.

CoDA is our preferred first line method for targeting a gene due to its rapidity, simplicity, and high efficacy. Using the ZiFiT program, we identify potential ZFN sites within the gene that can be targeted by CoDA. If potential sites are identified, we construct those ZFNs and test them in zebrafish.

If no CoDA sites are identified, we use ZiFiT to identify potential sites that can be targeted by OPEN. Although it has a high efficacy rate, OPEN is our second line choice due to the additional effort required to perform selections. For any potential OPEN ZFN sites that are identified, we first check the web-based Zinc Finger Database (ZiFDB) program (http://bindr.gdcb.iastate.edu:8080/ZiFDB/) (Fu *et al*., 2009) to see if OPEN or CoDA have already been used to make zinc finger arrays for any of the target half-sites; this can be easily done through hyperlinks to ZiFDB provided in the ZiFiT output. We also check to see whether CoDA can be used to make a zinc finger array for one of the two "half-sites" in each ZFN target site. These checks are worth making to avoid performing any unnecessary OPEN selections. OPEN selections are then performed for any other "half-sites" needed.

If no OPEN sites can be identified in the gene of interest, one can contemplate using modular assembly to engineer ZFNs. As noted above, modular assembly has been reported to have a low success rate. Therefore, to maximize chances of success,

we recommend: (1) use of all three publicly available module archives, (2) careful choice of target sites that have a high probability of success, and (3) testing of zinc finger arrays in the B2H reporter system prior to use as ZFNs in zebrafish. Recommendations (1) and (2) can be easily met by using the ZiFiT program, which identifies sites and provides protein sequences using all three module sites and provides a scoring function based on subsite composition (Ramirez *et al*., 2008) that indicates likelihood of success. Protocols for evaluating zinc finger arrays using the B2H reporter system have been previously published (Maeder *et al*., 2009; Wright *et al*., 2006).

For any of the methods above, it is desirable to target multiple potential sites within a gene to increase the chances of identifying a pair of ZFNs that can mutagenize efficiently *in vivo*. We recommend that users exhaust all possible sites that can be targeted with a given method before moving to the next method in the sequence described above.

As is evident from the above, the process of engineering ZFNs for a target gene is one that requires understanding and careful consideration of various options. The Zinc Finger Consortium, a group of academic researchers committed to the development of engineered zinc finger technology (http://www.zincfingers.org) (Pearson, 2008), also moderates an active newsgroup on issues related to the use of ZFNs that currently has over 300 members. Academic ZFN users are encouraged to join the newsgroup here: http://www.zincfingers.org/listserv.htm.

Acknowledgments

The authors acknowledge support from the National Institutes of Health (R01GM088040 (J.K.J. & R.T.P.), DP1OD006862 (J.K.J.), R01 CA0140188 (J.-R.J.Y), T32CA009216 (J.D.S.), and K01 AG031300 (J.-R.J.Y)) and the Massachusetts General Hospital (Claflin Distinguished Scholar Award (J.-R.J.Y)). We thank Frances Toneguzzo and Paul Cushing for helpful comments.

References

Carroll, D., Morton, J. J., Beumer, K. J., and Segal, D. J. (2006). Design, construction and *in vitro* testing of zinc finger nucleases. *Nat. Protoc.* **1**, 1329–1341.

Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N., Wolfe, S., and Giraldez, A. J. (2010). A novel miRNA processing pathway independent of dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694–1698.

Dove, S. L., Joung, J. K., and Hochschild, A. (1997). Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**, 627–630.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702–708.

Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171–1180.

Foley, J. E., Yeh, J. R., Maeder, M. L., Reyon, D., Sander, J. D., Peterson, R. T., and Joung, J. K. (2009). Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE* **4**, e4348.

Fu, F., Sander, J. D., Maeder, M., Thibodeau-Beganny, S., Joung, J. K., Dobbs, D., Miller, L., and Voytas, D. F. (2009). Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res.* **37**, D279–D283.

Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O., and Joung, J. K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl Acad. Sci. U. S. A.* **100**, 12271–12276.

Isalan, M., Choo, Y., and Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl Acad. Sci. U. S. A.* **94**, 5617–5621.

Isalan, M., Klug, A., and Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**, 12026–12033.

Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl Acad. Sci. U. S. A.* **97**, 7382–7387.

Joung, J. K., Voytas, D. F., and Cathomen, T. (2010). Reply to "Genome editing with modularly assembled zinc-finger nucleases". *Nat. Methods* **7**, 91–92.

Kim, H. J., Lee, H. J., Kim, H., Cho, S. W., and Kim, J. S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19**, 1279–1288.

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Muller-Lerch, F., Fu, F., Pearlberg, J., Gobel, C., Dassie, J. P., Pruett-Miller, S. M., Porteus, M. H., Sgroi, D. C., Iafrate, A. J., Dobbs, D., McCray Jr., P. B., Cathomen, T., Voytas, D. F., and Joung, J. K. (2008). Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294–301.

Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F., and Joung, J. K. (2009). Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat. Protoc.* **4**, 1471–1501.

Mandell, J. G., and Barbas 3rd, C. F. (2006). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.* **34**, W516–W523.

Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D., and Wolfe, S. A. (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., and Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785.

Pearson, H. (2008). The fate of fingers. *Nature* **455**, 160–164.

Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A., Cathomen, T., Voytas, D. F., and Joung, J. K. (2008). Unexpected failure rates for modular assembly of engineered zinc-fingers. *Nat. Methods* **5**, 374–375.

Reyon, D., Kirkpatrick, J. R., Sander, J. D., Zhang, F., Voytas, D. F., Joung, J. K., Dobbs, D., and Coffman, C. R. (2011). ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. *BMC Genomics* **12**, 83.

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y., Pierick, C. J., Hoffman, E., Maeder, M. L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D. M., Stupar, R. M., Giraldez, A. J., Voytas, D. F., Peterson, R. T., Yeh, J. R., and Joung, J. K. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* **8**, 67–69.

Sander, J. D., Maeder, M. L., Reyon, D., Voytas, D. F., Joung, J. K., and Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.* **38**(Suppl.), W462–W468.

Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res.* **35**, W599–W605.

Siekmann, A. F., Standley, C., Fogarty, K. E., Wolfe, S. A., and Lawson, N. D. (2009). Chemokine signaling guides regional patterning of the first embryonic artery. *Genes Dev.* **23**, 2272–2277.

Wolfe, S. A., Grant, R. A., Elrod-Erickson, M., and Pabo, C. O. (2001). Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure (Camb).* **9**, 717–723.

Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* **285**, 1917–1934.

Wright, D. A., Thibodeau-Beganny, S., Sander, J. D., Winfrey, R. J., Hirsh, A. S., Eichtinger, M., Fu, F., Porteus, M. H., Dobbs, D., Voytas, D. F., and Joung, J. K. (2006). Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.* **1**, 1637–1652.

CHAPTER 4

Retroviral-mediated Insertional Mutagenesis in Zebrafish

Adam Amsterdam[*], Gaurav Kumar Varshney[†] and Shawn Michael Burgess[†]

[*]Koch Institute for Integrative Cancer Research, Massachusetts Institute for Technology, Cambridge, Massachusetts, USA

[†]National Human Genome Research Institute, Bethesda, Maryland, USA

Abstract

Since the initial publication of this chapter in 2004, additional methodologies have been developed which could improve and/or complement the original retroviral-mediated insertional mutagenesis. Retroviral vectors have also been shown to be

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00004-5

useful for goals other than mutagenesis. In addition, retroviral-mediated insertional mutagenesis has been applied to zebrafish for use in reverse genetics as well as forward screening. Finally, the insertional mutant collection described herein has been screened by a number of labs to find a host of mutants (with genes already identified) with developmental and/or growth defects affecting the eye, liver, skin, craniofacial skeleton, kidney, myeloid cells, hematopoietic stem cells, and axon pathfinding, as well as mutants with defects in the cell cycle or DNA damage response, altered aging properties, and modulated cardiac repolarization. The major complementary approaches and new uses of this technique include:

- Pseudotyped retroviruses have been used to deliver enhancer trap vectors, which allows selection of insertions in or near genes with particular expression patterns. Although the mutagenicity of these vectors has yet to be determined, they are useful purely because they generate a large number of transgenic lines with visible reporters (e.g., Green Fluorescent Protein GFP) expressed in interesting patterns and they provide information regarding gene regulation in the context of genomic organization.
- Gain-of-function vectors have been designed to allow for dominant genetic screens. Thus, genes whose overexpression results in phenotypes of interest can be efficiently identified.
- Retroviruses can be used to make a library of insertions in which hundreds of thousands of mapped insertions can be recovered from frozen sperm samples. Such libraries could serve as on-the-shelf reverse genetic resources, whereby one can obtain a mutation in nearly any gene by simply recovering an insertion in that gene from the frozen sperm, similar to the use of gene-trap insertions and genome-wide gene targeting in ES cells in the mouse.
- Transposons have been shown to be nearly as effective transgenesis vectors as retroviruses and thus may be used in similar screens – both for mutagenicity and gene traps and enhancer traps. It is possible that retroviruses and transposons could have different insertion site biases, making them important complementary technologies for genome-wide screening.

I. Introduction

Large-scale chemical mutagenesis screens have resulted in the isolation of thousands of mutations in hundreds of genes that affect zebrafish embryonic development (Driever *et al*., 1996; Haffter *et al*., 1996). These screens have utilized an alkylating agent, ethyl nitrosourea (ENU), to induce mutations, primarily by causing base pair substitutions. Several hundred of the genes disrupted by these mutations have been isolated over the past 15 years, primarily via a candidate gene approach, and less frequently by pure positional cloning (Postlethwait and Talbot, 1997), and many other chapters in this volume are devoted to describing this task. However, positional cloning remains arduous.

Insertional mutagenesis is an alternative to chemical mutagenesis in which exogenous DNA is used as the mutagen. While insertional mutagenesis is usually less efficient than ENU, insertions serve as a molecular tag to aid in the isolation of the mutated genes. Several methods might be employed to insert DNA into the zebrafish genome, including DNA microinjection (Culp *et al*., 1991; Stuart *et al*., 1988), or microinjection of DNA aided by retroviral integrases (Ivics *et al*., 1993) or a transposable element's transposase (Davidson *et al*., 2003; Kawakami *et al*., 2000b; Raz *et al*., 1998); reviews and updates on these methods can be found in other chapters of this volume. However, to date, by far the most efficient way to make a large number of insertions in the zebrafish genome is to use a pseudotyped retrovirus.

Retroviruses have an RNA genome and, upon infection of a cell, will reverse transcribe their genome to a DNA molecule, the provirus. The provirus integrates into a host cell chromosome where it remains stable and thus is inherited by all of the descendants of that cell. Replication-defective retroviral vectors, unlike non-defective retroviruses, are infectious agents which can integrate into host DNA, but whose genetic material lacks the coding sequences for the proteins required to make progeny virions. Retroviral vectors are made in "split genome" packaging cells, in which the genome of the retroviral vector is expressed from one integrated set of viral sequences, while the retroviral genes required for packaging, infection, reverse-transcription, and integration are expressed from another locus. The most widely used retroviral vectors have been derived from a murine retrovirus, Moloney Murine Leukemia Virus (MoMLV), resulting in replication-defective viruses that can be produced at very high titers. Initially, these retroviruses were only capable of infecting mammalian cells, but their host range can be expanded as described below.

Retroviruses have a host range, or "tropism," that is frequently determined by their envelope protein, which recognizes and binds to some specific component, usually a protein, on the surface of the cell to be infected. Cell types that have an appropriate receptor can be infected by the retrovirus, while those that do not are refractory to infection. The host range of a virus can be changed by pseudotyping, a process in which virions acquire the genome and core proteins of one virus but the envelope protein of another. One way to enable this situation in split genome packaging cells is to simply substitute the gene encoding the alternative envelope protein for the usual one. While there is some specificity as to which envelope proteins can be pseudotyped with which viral genomes, one such combination that is particularly useful allows the MoMLV viral genomes and core proteins to be pseudotyped with the envelope glycoprotein (G-protein) of vesicular stomatitis virus (VSV) (Weiss *et al*., 1974). VSV is a rhabdovirus which is apparently pantropic; it can infect cells of species as diverse as insects and mammals (Wagner, 1972). MoMLV vectors pseudotyped with VSV-G possess two qualities essential for their use in high-frequency germline transgenesis in zebrafish: The extended host range allows for the infection of fish cells, and the VSV-G pseudotyped virions are unusually stable, which permits viruses to be concentrated 1000-fold by centrifugation (Burns *et al*., 1993).

When pseudotyped retroviral vectors are injected into zebrafish blastulae, many of the cells become independently infected, producing a mosaic organism in which different cells harbor proviral insertions at different chromosomal sites. When cells destined to give rise to the germline are infected, some proportion of the progeny of the injected fish will contain one or more insertions (Lin *et al*., 1994). When sufficiently high titer virus is used, one can infect a very high proportion of the germline of injected fish (Gaiano *et al*., 1996a). With very high titer virus, on average, about 25–30 independent insertions can be inherited from a single founder, though any given insertion will only be present in about 3–20% of the offspring (Chen *et al*., 2002). However, the progeny are non-mosaic for the insertions, and transmit them in a Mendelian fashion to 50% of their progeny. Furthermore, because more than one virus can infect a single cell, some germ cells contain multiple insertions, and thus offspring can be born with as many as 10–15 independently segregating insertions (Amsterdam *et al*., 1999; Chen *et al*., 2002; Gaiano *et al*., 1996a). This remarkable transgenesis rate has made it possible to conduct an insertional mutagenesis screen which has isolated over 500 insertional mutants and allowed the rapid cloning of the mutated genes (Amsterdam *et al*., 1999, 2004a; Golling *et al*., 2002).

II. Mutagenesis

In order to establish the frequency of mutagenesis with retroviral vectors in the zebrafish, we carried out a pilot screen in which we inbred over 500 individual proviral insertions, one at a time, and screened for recessive phenotypes which could be visually scored in the first five days of embryonic development. We found six recessive embryonic lethal mutations, a frequency of about one mutation per 80–100 insertions (Allende *et al*., 1996; Becker *et al*., 1998; Gaiano *et al*., 1996b; Young *et al*., 2002). We also found one viable dominant insertional mutation (Kawakami *et al*., 2000a). While this rate was too inefficient to conduct a large-scale screen by breeding one insertion at a time, utilization of the ability of founders to transmit multiple insertions to individual F_1 progeny enables screening of an average of 12 inserts per family, which allows the recovery of about one insertional mutation per seven families screened (Amsterdam *et al*., 1999). This is only 7–10-fold lower than the frequency observed in analogously performed (3 generation diploid) ENU screens (Driever *et al*., 1996; Haffter *et al*., 1996; Mullins *et al*., 1994; Solnica-Krezel *et al*., 1994). The strategy to produce, select, and breed the fish for such an insertional mutagenesis screen is outlined below.

A. Making Founder Fish that Transmit Proviral Inserts at High Frequency to their F_1 Progeny

Founders are produced by injecting late blastula stage (512–2000 cells) embryos. Virus must be injected into the space between the cells, and blastula stage embryos

ideally accommodate the injected fluid. At this time, there are four primordial germ cells, and these cells will divide two or three times over the course of the infection (Yoon *et al*., 1997). Thus, the injected embryos grow up to be founder fish (F0) with mosaic germlines. With very good viral stocks, individual founders can contain 25–30 different insertions in their germlines, with any given insertion present in 3–20% of the gametes (Chen *et al*., 2002). Individual F_1 fish can inherit up to ten different insertions, and, when founders are bred to each other, F_1 fish can be found with up to 20 different insertions. F_1 fish are not mosaic and will transmit all of their insertions in a Mendelian fashion.

Because the efficiency of the screen relies upon the generation of F_1 fish with a high number of inserts, it is essential to perform quality control assays on the viral stocks and founder injections before raising and breeding the founders. For every batch of injected embryos, several embryos are sacrificed for DNA preparation at 48 h post-injection and subjected to quantitative Southern analysis or real-time polymerase chain reaction (PCR) to determine the average number of insertions per cell in the entire infected embryo. This number is called the embryo assay value (EAV) (Amsterdam *et al*., 1999). In our experience, if the average EAV is above 15 and does not vary much amongst the individual analyzed embryos, the rest of the founders from that injection session transmit inserts at the rates mentioned above. Batches with average EAV below 15 transmit somewhat fewer inserts, and usually have greater founder-to-founder variation, and those with average EAV below 5 are quite inconsistent in transmitting multiple inserts to their progeny.

B. Breeding and Screening for Mutations

The breeding scheme for a diploid F_3 insertional mutagenesis screen is outlined in Fig. 1. In essence, the goal is to create families with a large number of independent insertions which can be screened simultaneously. This is achieved by selecting and breeding F_1 fish that inherit the most inserts from the mosaic founders.

Founder fish can be bred to each other or outcrossed to non-transgenic fish. For reasons that remain unclear, a majority of injected fish grow up to be males. Thus, it is most efficient to outcross the best male F_0 fish (those from batches with the highest EAV) and inbreed the rest. F_1 families of 30 fish are raised, and at six weeks of age the fish are fin-clipped for DNA preparation and analysis by quantitative Southern blotting or real-time PCR to determine which fish harbor the most insertions. Keeping up to the three top fish per family with at least five inserts strikes a balance between throwing away too many inserts (if fewer fish were kept) and keeping too many "repeat inserts" (i.e., the same insert inherited by sibling F_1 fish). The repeat insert rate is quite low if only three fish are kept, as the average mosaicism (i.e., proportion of F_1 inheriting a given insert) is about 8%. In the MIT screen (Golling *et al*., 2002), only 3% of the recovered mutations were caused by re-isolating such "repeat inserts."

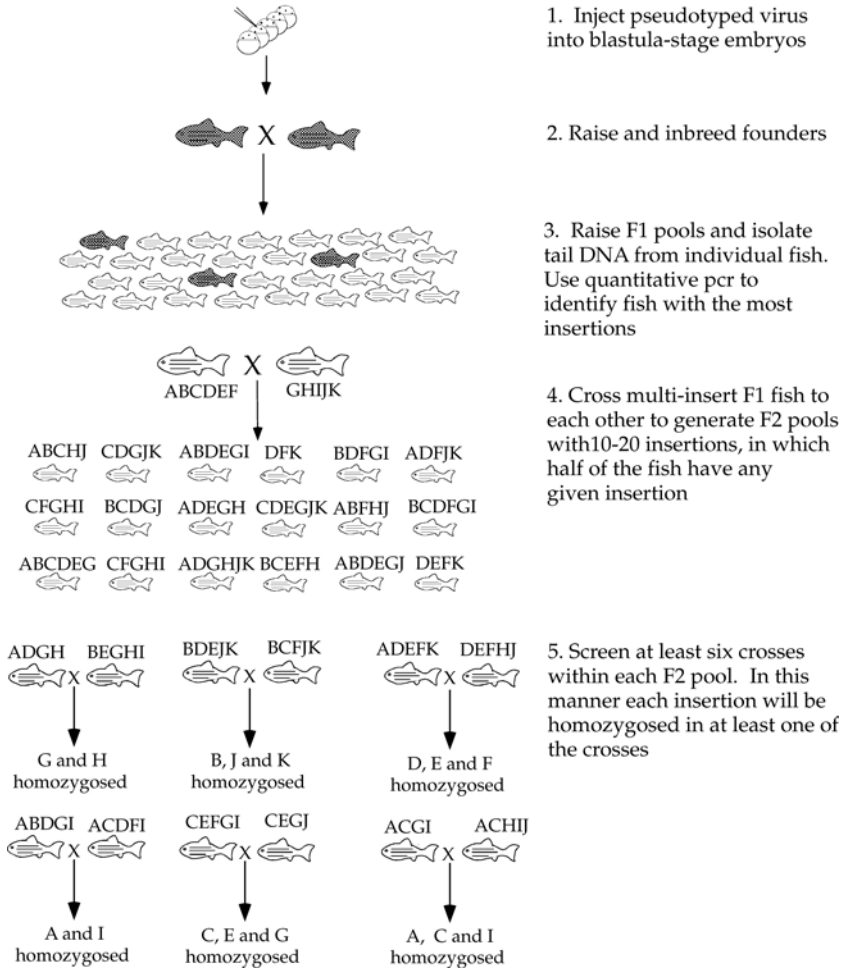


Fig. 1 Insertional mutagenesis breeding scheme.

The selected multi-insert F_1 fish are pooled together, and eventually bred to make F_2 families that harbor at least ten different independently segregating inserts, and in which each insert is present in half of the fish. Multiple sib-crosses are then conducted between the F_2 fish; since half the fish have any given insert, including those causing a mutation, each insert should be homozygosed in one-quarter of the crosses. On average, six crosses will homozygose 83% of the inserts in the family, and ten crosses will screen 95% of them. Every F_3 clutch from each F_2 family is screened for a phenotype in one-quarter of the embryos. In the MIT screen, embryos were scored for any morphological defect visible in a dissecting microscope at 1, 2, and 5 days post fertilization (dpf), as well as for defects in motility and touch response. One aid

to screening is that over 98% of these mutants fail to inflate their swim bladders by 5 dpf; since this is such a highly visible structure, a quick screen for clutches in which one-quarter of the embryos fail to inflate their swim bladder often signals the presence of a mutation.

III. Cloning the Mutated Genes

A. Identifying the Mutagenic Insert

The great advantage to using insertional mutagenesis over chemical mutagenesis is that the mutagenic insertion provides a molecular tag that can be used to identify the disrupted gene. However, because the mutagenesis screen described above utilizes multiple insertions to increase the rate of recovery of mutations, the first step after identifying a mutation is to determine which (if any) insertion appears to be responsible for the mutation. DNA is prepared from the tails of the parents of all of the crosses from the F_2 family and, using Southern analysis to distinguish the different insertions, one looks for an insertion that segregates with the phenotype (Fig. 2A). A linked insert (represented by a band of a specific size)

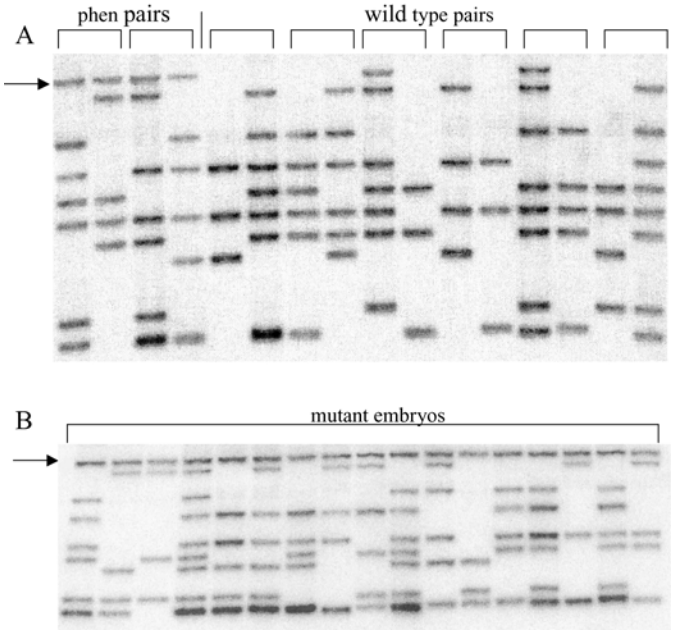


Fig. 2 Identification of the mutagenic insert. A. Southern analysis of DNA prepared from tail fins of F_2 fish; the arrow indicates an insert that is homozygosed in phenotypic pairs but not any of the wild-type pairs. B. Southern analysis of DNA from individual mutant embryos from the second phenotypic pair in 2A; the arrow indicates that the same insert is also present in all of the mutant embryos.

will be shared by both parents of every cross that had the phenotype and be in only one or neither of the parents of every cross which lacked the phenotype. Additionally, DNA prepared from the mutant embryos is subjected to the same analysis; while an unlinked insert would be in only three-fourths of the mutant embryos, a linked insert must be present in all of them (Fig. 2B). A similar analysis can be done with linker-mediated PCR, although one runs the risk that not every insertion will be successfully amplified and therefore a few inserts may remain "invisible."

It is possible that no insert will appear linked to the phenotype; in the MIT screen, we found that about one-quarter of the mutants recovered were not linked to a detectable insertion. Additionally, it is important to note that the identification of an insertion initially linked to the phenotype is not proof that the identified insert is *tightly* linked to the mutation; it is merely a way to either identify the insertion which is a *candidate* for causing the mutation or to conclude that the mutation is not linked to any insert if no insert meets the above criteria. This is because recombination rates in the male germline are much lower than in the female germline (Singer *et al*., 2002); thus an insert inherited from an F_1 male which is merely on the same chromosome as a non-insertional mutation will often meet the above criteria. The mutation and the insert will not have segregated in the F_2 generation, and because the mutant F_3 embryos must inherit the mutant locus from both parents, even if there is recombination in the female germline, all of the mutant embryos will receive the insert from their father. Thus, additional linkage experiments that can distinguish heterozygosity from homozygosity for the insert are required, but it is not possible to perform these until genomic DNA flanking the candidate insert is cloned.

Sometimes more than one insert meets the above criteria, and thus more than one is a candidate to have caused the mutant phenotype. This can be for one of several reasons. First, if more than one insert in the family is on the same chromosome, for the reasons described above, they may fail to segregate from each other. Often this can be resolved by outcrossing a female carrier and repeating the analysis in the next generation, either by further random sib-crosses followed by molecular analysis, or by using Southern analysis first to identify fish with one or the other insertion and then performing test crosses. Another possibility is that multiple copies of the virus have integrated in tandem, which happens about 3–4% of the time. Usually when this happens, there is a higher-intensity provirus-sized band (if the enzyme used cuts the insert only once) in addition to the band truly representing the junction between the provirus and endogenous fish DNA (junction fragment) Finally, there may be too many inserts in the family to accurately distinguish all of the inserts (greater than 15–20), and this may complicate the analysis. In these cases a female carrier fish, preferably already shown to have relatively few inserts, must be outcrossed and the analysis repeated in the next generation. It is essential not to focus on a single insertion as the cause of a mutation unless it is very clear that no other insertion could also be linked to the mutant phenotype.

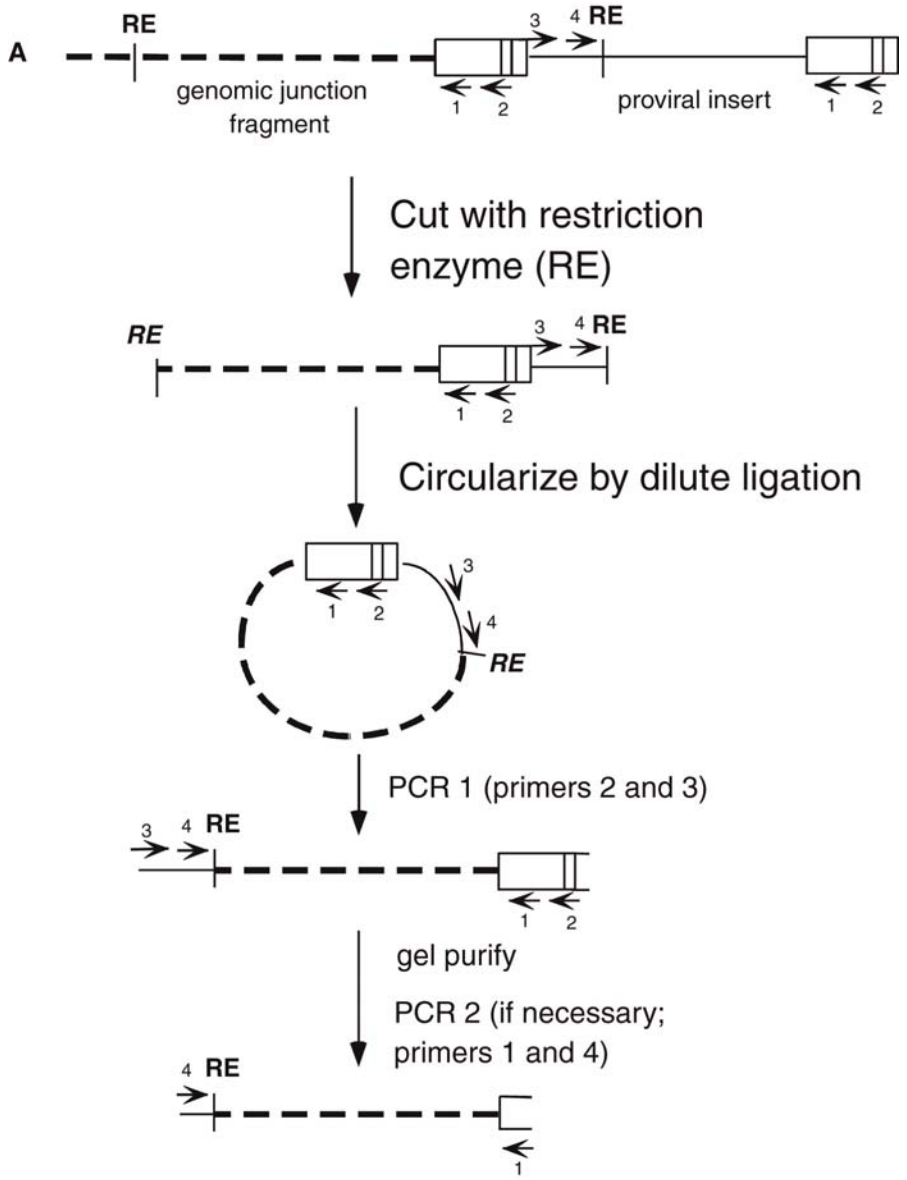


Fig. 3 Inverse PCR. A. Schematic of the inverse PCR process. B. Selection of the correct enzyme to use. For an insert already identified as in Fig. 2, DNA from several tails already known to be positive or negative for that insert is analyzed by Southern blotting after digestion with different restriction enzymes. In this example, while the junction fragments with *Bgl II* and *Nco I* are too big to successfully amplify by inverse PCR, the *Taq I* junction fragment should amplify easily. Arrows indicate the linked insertion.

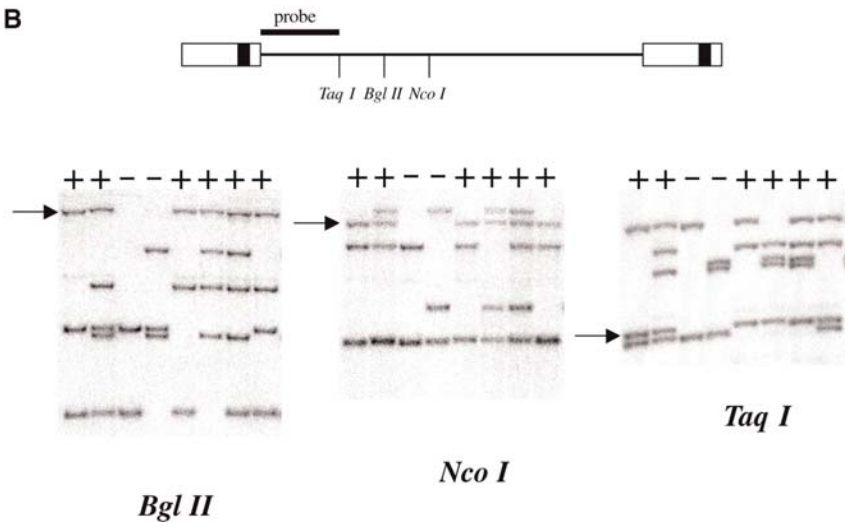


Fig. 3 *(Cont.)*

B. Cloning the Flanking Genomic DNA

After identifying the candidate mutagenic insertion, inverse PCR or linker-mediated PCR can be used to clone genomic DNA flanking one or both sides of the mutagenic provirus (Fig. 3a). Since all of the inserts have the same sequence, in order to clone the correct junction fragment, one must know the size expected for a given enzyme used. Often it is necessary to analyze the DNA samples by Southern analysis with several enzymes in order to identify which enzyme will be best for obtaining the desired insert (Fig. 3b). After cloning and sequencing the putative junction fragment, one can design a PCR primer in the sequence that points back at the provirus, and use PCR with this and a viral primer on DNA isolated from tail fin samples of fish known to be carriers or non-carriers to confirm that the correct junction was cloned.

After cloning the genomic DNA flanking one side of the virus, it is necessary (for reasons explained below) to obtain sequence on the other side of the insertion as well; it may also be desirable to obtain additional sequence extending further from the virus on the side originally cloned. One way to do this is to use the cloned sequence as an anchor for additional inverse PCR, or linker-mediated PCR. However, as the zebrafish genome assembly becomes increasingly complete, this step is becoming increasingly dispensable; often even a small amount of sequence adjacent to the virus is sufficient to place the insertion site on a large contig of known sequence.

One of the uses of the cloned sequence is that it allows one to perform an assay to distinguish transgenic and non-transgenic chromosomes in a co-dominant fashion. Such an assay is essential in order to demonstrate that the insertion is tightly linked to the mutation, and thus most likely is its cause. One method is to use the junction

fragment as a probe on a Southern blot, as the transgenic and non-transgenic chromosomes will each produce hybridizing fragments of a different size (Fig. 4a). Alternatively, PCR can be conducted with three primers, one on each side of the insert and one pointing out of the insert, such that different sized products will be amplified by insert-bearing and non-insert-bearing chromosomes (Fig. 4b). In either case, the assay is used to demonstrate that mutant embryos are invariably homozygous for the insertion, while wild-type embryos never are. Every mutant analyzed is the equivalent of observing one meiotic event (only counting the female germline); every wildtype analyzed is the equivalent of observing 1/3 of a meiosis. (In the wild-type case, only one in three recombination events between a mutation and a marker in a dihybrid cross will lead to a wild-type embryo which is homozygous for the marker; thus scoring for wildtypes which are homozygous for a marker only detects one third of the recombination events between these loci.). If any recombinants are observed between the mutation and the insertion, the insertion cannot be the cause of the

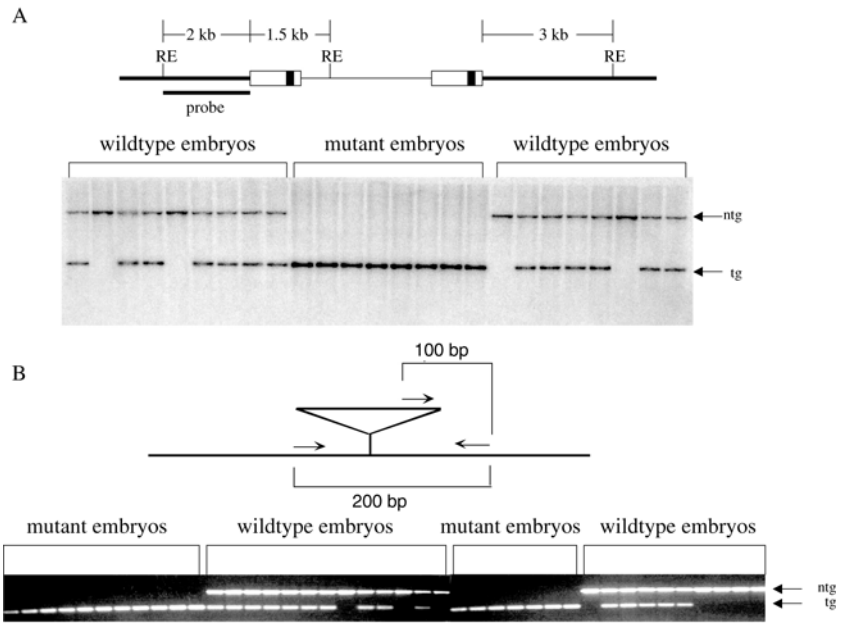


Fig. 4 Tight linkage assays. In either assay, mutant embryos should always be homozygous for the insert, while wild-type embryos never should be. A. Southern analysis of DNA prepared from individual wild-type or mutant embryos. The sequence of the junction fragment on one side of the virus is used as the probe. In this example, insert-bearing chromosomes (tg) will give a 3.5 kb band, while non-insert chromosomes (ntg) will give a 5 kb band. Thus each embryo can be genotyped as homozygous for the insertion (smaller band only), heterozygous (both bands), or homozygous non-insertion (larger band only). B. PCR analysis of DNA prepared from individual wild-type or mutant embryos. The PCR reaction is run with three primers, such that (as with the Southern method) the presence of either chromosome is indicated by a unique sized band.

mutation. We typically analyze 50–100 meioses in this fashion; while not absolute proof that the insert is the cause of the mutation, given the size of the genome, the relative rates of spontaneous and insertional mutations, and the average number of inserts in each family, fewer than 0.5% of mutations which meet this criteria will have a cause other than the insertion. Establishing tight linkage with more observed meioses can linearly decrease the likelihood that the mutation is not caused by the insertion, but linkage alone cannot reduce this likelihood to zero.

One exception to the requirement for absolute linkage is in cases where there is incomplete penetrance of the phenotype; thus by definition the phenotype and genotype do not always match. This is evident when consistently fewer than 25% of the embryos are phenotypic. In these cases, while all of the mutant embryos still must be homozygous for the insert, some of the phenotypically wild-type embryos will also be homozygous (Amsterdam *et al*., 1999; Golling *et al*., 2002).

C. Gene Identification

While the sequence of the junction fragment is useful for allowing genotypic identification of carriers and is required for the tight linkage experiments described above, its greatest utility of course is in the ability to identify the mutated gene. Given up to a few kilobases of sequence on either side of the insertion, in over 80% of the cases exon sequence can be found by BLAST analysis (Altschul *et al*., 1997) based upon either nucleotide identity to a zebrafish cDNA or expressed sequence tag (EST), or upon amino acid homology to known or predicted proteins from other organisms when translated. RT-PCR or 3′ and 5′ RACE can then be used to complete the cDNA if necessary. As the zebrafish genome assembly and annotation becomes more complete, merely BLASTing to the genome will be sufficient to identify the gene into which the virus has inserted.

It is important that the virus actually is in, as opposed to just near, the gene identified, or it is possible to identify the wrong gene. Zebrafish genes are sometimes very near each other, with as few as half a kilobase separating them. Thus, an insert could be less than one kilobase from an easily recognized or annotated gene, but actually disrupt another gene that was not found in the BLAST search. Zebrafish genes often have first exons that are entirely 5′ untranslated or include the coding sequence for only a few amino acids, and such initial exons could easily be missed by a BLAST search of the genomic DNA sequence. In the case of several of the insertional mutants, the gene originally recognized in the flanking sequence either began or ended about a kilobase from the provirus; only upon more careful inspection was it found that in fact another gene began between the originally identified gene and the virus. Analysis of mRNA expression in wildtype and mutant embryos in two of these cases demonstrated that only the proximal gene's expression was affected by the insertion. On the other hand, in the case of several insertional mutants, the insert is outside the affected gene, presumably in the promoter region, and does affect the transcription of the gene. Thus, the finding that the mutagenic

insert lies outside of a gene does not necessarily mean that the nearest identified gene is not the gene of interest, but rather that further analysis is required.

Cloning a gene that is proximal to a mutagenic insertion is not absolute proof that the correct gene has been identified. First, it is always possible, if very unlikely, that there is a non-insertional mutation very near the insert. Second, it is possible that the expression of a neighboring gene has been affected as well. We have not observed this yet in any of the recessive mutants that we have studied, but we have found this to be the case in at least one dominant mutant (Amsterdam *et al*., 2009). Similarly, even when a provirus lies between two exons of a gene and disrupts its expression, it is still possible that another gene (e.g., one lying within an intron of the first gene) might be the gene responsible for the mutant. Thus, to be absolutely certain that mutation of the identified gene is responsible for the phenotype, the gene identity needs to be confirmed by independent means. The finding of a second insertional allele, or non-complementation to a chemical allele with a demonstrated point mutation, would make the likelihood of a nearby non-insertional mutation exceedingly low. The ideal proof is rescue: Does reintroduction of the gene in trans into mutant embryos rescue the phenotype? Rescue is not easily accomplished in stable transgenics but can be done transiently for some mutations. Alternatively, phenocopy of the phenotype by morpholino injection (Nasevicius and Ekker, 2000) can independently verify that mutation of this gene leads to this phenotype.

D. Phenotypic Consequence of Insertions

The insertion of a provirus into a gene can affect that gene in a number of different ways. Unlike chemical mutagenesis, which by causing point mutations has the potential to create hypomorphic or neomorphic alleles by amino acid substitution, insertional mutagenesis generally works by more broadly knocking down or out gene expression, although there are exceptions. While only about one third of the mutagenic inserts actually interrupt exons, which would obviously impair gene expression, nearly half land in the first intron and the rest in downstream introns (Fig. 5a). We have used Northern analysis and/or quantitative RT-PCR to analyze gene expression in mutants in many of these cases, and the most common effect that we have seen is a reduction or elimination of mRNA, anywhere from fivefold to undetectable levels (Golling *et al*., 2002). In a study of randomly selected insertions (i.e., isolated independent of causing a phenotype), most insertions in the first intron of genes were found to similarly reduce steady-state mRNA levels, while insertions in later introns were less likely to have any effect (Wang *et al*., 2007). Thus, while some insertional mutants are nulls, some are hypomorphs, as expression has not been completely abrogated.

Additionally, some insertional mutants can affect the nature of the message (and the protein it produces) rather than merely the level of expression. First, some mutations cause exon skips instead of down-regulation of expression. For example, we have three insertional alleles of the vHNF1 gene, two in the first or second exon,

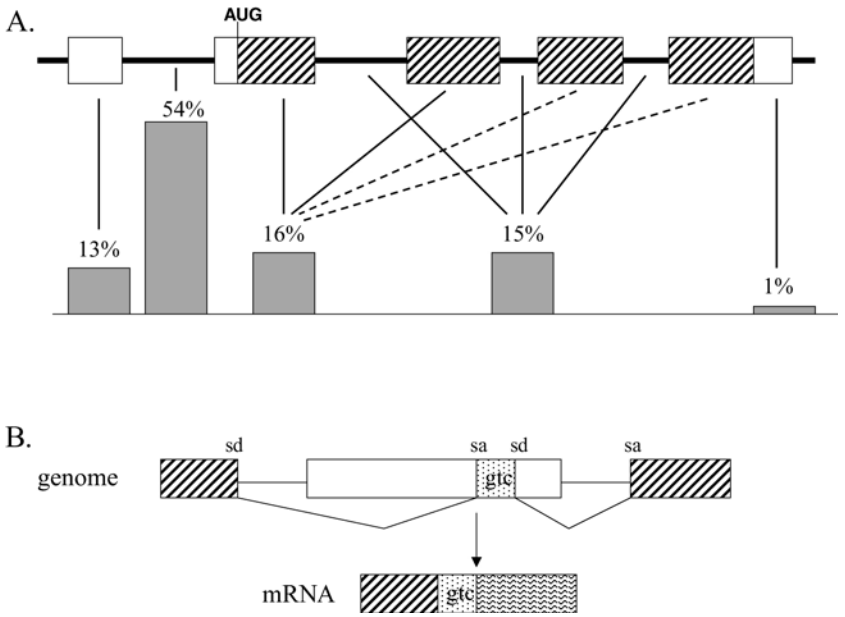


Fig. 5 Mutagenic insertion sites. A: distribution of mutagenic insertions from 413 insertional mutants in 298 different genes. White boxes indicate 5′ and 3′ untranslated regions (UTR), striped boxes indicate coding exons, lines between boxes indicate introns. While the sample gene here has an untranslated first intron (the first white box), in about half the cases the first exon contains the initiation codon. The percentage of mutagenic insertions which lie in the 5′ UTR or promoter, the first intron, coding exons, downstream introns, or the 3′ UTR, is shown below the gene. B: consequence of gene-trap event. If the insertion is in an intron in the correct orientation, the splice donor from the previous exon can splice to the gene-trap cassette, and then splice out to the next endogenous exon, thus creating a frameshift. Striped boxes, exons; large white box, provirus; stipled box (gtc), gene-trap cassette; sd, splice donor; sa, splice acceptor; wavy-lined box, exon with frameshift mutation.

which appear to be nulls, and one in the fifth intron, which leads to two different splice variants, skipping either the fourth or third and fourth exons (Sun and Hopkins, 2001). This allele is predicted to make a truncated protein, and in fact has a less severe phenotype. Second, the virus used to make most of the insertional mutations contains a splice-in, splice-out, frameshift-producing gene-trap cassette (Chen *et al*., 2002). When the virus inserts in an intron in the correct orientation, it is possible for the preceding splice donor to splice to this exon in the provirus, and splice out to the next endogenous exon, thus creating a frameshift and presumably a truncated protein (Fig. 5b). If this happens, analysis of mRNA by RT-PCR will show the presence of an increased-sized band, indicating the inclusion of the trapped exon. This is likely to lead to a truncated protein, and thus could act as a hypomorph or neomorph. While we have not found this to be a common mechanism of mutation, it has occurred in some of the mutants.

IV. Retroviral-mediated Insertional Mutagenesis as a Reverse Genetic Tool

A. Demonstrating the Utility of Retroviral Mutagenesis to Reverse Genetics Approaches

As described earlier, retroviruses have been used in forward genetic screens to isolate mutations affecting embryonic development (Amsterdam *et al*., 1999). Forward genetics is a process where mutational studies are done to determine the genetic basis of associated phenotypes. Forward genetic screens are effective in identifying mutations with visible phenotypes, but are limited by issues of gene redundancy or the ability to detect phenotypes blindly. With the recent advancement in sequencing technologies and availability of high-quality whole genomes, researchers have access to rich gene information in many model organisms and it becomes possible to systematically screen by first identifying mutations in genes, and then testing homozygous mutant animals for phenotypes, i.e. "reverse" genetics.

The genes can be disrupted either in a targeted manner (homologous recombination or RNAi-mediated knockdown) or by using non-targeted approaches (insertional mutagenesis or chemical-induced mutagenesis for example).

The most widely used reverse genetic approach in zebrafish is Targeted Induced Local Lesion IN Genomes (TILLING), which has been used to identify a desired mutation from a pool of ENU-induced mutagenized fish either through CEL1 nuclease assay or DNA sequencing (Stemple, 2004; Wienholds *et al*., 2002). Although effective, this approach is expensive and labor-intensive.

It has been shown that mutagenesis using MoMLV as an insertional mutagen followed by integration mapping and recovery is possible, and thus could be an alternative reverse genetics approach (Wang *et al*., 2007). A public effort is underway in the Burgess lab in collaboration with Dr. Shou Lin's lab at UCLA to systematically disrupt every protein-coding gene in zebrafish genome using pseudotyped MoMLV. In this approach we infect zebrafish embryos with MoMLV, and proviral integrations in the F_1 generation of infected fish are mapped by PCR amplification and sequencing (Fig. 6). Using this strategy it is possible to identify a mutagenic event for every 30 sequencing reactions done on the F_1 fish, thus showing up to 30-fold increase in efficiency compared to TILLING (Wang *et al*., 2007). The gene disruption pipeline begins with the production of high-titer pseudotyped MoMLV retroviruses using an improved method previously described in Jao and Burgess (2009), which is then injected into blastula stage embryos (as described earlier) to produce founder fish. We raise the founders and outcross them with wild-type fish to obtain the F_1 generation. We raise only the male F_1 fish, sacrifice them to cryopreserve their testes, and collect corresponding tail cuts to isolate genomic DNA (gDNA) for integration mapping. In order to map retroviral integrations, gDNA is subjected to restriction digestion and then linkers are ligated to the digested fragments. We use a linker-specific and proviral-LTR (long terminal repeat)-specific primer to amplify proviral-gDNA-LTR fragments. The amplified PCR products are

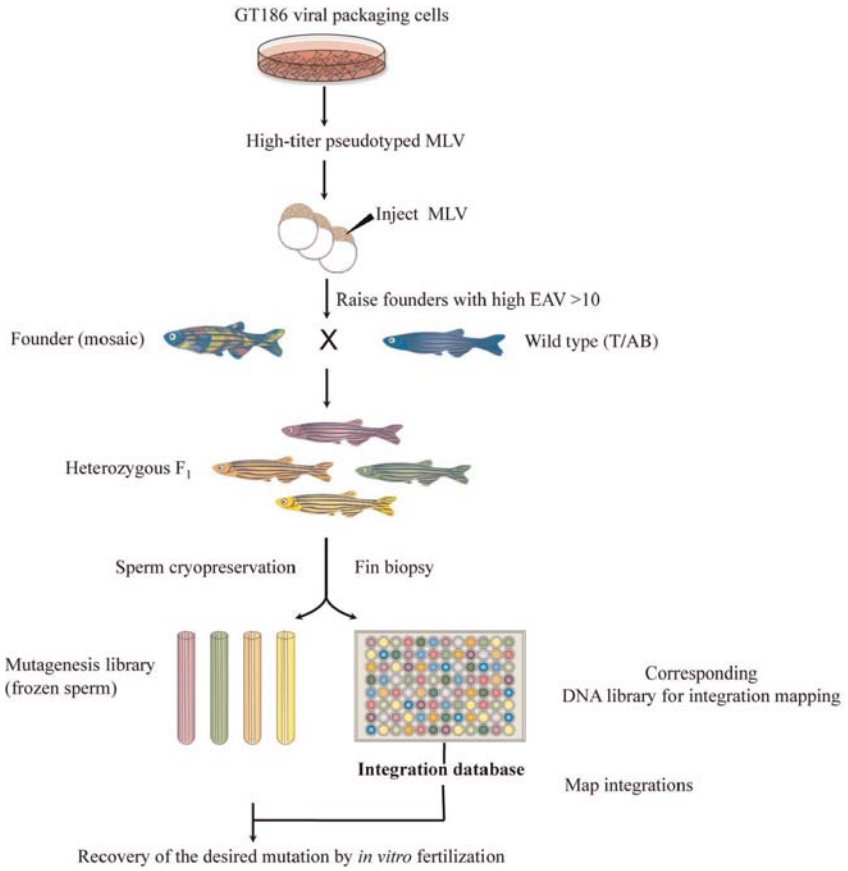


Fig. 6 Retroviral mapping pipeline. Blastula stage embryos are injected with the pseudotyped MoMLV. Infection rate is determined by quantitative PCR (qPCR) and founder fish with high infection rate are raised. Founders are crossed to wild-type fish and male F_1 fish are raised. Sperm samples are cryopreserved from the F_1 fish, integrations are mapped in the corresponding DNA library, integrations are assigned to the corresponding sperm samples. Desired mutations are recovered by *in vitro* fertilization using the corresponding frozen sperm.

sub-cloned and to date have been sequenced by Sanger sequencing. The sequences are mapped back to zebrafish whole genome assembly to get integration information. Our goal is to generate a genome-wide knockout library and mutagenize to saturation in the next 5 years.

Our lab originally investigated the distribution profile of proviral integration sites in the zebrafish genome by sequencing integrations from 900 F_1 fish. We obtained 933 unique sequences; of these, 599 integration sites were mapped to the zebrafish genome. Sixty-five percent of mapped integrations either landed in annotated Ensembl genes, or within 3 kb of genes, with a strong preference for the first intron

of genes (Wang *et al*., 2007). The preference of MoMLV integration for the first intron is consistent with the results in mouse (Mooslehner *et al*., 1990; Scherdin *et al*., 1990) and in human tissue culture cells (Wu *et al*., 2003). We showed that not only are integrations in the exons mutagenic but also integrations landing in the first intron of genes are also highly mutagenic (Golling *et al*., 2002; Wang *et al*., 2007). The gene expression level in 80% of the first-intron hits reduced mRNA level to $<$10% of wild-type level. Our prediction is that one in five retroviral integrations will result in a gene disruption that reduces the gene expression level to 10% or less of the wild-type level.

B. Adapting Retroviral Mapping Strategies to Second Generation Sequencing Technologies

The advantage of retroviral mutagenesis over ENU is that mutated genes are molecularly tagged by proviral sequences. By identifying genomic sequences flanking proviral integrations, these integrations can be identified in a relatively short time compared to ENU-mediated mutagenesis (Gaiano *et al*., 1996b; Golling *et al*., 2002). Our lab has established a more efficient method of identifying these integrations. The identification of the MoMLV integration by linker-mediated-PCR (LM-PCR) starts by fragmenting gDNA using a frequent cutter restriction enzyme, followed by ligating linkers and linker-mediated PCR, shotgun cloning, and Sanger sequencing. However, this method has limitations; we identify fewer integrations than are predicted based on qPCR data.

Limited recovery is not an issue for phenotype-based screens, as only specific integrations that are linked with the phenotype need to be recovered. But in order to implement the reverse genetic strategy to generate a genome-wide knockout library, it is imperative to recover as many integrations as possible. The published method (Wang *et al*., 2007) for identifying integrations is labor-intensive, expensive, and low-throughput because it requires significant sub-cloning and depends on Sanger sequencing. To overcome these limitations and improve the mapping process, we have modified the strategy by adapting it to second generation sequencing technologies (Fig. 7). The modified method uses three pairs of enzymes (*Mse1* and *Pst1*), (*Bfa1* and *BanII*), and (*Csp6I* and *BanII*). *PstI* or *BanII* is used to suppress the amplification of an internal band that would occur from the 5′ LTR (which has identical sequence to the 3′ LTR). *Mse1*, *Csp6I*, and *Bfa1* recognize 4-base TTAA, CTAG, and GTAC sites in the genome, respectively, and three enzymes generate the same 5′-TA overhang in the digested DNA. The linkers, consisting of a short and long strand, are ligated onto the overhanging ends of the gDNA. When these strands are annealed, the short strand contains non-complimentary sequences that will form a hairpin. This design prevents unwanted linker-to-linker amplification. The long strand of the linker does not contain a binding site for the linker PCR primer; instead it contains an exact match to the linker primer sequence and will only work after its complimentary sequence is synthesized. As a result, only gDNA containing LTR-gDNA-linker sequences will be amplified. In order to make the mapping pipeline

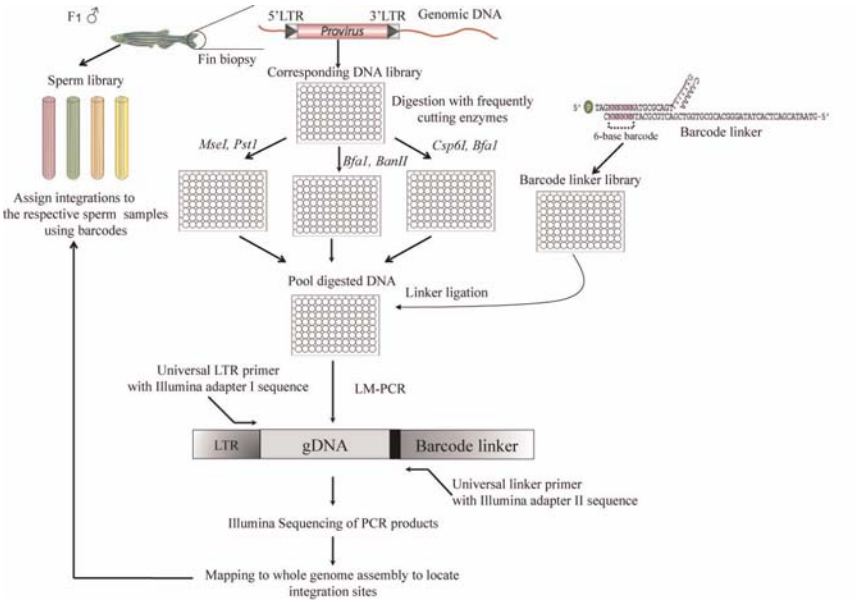


Fig. 7 High-throughput strategy to identify retroviral integrations using next-generation sequencing. DNA samples corresponding to individual F_1 fish are separately digested with three sets of restriction enzymes to fragment genomic DNA. The digested samples are then ligated with a DNA linker containing a unique 6-bp barcode to index each F_1 fish. The linker-ligated DNA samples are amplified by LM-PCR using LTR and linker-specific primers to amplify the adjacent genomic DNA sequences. The 5′ ends of the LTR and linker primers also contain adapter sequences required to bind to the Illumina flow cell for sequencing. Illumina sequencing can generate up to 150 bp of mappable sequences using the paired-end method. These paired-end sequences are aligned to the latest assembly of the zebrafish genome and integrations are assigned to corresponding sperm samples using the index of unique barcodes.

cost-effective and high-throughput, we have replaced the multi-stepped shotgun cloning and Sanger sequencing process with second generation sequencing technology. There are several commercial next-generation sequencing platforms available, each with some advantages and disadvantages. We utilize the Illumina Genome Analyzer (GAxII) platform because of high sequence read numbers, relatively low costs, and simplified template preparation. By using the GAxII, we overcome the problem of limited sampling as hundreds of samples are pooled together in a single lane of sequencing and each sample (i.e., fish) will still have tens of thousands of sequence reads each. However, pooled samples pose a challenge as how to assign the mapped integrations to their respective fish. We developed a 6-base molecular barcode strategy and embedded 960 different barcode sequences into linkers; thus the digested DNA from each F_1 fish can be coded individually. This barcoding strategy allows us to minimize the sample handling and reagent consumption in LM-PCR reactions as barcoded samples can be pooled together, bringing mapping

costs down. We added Illumina adapter sequences to the LTR and linker-specific primers and these adapter sequences are used to bind the amplicons to the GAxII flow cells. The LM-PCR products are sequenced from both sides using "paired-end" sequencing. We obtain 24 bp of LTR sequence and 76 bp of genomic sequence on one side, and 24 bp of linker, the 6 bp barcode, and 70 bp of genomic sequence from the opposite side. It is therefore possible to get up to 150 bp of genomic sequence for mapping the retroviral integration, although the majority of the reads in the sequencing are shorter (∼60–70 bp). The sequence reads are filtered to obtain only perfect LTR and linker sequences, and the resulting subset of reads are aligned using the short-read aligner program Bowtie (Langmead *et al*., 2009) and mapped to the latest assembly of the zebrafish genome.

We have compared our traditional Sanger sequencing method to the barcoded Illumina method. We recovered ∼57% more mappable integrations using the latter. The majority of this increase in recovery can be explained by the use of the three sets of restriction enzymes and the exhaustive sequence sampling. To date, we have mapped more than 6000 integrations from 2000 fish in a very high-throughput and cost-effective manner.

V. Future Directions

The current insertional mutant collection from the Hopkins lab includes over 500 mutants, with mutations in nearly 400 genes; so far we have identified 340 of these genes. We believe that this represents approximately 25% of the genes which can be mutated to a visible (and usually lethal) phenotype in the fish (Amsterdam *et al*., 2004a). While the characterization of the phenotypes is somewhat rudimentary at present, numerous "shelf screens" of this collection have been conducted, including staining with various antibodies, *in situ* hybridization markers, and other reagents that illuminate the patterning and development of specific tissues, as well as the assessment of various physiological functions. Thus a substantial portion of the genes required for the proper formation of all of these structures will be identified.

The existing collection can also be used to monitor the long-term effects of heterozygosity of these genes in adults, as mutations that have recessive embryonic phenotypes might predispose adults to disease in the heterozygous state. The genes mutated in many of the mutants are known to be autosomal dominant disease genes in humans. For example, mutations in the zebrafish vHNF1 gene affect both kidney and pancreas development (Sun and Hopkins, 2001). vHNF1 mutations in the heterozygous state, while not yet investigated in fish, can lead to either kidney disease or diabetes in humans (Horikawa *et al*., 1997; Nishigori *et al*., 1998). Similarly, the *jellyfish* mutation, which results in defective cartilage differentiation and morphogenesis, is caused by disruption of the sox9a gene (Yan *et al*., 2002), while heterozygosity for sox9 mutations in humans leads to campomelic dysplasia (Foster *et al*., 1994). Thus, zebrafish heterozygous for insertional mutants could be screened for susceptibility to a variety of diseases. One example that illustrates the

utility of this strategy is the identification of insertional mutants with aging phenotypes; heterozygotes for mutations in either the *spinster* homolog or *telomerase-binding factor 2* have a shortened lifespan as well as some other markers of aging such as lipofuscin accumulation and premature retinal degeneration (Kishi *et al*., 2008).

Similarly, many genes whose mutation in the heterozygous state is known to predispose mammals to cancer (tumor suppressor genes) cause prenatal death in mice when homozygous (Jacks, 1996). An insertional mutation of the *NF2a* gene, an ortholog of a known human tumor suppressor gene (Ruttledge *et al*., 1994; Trofatter *et al*., 1993), is embryonic lethal in zebrafish. Heterozygosity for this *NF2a* mutation in zebrafish predisposes them to development of tumors of the nervous system (Amsterdam *et al*., 2004b), as it does in mammals. Among a large collection of recessive embryonic lethal mutations one might expect to find other mutations in which heterozygotes are more prone to develop cancer, and in fact, a number of the insertional mutations have increased rates of tumorigenesis as heterozygous adults (Amsterdam *et al*., 2004b; Lai *et al*., 2009).

There is also room for newer vectors or other methods of transgenesis to improve insertional mutagenesis, many of which are described in other chapters in this volume. Vectors employing a gene trap or enhancer trap with a visible marker, such as GFP, could preselect for insertions in genes with expression patterns of interest (Ellingsen *et al*., 2005; Kawakami *et al*., 2004). Such insertions could be selected in F_1 fish after passage through the germline, or gene-trap events could be selected *in vitro*, as is often done in ES cells in mice, and then cloned by nuclear transfer (Lee *et al*., 2002). It has also been suggested that vectors including transcriptional regulatory elements, such as the tetracycline-responsive promoter or the binary Gal4/UAS system, could create inducible gain-of-function mutations (Chen *et al*., 2002). One such screen utilized a virus containing multiple Gal4-responsive UAS sequences. Upon crossing fish harboring insertions of this virus to a transgenic line expressing Gal4-VP16 in a subset of neurons, genes at the insertion site of the virus were overexpressed in those neurons (Maddison *et al*., 2009). Thus, a single generation screen can identify genes whose overexpression in a particular cell type causes a phenotype, and this method can be repeated with many different Gal4-VP16 expressing lines. Furthermore, retroviruses need not be the only insertional tool. Several transposons have been shown to integrate into the zebrafish genome (Davidson *et al*., 2003; Kawakami *et al*., 2000a; Raz *et al*., 1998); if they could be mobilized in a controlled fashion similar to P elements in flies (Cooley *et al*., 1988; Spradling *et al*., 1999), they could prove to be a very effective mutagenic tool. Transposons have also been shown to be effective in delivering gene traps and enhancer traps, although the frequency with which these vectors mutate the gene they trap is still unclear (Balciunas *et al*., 2004; Kawakami *et al*., 2004; Parinov *et al*., 2004; Sivasubbu *et al*., 2006). Recent unpublished results suggest these vectors are becoming very promising as mutagens.

Finally, insertional technologies, be they retroviral or transposon based, could also be used for reverse genetics by generating a library of insertions, as has

become popular in other model organisms. Thousands of P-element fly lines have been cataloged with their junction sequences, and thus chromosomal location, now that the *Drosophila* genome is complete (FlyBase Consortium, 2002). Similarly, a consortium of several labs is producing thousands of murine ES cell clones and identifying the trapped genes by inverse PCR or RACE (Stanford *et al*., 2001; Wiles *et al*., 2000). Furthermore, hundreds of Sleeping Beauty insertions in live mouse lines are being mapped (Roberg-Perez *et al*., 2003). In *C. elegans*, hundreds of thousands of Tc1 insertion lines have been isolated and their genomic DNAs have been arrayed to provide a PCR-screenable panel for insertions in any gene of interest (Zwaal *et al*., 1993). As described above, we have taken an analogous approach in zebrafish, as a publically funded effort to create a library of mapped F_1s (Wang *et al*., 2007 and Section V). Screening such libraries provides an alternative to screening chemically mutagenized sperm libraries (Stemple, 2004; Wienholds *et al.*, 2002).

A powerful aspect of large mutant collections is the ability to systematically screen every gene in the zebrafish genome for phenotypes. Prior knowledge of the mutated gene enables researchers to use more "targeted" phenotype assessment. For example, genes expressed in the sensory epithelium of the ear can be tested for hearing defects in adults if no obvious phenotype is seen in the developing embryo. Testing of double mutations in duplicated genes would be possible as well. Finally, the strategies described here would make it possible to systematically track homozygous mutations for adult phenotypes, rather than being limited to testing haploinsufficiency phenotypes in adults.

Acknowledgments

We thank past and present members of the Hopkins, Lin, and Burgess laboratories for their contribution to the insertional mutagenesis screen and mutant library efforts. This research was supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (SB), and by Amgen Inc. and the NIH extramural program (AA).

References

Allende, M. L., Amsterdam, A., Becker, T., Kawakami, K., Gaiano, N., and Hopkins, N. (1996). Insertional mutagenesis in zebrafish identifies two novel genes, pescadillo and dead eye, essential for embryonic development. *Genes Dev.* **10**, 3141–3155.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M., and Hopkins, N. (1999). A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* **13**, 2713–2724.

Amsterdam, A., Lai, K., Komisarczuk, A. Z., Becker, T. S., Bronson, R. T., Hopkins, N., and Lees, J. A. (2009). Zebrafish Hagoromo mutants upregulate fgf8 post-embryonically and develop neuroblastoma. *Mol. Cancer Res.* **7**, 841–850.

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., and Hopkins, N. (2004a). Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. U. S. A.* **101**, 12792–12797.

Amsterdam, A., Sadler, K. C., Lai, K., Farrington, S., Bronson, R. T., Lees, J. A., and Hopkins, N. (2004 b). Many ribosomal protein genes are cancer genes in zebrafish. *PLoS Biol.* **2**, E139.

Balciunas, D., Davidson, A. E., Sivasubbu, S., Hermanson, S. B., Welle, Z., and Ekker, S. C. (2004). Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics* **5**, 62.

Becker, T. S., Burgess, S. M., Amsterdam, A. H., Allende, M. L., and Hopkins, N. (1998). Not really finished is crucial for development of the zebrafish outer retina and encodes a transcription factor highly homologous to human Nuclear Respiratory Factor-1 and avian Initiation Binding Repressor. *Development* **125**, 4369–4378.

Burns, J. C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J. K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc. Natl Acad. Sci. U. S. A.* **90**, 8033–8037.

Chen, W., Burgess, S., Golling, G., Amsterdam, A., and Hopkins, N. (2002). High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish. *J. Virol.* **76**, 2192–2198.

Cooley, L., Kelley, R., and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P elements. *Science* **239**, 1121–1128.

Culp, P., Nusslein-Volhard, C., and Hopkins, N. (1991). High-frequency germ-line transmission of plasmid DNA sequences injected into fertilized zebrafish eggs. *Proc. Natl Acad. Sci. U. S. A.* **88**, 7953–7957.

Davidson, A. E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., Sivasubbu, S., Cliff, M. P., Hackett, P. B., and Ekker, S. C. (2003). Efficient gene delivery and gene expression in zebrafish using the Sleeping Beauty transposon. *Dev. Biol.* **263**, 191–202.

Driever, W., *et al*. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46.

Ellingsen, S., Laplante, M. A., Konig, M., Kikuta, H., Furmanek, T., Hoivik, E. A., and Becker, T. S. (2005). Large-scale enhancer detection in the zebrafish genome. *Development* **132**, 3799–3811.

FlyBase Consortium(2002). The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Res.* **30**, 106–108.

Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., Kwok, C., Weller, P. A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I. D., Goodfellow, P. N., Brook, J. D., and Schafer, A. J. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**, 525–530.

Gaiano, N., Allende, M., Amsterdam, A., Kawakami, K., and Hopkins, N. (1996a). Highly efficient germ-line transmission of proviral insertions in zebrafish. *Proc. Natl Acad. Sci. U. S. A.* **93**, 7777–7782.

Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T., and Hopkins, N. (1996b). Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383**, 829–832.

Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado, E., Chen, W., Burgess, S., Haldi, M., Artzt, K., Farrington, S., Lin, S. -Y., Nissen, R. M., and Hopkins, N. (2002). Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat. Genet.* **31**, 135–140.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J. M., Jiang, Y. -J., Heisenberg, C. -P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., and Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development* **123**, 1–36.

Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B. N., Lindner, T., Yamagata, K., Ogata, M., Tomonaga, O., Kuroki, H., Kasahara, T., Iwamoto, Y., and Bell, G. I. (1997). Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat. Genet.* **17**, 384–385.

Ivics, Z., Izsvák, Z., and Hackett, P. B. (1993). Enhanced incorporation of transgenic DNA into zebrafish chromosomes by a retroviral integration protein. *Mol. Mar. Biol. Biotech.* **2**, 162–173.

Jacks, T. (1996). Tumor suppressor gene mutations in mice. *Annu. Rev. Genet.* **30**, 603–636.

Jao, L. E., and Burgess, S. M. (2009). Production of pseudotyped retrovirus and the generation of proviral transgenic zebrafish. *Methods Mol. Biol.* **546**, 13–30.

Kawakami, K., Amsterdam, A., Shimoda, N., Becker, T., Mugg, J., Shima, A., and Hopkins, N. (2000a). Proviral insertions in the zebrafish hagoromo gene, encoding an F-box/WD40-repeat protein, cause stripe pattern anomalies. *Curr. Biol.* **10**, 463–466.

Kawakami, K., Shima, A., and Kawakami, N. (2000b). Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. U. S. A.* **97**, 11403–11408.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144.

Kishi, S., Bayliss, P. E., Uchiyama, J., Koshimizu, E., Qi, J., Nanjappa, P., Imamura, S., Islam, A., Neuberg, D., Amsterdam, A., and Roberts, T. M. (2008). The identification of zebrafish mutants showing alterations in senescence-associated biomarkers. *PLoS Genet.* **4**, e1000152.

Lai, K., Amsterdam, A., Farrington, S., Bronson, R. T., Hopkins, N., and Lees, J. A. (2009). Many ribosomal protein mutations are associated with growth impairment and tumor predisposition in zebrafish. *Dev. Dyn.* **238**, 76–85.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Lee, K. Y., Huang, H., Ju, B., Yang, Z., and Lin, S. (2002). Cloned zebrafish by nuclear transfer from long-term-cultured cells. *Nat. Biotechnol.* **20**, 795–799.

Lin, S., Gaiano, N., Culp, P., Burns, J. C., Friedmann, T., Yee, J. K., and Hopkins, N. (1994). Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**, 666–669.

Maddison, L. A., Lu, J., Victoroff, T., Scott, E., Baier, H., and Chen, W. (2009). A gain-of-function screen in zebrafish identifies a guanylate cyclase with a role in neuronal degeneration. *Mol. Genet. Genomics* **281**, 551–563.

Mooslehner, K., Karls, U., and Harbers, K. (1990). Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J. Virol.* **64**, 3056–3058.

Mullins, M. C., Hammerschmidt, M., Haffter, P., and Nusslein-Volhard, C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* **4**, 189–202.

Nasevicius, A., and Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**, 216–220.

Nishigori, H., Yamada, S., Kohama, T., Tomura, H., Sho, K., Horikawa, Y., Bell, G. I., Takeuchi, T., and Takeda, J. (1998). Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction. *Diabetes* **47**, 1354–1355.

Parinov, S., Kondrichin, I., Korzh, V., and Emelyanov, A. (2004). Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes *in vivo*. *Dev. Dyn.* **231**, 449–459.

Postlethwait, J. H., and Talbot, W. S. (1997). Zebrafish genomics: from mutants to genes. *Trends Genet.* **13**, 183–190.

Raz, E., van Luenen, H. G., Schaerringer, B., Plasterk, R. H., and Driever, W. (1998). Transposition of the nematode Caenorhabditis elegans Tc3 element in the zebrafish Danio rerio. *Curr. Biol.* **8**, 82–88.

Roberg-Perez, K., Carlson, C. M., and Largaespada, D. A. (2003). MTID: a database of Sleeping Beauty transposon insertions in mice. *Nucleic Acids Res.* **31**, 78–81.

Ruttledge, M. H., Sarrazin, J., Rangarantnam, S., Phelan, C. M., Twist, E., Merel, P., Delattre, O., Thomas, G., Nordenskjold, M., Collins, P., Dumanski, J. P., and Rouleau, G. A. (1994). Evidence for the complete inactivation of the NF2 gene in the majority of sporadic meningiomas. *Nat. Genet.* **6**, 180–184.

Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64**, 907–912.

Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B., and Postlethwait, J. (2002). Sex-specific recombination rates in zebrafish (Danio rerio). *Genetics* **160**, 649–657.

Sivasubbu, S., Balciunas, D., Davidson, A. E., Pickart, M. A., Hermanson, S. B., Wangensteen, K. J., Wolbrink, D. C., and Ekker, S. C. (2006). Gene-breaking transposon mutagenesis reveals an essential role for histone H2afza in zebrafish larval development. *Mech. Dev.* **123**, 513–529.

Solnica-Krezel, L., Schier, A. F., and Driever, W. (1994). Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**, 1401–1420.

Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverty, T., Mozden, N., Misra, S., and Rubin, G. M. (1999). The Berkeley Drosophila Genome Project gene disruption project: single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135–177.

Stanford, W. L., Cohn, J. B., and Cordes, S. P. (2001). Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2**, 756–768.

Stemple, D. L. (2004). TILLING – a high-throughput harvest for functional genomics. *Nat. Rev. Genet.* **5**, 145–150.

Stuart, G. W., McMurray, J. V., and Westerfield, M. (1988). Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**, 403–412.

Sun, Z., and Hopkins, N. (2001). vhnf1, the MODY5 and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, pronephros, and hindbrain. *Genes Dev.* **15**, 3217–3229.

Trofatter, J. A., MacCollin, M. M., Rutter, J. L., Murrell, J. R., Duyao, M. P., Parry, D. M., Eldridge, R., Kley, N., Menon, A. G., and Pulaski, K., *et al*. (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* **72**, 791–800.

Wagner, R. R. (1972). Rhabdoviridae and their replication. *In* "Fundamental Virology," (B. N. Fields, and D. M. Knipe, eds.), pp. 489–503. Raven Press, Ltd, New York.

Wang, D., Jao, L. E., Zheng, N., Dolan, K., Ivey, J., Zonies, S., Wu, X., Wu, K., Yang, H., Meng, Q., Zhu, Z., Zhang, B., Lin, S., and Burgess, S. M. (2007). Efficient genome-wide mutagenesis of zebrafish genes by retroviral insertions. *Proc. Natl. Acad. Sci. USA* **104**, 12428–12433.

Weiss, R. A., Boettiger, D., and Love, D. N. (1974). Phenotypic mixing between vesicular stomatitis virus and avian RNA tumor viruses, Cold Spring Harbor Symp. *Quant. Biol.* **39**, 913–918.

Wienholds, E., Schulte-Merker, S., Walderich, B., and Plasterk, R. H. (2002). Target-selected inactivation of the zebrafish rag1 gene. *Science* **297**, 99–102.

Wiles, M. V., Vauti, F., Otte, J., Fuchtbauer, E. M., Ruiz, P., Fuchtbauer, A., Arnold, H. H., Lehrach, H., Metz, T., von Melchner, H., and Wurst, W. (2000). Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nat. Genet.* **24**, 13–14.

Wu, X., Li, Y., Crise, B., and Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751.

Yan, Y. L., Miller, C. T., Nissen, R. M., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P. A., Amsterdam, A., Chung, B. C., Westerfield, M., Haffter, P., Hopkins, N., Kimmel, C., and Postlethwait, J. H. (2002). A zebrafish sox9 gene required for cartilage morphogenesis. *Development* **129**, 5065–5079.

Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. *Development* **124**, 3157–3165.

Young, R. M., Marty, S., Nakano, Y., Wang, H., Yamamoto, D., Lin, S., and Allende, M. L. (2002). Zebrafish yolk-specific not really started (NRS) gene is a vertebrate homolog of the Drosophila spinster gene and is essential for embryogenesis. *Dev. Dyn.* **223**, 298–305.

Zwaal, R. R., Broeks, A., van Meurs, J., Groenen, J. T., and Plasterk, R. H. (1993). Target-selected gene inactivation in Caenorhabditis elegans by using a frozen transposon insertion mutant bank. *Proc. Natl Acad. Sci. U. S. A.* **90**, 7431–7435.

CHAPTER 5

Genetic Screens for Mutations Affecting Adult Traits and Parental-effect Genes

Francisco Pelegri* **and Mary C. Mullins†**

*University of Wisconsin – Madison, Laboratory of Genetics, Madison, Wisconsin, USA

†Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00005-7

Abstract

Forward genetic analysis in the zebrafish has largely until now been restricted to the developmental period from the time of zygotic genome activation through the end of embryogenesis. However, the use of the zebrafish as a model system for the analysis of larval, juvenile and adult traits, including fertility and maternal and paternal effects, is gaining momentum. Here, we describe two different approaches, an F3-extended family and a gynogenesis-based approach, that allow genetic screening for and recovery of mutations affecting post-embryonic stages, including adult traits, fertility, and parental effects. For each approach, we also describe strategies to maintain and map the identified mutations.

I. Introduction

A. Extending the Range of Forward Genetics Analysis in Zebrafish Screens

The zebrafish continues to gain momentum as a model system to understand a wide range of biological processes. In spite of its potential range, zebrafish genetic research in the previous two decades has largely focused on the identification and analysis of zygotically driven genetic networks involved in embryonic development. The range of this more thoroughly studied time period has been determined by both biological and practical reasons: The earlier time point corresponds to about 3 h post fertilization (hpf), the time at which embryos begin to rely on zygotic gene expression. The latter limit of this range is at ~5 days post fertilization (dpf), the latest point that the embryo is nutritionally self-sufficient and does not require an externally supplied food source, which facilitates analysis by the investigator. Intensive studies

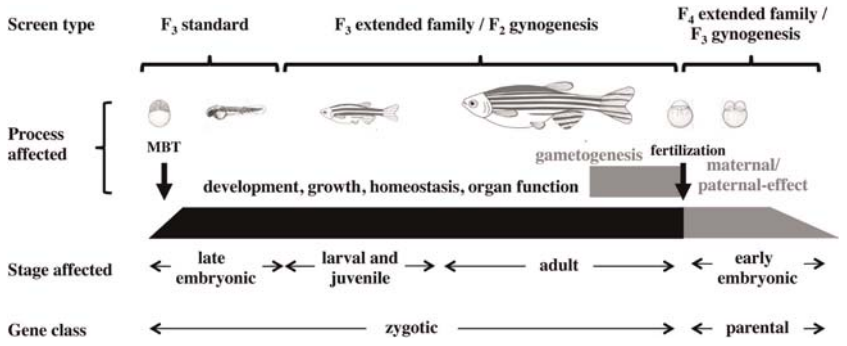


Fig. 1 Use of various genetic schemes to target different developmental stages in the zebrafish. Standard inbreeding approaches have been used to identify recessive zygotic mutations affecting the period between the activation of the zygotic genome at MBT (ca. 3 hpf) until 5 dpf. The identification of recessive mutations in stages past 5 dpf and into adulthood has been facilitated by the use of strategies described in this chapter, such as extended family- and gynogenesis-based approaches. The identification of parental-effect genes, which act prior to the MBT, requires producing an additional generation. Mutations affecting gametogenesis can be identified using strategies targetting adult or parental-effect genes. See text for details.

of developmental processes in this limited developmental time period have generated an enormous amount of crucial information to understand key biological questions, including aspects of cell fate diversification, the behavior and movement of cells and tissues during normal development, and the intricate interactions between these two types of processes during the development of the body plan and organ systems. However, development encompasses a much broader period that extends after the embryo exhausts its yolk supply at 5 dpf and enters larval, juvenile, and adult stages (Fig. 1) (Gupta and Mullins, 2010). Moreover, development of gonads and gametogenesis in adults is essential for the generation of functional gametes and to provide products essential in the next generation for the cleavage stage embryo prior to the activation of the zygotic genome, closing full circle the generational life cycle.

Because of the repeated use of developmental programs throughout the lifespan of an organism, much of the knowledge gained from the genetic analysis of zygotic genes in embryonic development has shed light on understanding related processes at other developmental stages. Thus, zygotically driven signaling pathways involved in early development have been shown in various model organisms to be reused during subsequent growth, in adulthood and aging, and the same pathways are affected in disease conditions such as cancer. However, a large gap remains in understanding the genetic circuitry acting at these other stages of development. The direct genetic dissection of juvenile, adult, and parental-effect traits through forward genetics promises to provide important new clues to these processes.

The genetic attributes of the zebrafish allow the powerful method of forward genetics, so effective in other model systems, to be applied in a vertebrate to identify and study the functions of genes through loss-of-function analysis. Such an approach can identify in a systematic and unbiased manner a myriad of genes that are essential for a given process, establishing gene pathways within a molecular framework,, or filling gaps in our knowledge in particular processes, often providing insights in unpredicted ways.

This chapter presents forward genetic methods for the analysis of juvenile, adult, and maternal- and paternal-effect traits. These methods were originally developed and carried out primarily to identify maternal-effect genes and previous versions of this chapter have focused solely on this topic. Such methods were subsequently shown also to allow the identification of paternal-effect genes and adult traits. However, maternal and paternal effects, as well as defects during gametogenesis also leading to sterility, can be regarded as a specific type of adult trait (involved in the production of viable progeny). In this chapter, we have expanded this initial focus to include additionally the genetic analysis of juvenile and adult traits, because the basic approaches are similar.

B. Application of Genetic Screens Beyond the Embryonic Lethal Range

Screening for juvenile and adult phenotypic traits requires screening during the growth and/or aging of the potentially affected clutches (Fig. 1). In addition, and somewhat counterintuitively, screens for the earliest stages of embryonic development must be carried out through genetic schemes that utilize an additional generation of crosses (Fig. 1). This is because the earliest stages of embryonic

development rely on maternally and paternally derived products, and homozygous mutant adults must be generated to observe the effect of mutations in these inherited products. We note that mutations causing defects in gametogenesis that lead to sterility can be screened for either directly in the homozygous mutant individuals (as a *bona fide* adult trait) or after breeding, preselecting mutant individuals that fail to produce viable offspring (in a maternal- or paternal-effect screen) followed by the analysis of the gonads of the affected parents.

1. Post-embryonic Development: Juvenile and Adults Stages

Depletion of nutritional sources in the yolk cell and the development of the digestive system, swim bladder inflation, and the appearance of foraging behavior are synchronized in the zebrafish embryo, such that by day 5 of development larvae no longer rely on maternal food resources and must find and process extraneous nutrients (Kimmel *et al*., 1995; Parichy *et al*., 2009). Genetic screens probing stages beyond 5 dpf necessitate the raising of the clutch through a full schedule of feeding in an aquarium system. Since genome-wide screening in the zebrafish is carried out by random inbreeding of a sufficiently large number of crosses to maximally homozygose the mutagenized genome, this extra generation and rearing requirements generate a challenge, as now a large number of clutches for each mutagenized genome need to be maintained and grown until testing.

Nevertheless, many of the phenotypes expected to be of high biomedical relevance (e.g. those relevant to organ function, physiology, and propensity to disease) will be manifested in juveniles and adults (for simplicity, hereafter referred to as adults) (Gupta and Mullins, 2010). Thus it would be advantageous to dissect such traits using forward genetics. A handful of genetic screens have identified recessive mutations affecting adult traits, including body shape, pigmentation, and eye and fin morphology (Haffter *et al*., 1996; Johnson *et al*., 1995; Lang *et al*., 2009; Parichy and Turner, 2003; Wagner *et al*., 2004), skeletal structures (Andreeva *et al*., 2011; Harris *et al*., 2008), ocular development (Tschopp *et al*., 2010), fin regeneration (Johnson *et al*., 1995), scale formation (Rohner *et al*., 2009) and gonad formation (Bauer and Goetz, 2001), and other adult morphological features (Wagner *et al*., 2004).

2. Pre-midblastula Development: Maternal- and Paternal-Effect Genes

In all animals, development from fertilization to the activation of the zygotic genome at the mid-blastula transition (MBT, Newport and Kirschner, 1982a,b; Signoret and Lefresne, 1971) depends on maternal factors made during oogenesis and activated upon fertilization. By necessity, all cellular and developmental processes acting during this time window are carried out solely by such maternal factors. Although the activation of widespread zygotic gene expression at the MBT marks the beginning of zygotic gene control during development, it does not imply an absolute shift between the use of maternal and zygotic products. Rather, in many instances perduring maternal products interact with newly expressed zygotic

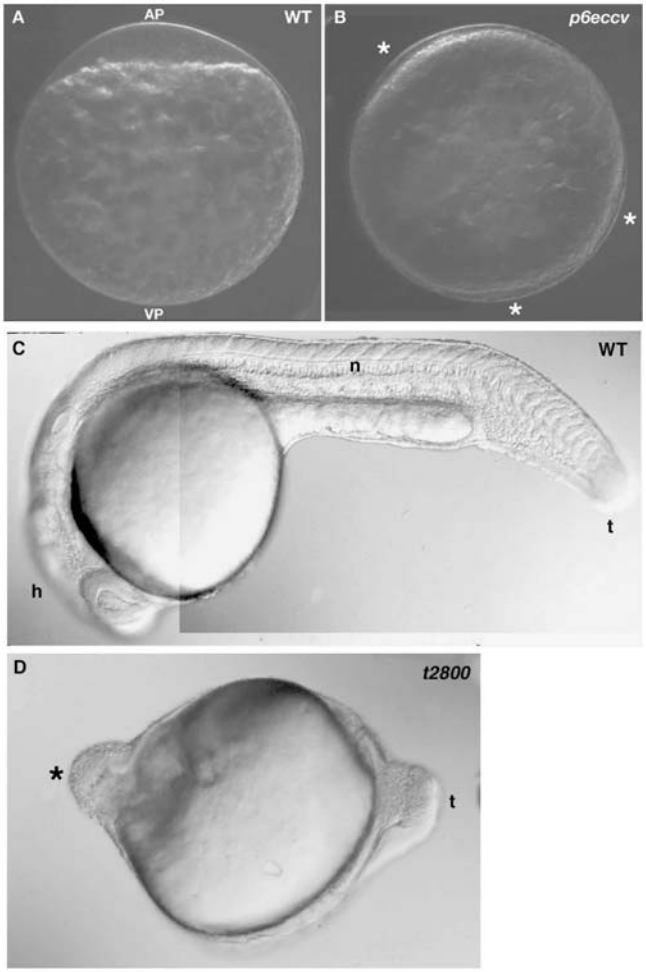


Fig. 2 Two mutants identified and recovered in systematic screens for maternal-effect mutants. (A,B) The animal–vegetal polarity mutant *magellan*^{p6cv}*(p6eccv)* was identified in an F_4 natural crosses screen (Dosch *et al*., 2004; Gupta *et al*., 2010). (A) A wild-type egg shortly after fertilization displaying the blastodisc prominently at the animal pole. (B) In contrast, in the p6eccv mutant the cytoplasm segregates to multiple locations around the circumference of the egg (asterisks). AP and VP are the animal and vegetal poles, respectively (A,B photos courtesy of Florence Marlow). (C, D) The mutation *hecate*^{t2800} (*t2800*), recovered in an EP-based screen (Pelegri *et al*., 2004), results in defects in the induction of dorsoanterior cell fates. (C) A wild-type embryo 24 hpf shows the normal body plan, including the head (h) and the notochord (n), a dorsal mesoderm derivative. (D) An embryo from *hecate* mutant mothers lacks anterodorsal structures and is radially symmetric (Lyman-Gingerich *et al*., 2005, 2006). (t) indicates the tail region in both wild-type and mutant embryos, which is less extended and contains multiple folds in the mutant. The asterisk indicates a group of cells that accumulates at the anterior of the embryo.

products to control developmental processes even after the activation of the zygotic genome. Maternal products are instrumental for the generation of the embryonic body plan, for example by establishing both the animal–vegetal and dorsal–ventral axes (Fig. 2). The animal–vegetal axis is established during oogenesis and marks the prospective anterior–posterior axis of the embryo, whereas the dorsal–ventral axis forms during the early cleavage stages prior to the MBT.

Genetic analysis in invertebrate model organisms, such as *Drosophila* and *Caenorhabditis elegans*, has revealed networks of maternal factors involved in basic cellular functions, establishment of egg polarity, and the regulation of cell fates (Kemphues and Strome, 1997; Schnabel and Priess, 1997; St. Johnston and Nüsslein-Volhard, 1992). In vertebrates, a much more limited knowledge of maternal gene function has been acquired in model organisms such as Xenopus through embryological and molecular approaches, and in mouse through targetted gene mutation approaches. Studies in teleost fish including the zebrafish have also begun to address the requirement of maternally driven genes in early development. Such maternal processes span basic cellular functions such as fertilization, egg activation, and the early cellular and nuclear divisions, as well as the induction of embryonic cell fates and the execution of morphogenetic movements during gastrulation (reviewed in Abrams and Mullins, 2009; Lindeman and Pelegri, 2010; Putiri and Pelegri, 2008). Genetically, mutations in such maternal-effect genes manifest themselves in the embryos from females homozygous mutant for these genes (for recessive alleles), regardless of the genetic composition of the sperm. The genotype of the sperm is not important because the function of the gene *in the embryo* depends on its expression *in the mother* during oogenesis and in this way the gene function is supplied to the egg. For this same reason, the mutant phenotype is expressed in 100% of the progeny of mutant females (in cases of fully penetrant mutations), because all eggs are made from the same mutant ovarian germline. The first forward genetic screens to isolate recessive maternal-effect mutations in vertebrate species have been performed in the zebrafish (Dekens *et al*., 2003; Dosch *et al*., 2004; Kishimoto *et al*., 2004; Pelegri and Schulte-Merker, 1999; Pelegri *et al*., 1999, 2004; Wagner *et al*., 2004) and have led to a number of important new findings (reviewed in Abrams and Mullins, 2009; Lindeman and Pelegri, 2010; Putiri and Pelegri, 2008).

Testing of putative homozygous males in the Mullins screen also led to the identification of a surprisingly large number of paternal-effect mutations (Dosch *et al*., 2004; Wagner *et al*., 2004; Yabe *et al*., 2007). Paternal-effect genes were manifested in the progeny of homozygous mutant males, irrespective of the genotype of the mother. The identification of paternal-effect genes was somewhat unexpected considering the well-established view that most of the materials inherited by the embryo from its parents are stored in the large, sessile oocyte, as opposed to the much smaller, motile sperm cell. However, the identification of these mutations is consistent with the expected requirement for male germline genes in chromosomal segregation during meiosis in spermatogenesis (Poss *et al*., 2004), as well as the inheritance of a pair of centrioles in the sperm cells for the centriole-less oocyte (Yabe *et al*., 2007), and the isolation of paternal-effect mutants in Drosophila (Wakimoto *et al*., 2004).

In spite of this progress, our knowledge of the functions of maternal- and paternal-effect genes (which for simplicity we will refer to as parental-effect) in early development in zebrafish and vertebrates in general remains superficial and disconnected. Ongoing and future screens in zebrafish, together with studies of the current mutants in zebrafish and mouse, and further molecular investigations in other model systems such as Xenopus should help fill these gaps of knowledge.

II. Strategies for Adult Trait and Parental-effect Screens

There are two approaches that have been used to identify *de novo* recessive mutations affecting adult and parental-effect genes in the zebrafish: a) F_3 screen for adult traits and F_4 for parental-effect traits based solely on natural crosses and b) an F_2 screen for adult and F_3 for parental-effect based on gynogenesis, specifically the technique of Early-Pressure (EP). These alternatives differ in various important ways, which are summarized in Table I and described throughout this chapter. Here, we discuss these alternatives and provide detailed protocols to implement them.

A. F_3 Screen for Adult Traits and F_4 for Parental-effect Traits Based on Natural Crosses that Integrates a Mapping Strategy

One method to produce recessive, homozygous maternal-effect mutants is through a three-generation inbreeding strategy. Although at first glance this method is expected to occupy an enormous amount of tank space, consolidating the F_3 generation into a

Table I
Summary between an F_4 screen based solely on natural crosses and an F_3 screen based on gynogenesis (EP)

	Genetic approach	
	F_4 based on natural crosses, (F3-extended family)	F_3 based on gynogenesis
Background strain used in screen	Lethal- and sterile free	Lethal- and sterile-free. Amenable to IVF and EP-based gynogenesis. Needs to produce females under gynogenetic conditions
Basic methodology	Sibling pair matings	Induction of gynogenetic clutches using EP
Number of generations needed	Three, plus maternal-effect test	Two, plus maternal-effect test
Amount of space needed	Large to moderate	Moderate to small
Fraction of females within a family expected to exhibit maternal-effect phenotype	1/16	Variable, between 50% toward 0%, depending on distance between the locus to the centromere
Incorporation of mapping scheme	Feasible	Difficult, due to low number of surviving gynogenotes
Identification of maternal-zygotic and paternal-effect mutations	Feasible	Feasible, but more difficult due to reduced fertility of gynogenotes

single tank comprised of an "F_3-extended family" makes this approach considerably more practical. This scheme begins similarly to F_3 zygotic screens (Driever *et al*., 1996; Haffter *et al*., 1996; Mullins *et al*., 1994). As shown in Fig. 3A, G_0 male fish are mutagenized with 3 or 3.3 mM ENU as described (Mullins *et al*., 1994), crossed to wild-type females, and their F_1 progeny raised. Each F_1 fish carries a different set of mutagenized genes derived from the independently mutagenized spermatogonial cells of their fathers. F_1 fish are interbred and the F_2 progeny, referred to as an F_2 family, are raised to adulthood. Each F_2 family contains two mutagenized genomes, one from its mother and one from its father. In a zygotic screen, the fish of a given F_2 family are intercrossed, and the F_3 embryos are screened for recessive mutant defects. To identify adult and parental-effect mutations, the F_3 progeny are instead raised to adulthood. As with a zygotic mutation, for an adult or parental-effect mutation that is present in an F_2 family, one-quarter of the F_2 intercrosses will yield F_3 families containing 25% mutant adults or parental-effect mutants.

Similar to a zygotic mutation, if a recessive, adult, or parental-effect mutation exists in an F_2 family, the probability of identifying it directly depends on the number of F_3 families generated from F_2 intercrosses and the number of F_3 females screened from each F_3 family (see Section II.A.3. for formulas). To obtain a 90% probability of making a mutation homozygous, eight F_2 intercrosses are required, corresponding to eight F_3 families raised from each F_2 family. If these eight F_3 families each occupy a separate tank, then an eightfold increase in tank space is required beyond that needed for an F3 zygotic screen. To make such a maternal-effect screen practical for the moderately sized fish facility, multiple F_3 crosses from a single F_2 family are pooled into an "F_3-extended family," comprised of siblings and cousins, and raised in a single tank. Specifically, if eight crosses from each F_2 family are desired, then equal numbers of each cross are pooled and raised together. If it is necessary to set up the F_2 family more than once to obtain the desired eight crosses, then separate pools are generated on different days, which are then pooled at a later point.

In an F_3-extended family, recessive adult or parental-effect mutants represent one-sixteenth of the total adults or males or females, rather than one-quarter because F_2 intercrosses generating mutant and non-mutant progeny are pooled. The advantage is that only one tank is occupied and screened, rather than 8, which is significant savings in space. However, the same number of adults is screened whether they are in eight separate or one consolidated tank. To obtain an 80% probability of identifying a mutant if it exists in an F_3-extended family (see also Section II.A.3), 25 F_3 individuals must be screened. F_3 individuals are screened for phenotypes using the desired assays (adult phenotypes) or by crossing them to sibling or wild-type individuals and examining their F_4 progeny for defects (parental-effect phenotypes). Far fewer F_4 embryos are examined for defects in a parental-effect screen than a zygotic screen, because all or nearly all embryos are affected, in contrast to 25% for a zygotic mutant phenotype.

1. An F_3-Extended Family Approach with Integrated Mapping

A chromosomal mapping cross can be integrated into an F_3 adult or F_4 parental-effect screen using a natural crosses approach, which allows one to map the mutation

to a chromosomal position in the F_3 generation. This integrated mapping strategy is particularly beneficial in the case of parental-effect mutations, which are difficult to propagate because they typically produce all non-viable progeny. However, a similar difficulty would arise if the mutations are detrimental during the juvenile or adult stages to the extent that mutant carriers cannot reproduce. In these cases it is necessary to identify heterozygous females and males. This can be greatly facilitated by mapping the mutation to a chromosomal position, which has the additional value of initiating the molecular isolation of the mutated gene. The mapping strategy is discussed further in Section V.A.

2. Identification of Maternal-Zygotic, Male Sterile, and Paternal-effect Mutations

Crossing F_3 females with F_3 sibling/cousin males permits the simultaneous isolation of maternal-zygotic, zygotic, as well as male sterile and paternal-effect mutations. F_3 intercrosses provide a ~50% probability of detecting a mutant that requires loss of both maternal and zygotic gene activity, which is not possible if the F_3 female is crossed with a wild-type male. All zygotic mutations in the F_2 family are still present in the F_3 family, so the F_4 embryos from F_3 intercrosses can be screened not only for maternal-zygotic, but also for zygotic mutants, if desired. In F_3 sibling/cousin crosses, paternal-effect and male-sterile mutations are also revealed.

If a mutant is identified in the F_4 embryos, the type of mutation induced can be distinguished by crossing the F_3 female and male parental fish separately with wild-type fish and examining the progeny for defects. If the mutation is a maternal- or paternal-effect or a female or male-sterile mutation, then the defect is evident in the F_4 embryos, even when the F_3 fish is crossed with wild-type. For a maternal-effect or female-sterile mutation, the F_3 female will be the cause of the defective embryos, whereas for a paternal-effect or male-sterile mutation, the F_3 male parent will be the sole cause of the defect. For recessive zygotic and maternal-zygotic mutations, the F_4 embryonic defect will depend on both F_3 parents. A different fraction of affected F_4 embryos is expected for fully penetrant maternal-zygotic versus zygotic mutations. For a maternal-zygotic mutant, the F_3 female is homozygous and the F_3 male heterozygous for the mutation resulting in 50% mutant progeny, whereas a zygotic mutation yields 25% F_4 mutants.

3. Assessment of the F_4 Natural Crosses Screen for Maternal-effect Mutations

Typical values for several parameters of an F_4 natural crosses screen are shown in Table II. The results of an F_4 natural crosses strategy, using an F_3-extended family and incorporating a mapping cross, have been published (Dosch *et al*., 2004; Wagner *et al*., 2004). In such an F_4 screen, the number of mutagenized genomes (G) screened contributed by a given F3-extended family is determined by the following formula: $G = (1–0.9375^n) \times 2 \times (1–0.75^m)$. The term $(1–0.9375^n)$ is the probability of identifying an F_3, recessive maternal-effect mutant female present in an F_3-extended family if "n" females are screened within that family. The factor "2" represents the two mutagenized genomes derived from the two F_1 fish. The term $(1–0.75^m)$ is the fraction

Table II
Statistics in an F_4 natural crosses screen

% F_2 families used to generate F3 families[a]	85%
# F_2 intercrosses generating an F3-extended family	>8
Fraction of F_1 mutagenized genomes homozygous in F_3-extended family	>90%
# F_3 females screened/F_3-extended family	24
Fraction of genomes screened in F_3 family	79%
# haploid genomes screened /# F_3-extended families[b]	1.4
Maternal-effect mutants identified/genome screened	0.11 genomes
Fraction of candidate mutations that are recovered	95%

[a] Due to the female bias in hybrid strains that are needed to make the mapping cross, a fraction of the F_2 families have three or less males. These families are difficult to work with and so are discarded.

[b] Since F_1 fish are interbred to make the F_2 genome, there are two mutagenized genomes present in the F_2 family and therefore greater than 1 mutagenized genome is ultimately screened in each F_3-extended family.

of the two mutagenized genomes expected to be homozygous in the F_3 generation, where "m" is the number of F_2 crosses that comprise an F_3-extended family.

B. F_2 Adult or F_3 Parental-effect Trait Screens Based on EP-induced Gynogenesis

Artificially induced gynogenesis in zebrafish involves diploidization of the maternal haploid genome, producing viable offspring with solely a maternal genetic contribution (Streisinger *et al*., 1981). The incorporation of gynogenesis into a genetic scheme for adult or parental-effect mutations allows the direct production of homozygotes for induced mutations from a single heterozygous F_1 carrier, bypassing one generation in comparison to a scheme based solely on natural crosses (compare Fig. 3A and B). Due to the large number of chromosomes present in the zebrafish, genetic screens in this organism involve the whole genome and are essentially blind, screening all mutagenized chromosomes simultaneously, rather than individually as done in the fly and worm. Therefore each generation in a screen generates an exponentially increasing number of crosses. Thus, bypassing one generation through the use of EP allows a significant reduction of the time and space required to carry out an adult or parental-effect screen.

In a basic gynogenesis-based scheme (Fig. 3B), mutations are induced in the germline of parental (P) males by exposing them to the point-mutagen N-ethyl-N-nitroso-urea (ENU) (Mullins *et al*., 1994; Solnica-Krezel *et al*., 1994; van Eeden *et al*., 1999). P males are then crossed to produce F1 progeny heterozygous for induced mutations. Eggs are extruded from the F_1 females and gynogenesis is induced. This allows newly induced mutations to become homozygous in up to 50% of the gynogenetic F_2 generation (see Section II.B.1). Adult F_2 individuals can be screened directly for mutations affecting adult traits as well as parental-effects by testing their F_3 progeny for embryonic phenotypes.

In the EP-based screen, the production of F_3 clutches during a parental-effect screen is best achieved by *in vitro* fertilization (IVF) using wild-type sperm. This is

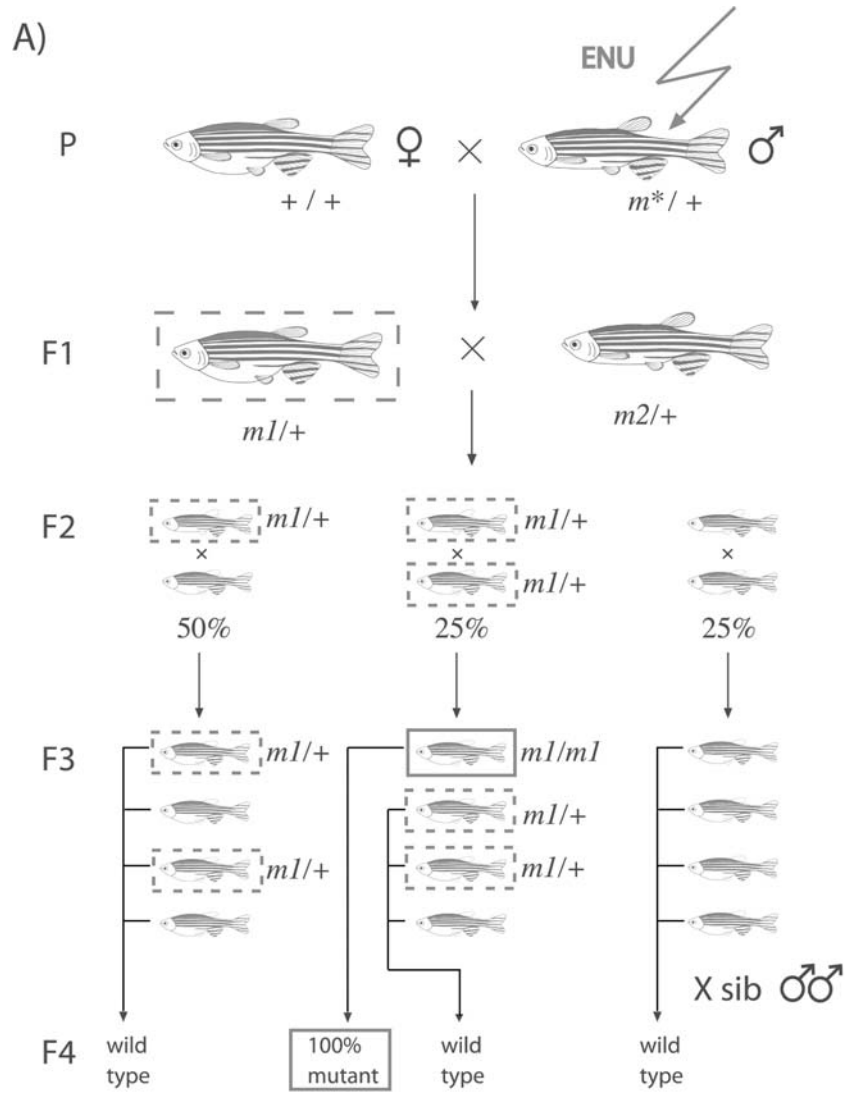


Fig. 3 (A) F_4 natural crosses screen strategy. Males of the parental generation (P) are mutagenized with ethylnitrosourea (ENU), to induce new mutations *(m*)* and crossed to wild-type females. F_1 fish are raised, each of which carries a different set of mutagenized genes. Two mutations are shown, *m1* and *m2*, each carried by one of the two F_1 fish. Only mutation *m1* is followed in subsequent generations for simplicity. Two F_1 fish are intercrossed and an F_2 family raised. Half of the individuals of the F_2 family are heterozygous for *m1*. The F_2 fish are intercrossed to make an F_3-extended family (see text), composed of equal numbers of F_3 fish from each of the F_2 intercrosses of one family. One-quarter of the F_2 intercrosses are between *m1* heterozygotes, producing *m1* homozygotes in 25% of their F_3 progeny. F_3 individuals can be tested for larval, juvenile, and adult phenotypes, as well as for parental effects in the F_4 generation. (B) EP-based screen strategy. F_1 heterozygous females carrying newly induced mutations (boxed with hatched lines; *m** and *m1* as in (A)) are treated to induce gynogenetic F_2 clutches, which can contain homozygotes for maternal-effect mutations (boxed with solid lines). F_2 individuals can be tested for larval, juvenile, and adult phenotypes, as well as for parental effects in the F_3 generation. A fraction of EP-derived progeny will be heterozygous for the mutation. Hypothetical results are shown using a gene with an average centromere–locus distance (see text).

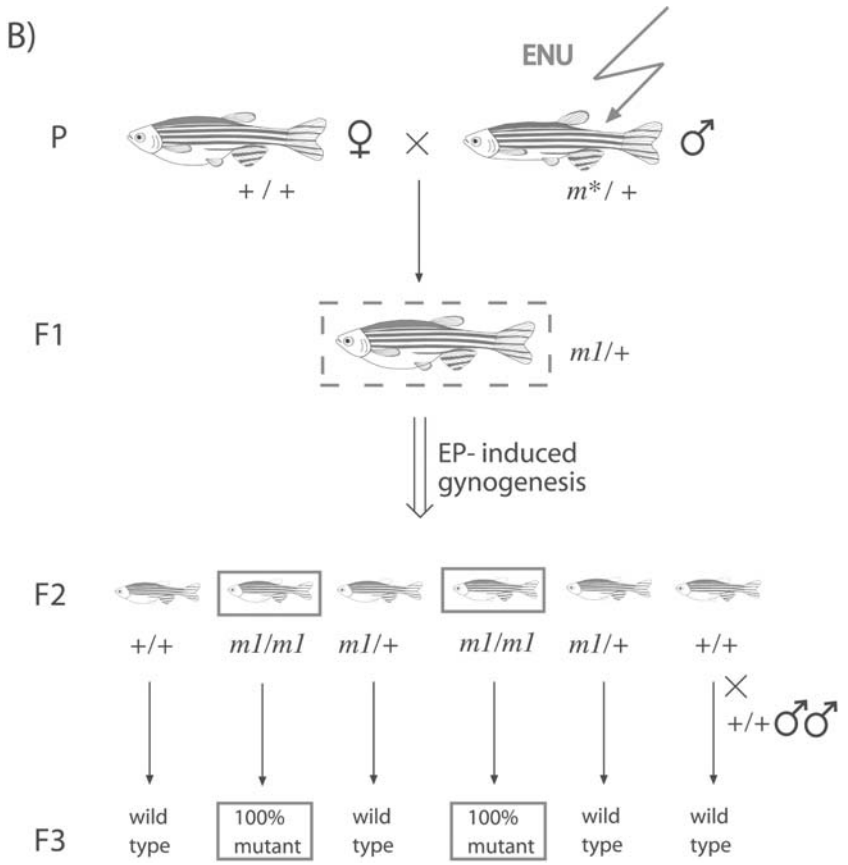


Fig. 3 ***(Cont.)***

because sib–sib crosses mate inefficiently due to the semisterility of sibling males caused by both the female-rich genetic background (see Section III.B.2) and EP-induced inbreeding. On the other hand, IVF is facilitated by the fact that females from the background line are easily stripped of eggs. The use of IVF precludes the possibility of identifying maternal-zygotic mutations (see Section II.A.2), although the use of selected lines may allow this in the future. However, IVF has the advantage that it allows the production of F_3 clutches that can be immediately observed and followed synchronously, thus facilitating the identification of early phenotypes (see Section II.C). In addition, because all clutches are fertilized by the same batch of sperm solution, IVF allows the rapid identification of maternal mutations that affect the ability of the egg to become fertilized (which requires re-testing in a natural crosses screen approach, because unfertilized clutches are sporadically observed in natural matings from wild-type parents). Because of the relatively high incidence of semisterility in EP-derived clutches, this strategy is not recommended for the

isolation of male-sterile mutations causing a lack of fertilization (e.g., resulting from defects in spermatogenesis), although this approach is in principle amenable to the identification of paternal-effect phenotypes in fertilized embryos (e.g., defects in chromosome segregation during meiosis (Poss *et al*., 2004), centriole complement defects (Yabe *et al*., 2007)) that lead to embryonic lethality.

Because of the relatively small number of surviving individuals in the F_2 gynogenetic clutches, it is not currently practical to incorporate a mapping strategy within an EP-based genetic screen as described above for screens based solely on natural crosses (see Section II.A.1 and V.A), although this may be possible in the future through the use of selected polymorphic lines (see Section III.B). Rather, in an EP-based screen, mapping crosses are currently initiated after recovery of the mutation (see Section V.B).

Although the incorporation of gynogenesis can simplify an adult and parental-effect genetic screen, gynogenesis itself is only efficiently induced under specific conditions. In a scheme for such a gynogenesis-based screen, the main goal is the efficient production of fertile gynogenetic F_2 adults that are homozygous for newly induced mutations. A number of variables need to be optimized to carry out this procedure, which we discuss below. First, a suitable method of gynogenesis needs to be selected. Secondly, an appropriate mutagenesis dosage needs to be chosen to induce a reasonably high rate of mutations, while allowing the production of viable homozygous adult mutants. In addition, lines amenable to gynogenetic procedures need to be selected (see Section III.B).

1. Choice of Gynogenetic Method

There are two main techniques for the artificial induction of gynogenesis in the zebrafish: Early Pressure (EP) and Heat Shock (HS) (Streisinger *et al*., 1981; see also Fig. 3 in Pelegri and Schulte-Merker, 1999). In both methods, eggs are first artificially fertilized with sperm whose genetic material has been inactivated by UV irradiation. In the absence of further treatment, these eggs would develop into haploid embryos that are inviable. Both EP and HS lead to the diploidization of the genetic content of the egg, thus producing viable, diploid embryos.

In EP, diploidization is induced by the application of hydrostatic pressure between 1.33 and 6 min after egg activation (see Section VI.C.6). This treatment inhibits completion of the second meiotic division and the expulsion of the 2 ° polar body, resulting in a diploid egg. HS, on the other hand, inhibits cytokinesis of the first mitotic division of haploid embryos by applying a heat pulse during 13–15 min after egg activation (see Section VI.C.5), transforming haploid embryos into diploid ones. Hydrostatic pressure, applied late, has also been used as an alternate method to inhibit the first mitosis, although it has been found to be less effective and more cumbersome than HS (Streisinger *et al*., 1981).

In theory, HS is more efficient than EP in the direct induction of homozygosity and therefore might be the technique of choice in a maternal-effect screen. This is because HS-derived progeny are homozygous at every single locus, and therefore

50% of HS-derived F_2 progeny are homozygous for a mutation present in heterozygous form in the F_1 mother. EP, on the other hand, due to recombination during meiosis, leads to a variable degree of homozygosity ranging from 50% toward 0% depending, respectively, on whether loci are linked to the centromere or are distally located. Thus, HS would, in principle, provide the highest possible yield of homozygous mutant adults for all loci regardless of their chromosomal location. Moreover, the expectation of a fixed percentage of homozygous mutant females would aid in the assessment of newly identified phenotypes.

In spite of these obvious theoretical advantages of HS over EP, in practice EP is superior to HS as a gynogenetic method for a number of reasons. First, EP is about twofold more efficient than HS in inducing diploid, viable gynogenotes (Table III; see also Streisinger *et al*., 1981), presumably due to a greater intrinsic ease of inhibiting the extrusion of the polar body during meiosis rather than cytokinesis during the first mitosis. Moreover, EP-derived adults, probably due to their higher heterozygosity, show viability and fertility rates that combined are about fourfold higher than those in HS-derived clutches (Table III). Thus, the final yield of fertile adult gynogenotes derived from EP is about eightfold higher than that derived from HS.

Higher levels of heterozygosity in EP-derived gynogenotes are beneficial for additional reasons. First, under mutagenic conditions, the yield of HS-gynogenetic clutches is expected to be further reduced by a factor of 0.5 per induced zygotic recessive-lethal mutation, while EP-derived clutches are expected to be reduced by an average factor of 0.23 (see Section II.B.2). Secondly, in the case of mutations that affect fertility, the increased heterozygosity of EP gynogenotes improves the odds of recovering newly identified mutations. This occurs because the overall fraction of *fertile* siblings that carry a given mutation, due to a decrease in the fraction of the (sterile) homozygous mutant females, is greater in EP-derived clutches than in HS-derived ones (see Section IV.A.2).

Table III
Comparison of Heat Shock- and Early Pressure-induced gynogenesis (gol-mix line)

	Heat Shock	Early Pressure
Viability at d5	0.09	0.21
(viable/fertilized eggs)[a]	$n = 3590$	$n = 4368$
Fraction clutches with > 6 viable d5 fish	0.41	0.93
	$n = 29$	$n = 29$
Clutch size	10	37
(viable d5 fish/clutch)	$n = 29$	$n = 29$
Adult viability	0.53	0.66
(viable at 3 mo./d5 viable)	$n = 324$	$n = 218$
Fertility	0.23	0.65[b]
(fertile adults/total adults)	$n = 13$	$n = 226$

[a] Viable at day 5 are defined as fish that can inflate their swim bladders.
[b] Value from F_2 descendants of P males mutagenized with 2 mM.

The main drawback of the higher heterozygosity of EP gynogenotes is that it leads to an intrinsic bias against the identification of distally located mutations. However, measurements of the frequency of homozygosity (F_m) of random zygotic mutations after EP-induced diploidization range from 0.50 to 0.04, with an average value of 0.23 (16 loci; Neuhauss, 1996; Streisinger *et al*., 1986). With the assumption that maternal genes are similarly distributed throughout the chromosomes, these data suggest that the majority of these genes are sufficiently close to a centromere to be identified through an EP-based screen.

For these reasons, we chose EP over HS as a gynogenetic method for our screen, although it is possible that HS may become applicable in the future with the use of selected lines.

2. Mutagenesis Dose

In the F_2 gynogenetic generation, homozygosity for mutations in essential zygotic genes will lead to a decreased survival of gynogenotes. For example, the mutagenic dosage used in large-scale zygotic screens (3 × 1 h treatments with 3 mM ENU) is expected to induce about one embryonic lethal and one larval lethal per haploid genome (Haffter *et al*., 1996; Mullins *et al*., 1994; Solnica-Krezel *et al*., 1994). This implies that under this mutagenic condition, only 59% of what would be otherwise viable EP-derived gynogenotes (25% using HS) would survive to adulthood. Thus, we reduced the ENU dosage in our maternal-effect screen experiments. Similar reductions in the strength of mutagenic treatments were adopted for maternal-effect screens in Drosophila and *C. elegans* (see, for example, Kemphues *et al*., 1988; Lehmann and Nüsslein-Volhard, 1986). We have observed that a mutagenic dosage of 3 × 1 h 2 mM ENU treatments begins to have a mild effect on the viability of F_2 gynogenetic clutches (not shown). These conditions lead to a mutagenic rate, as assayed by the frequency of newly induced *albino* alleles, estimated to be about one-third of the rate induced by the standard (3 mM ENU) treatment (Mullins *et al*., 1994; Pelegri and Schulte-Merker, 1999; Solnica-Krezel *et al*., 1994), or about 0.3–0.4 embryonic lethal mutations per haploid genome. We chose for our screen the ENU concentration of 2 mM as a compromise between a moderate mutagenic rate and a practical level of viability.

3. Assessment of EP-based Screens for Maternal-effect Mutations

The results from genetic screens using an EP-based method have been described elsewhere (Pelegri and Schulte-Merker, 1999; Pelegri *et al*., 2004, see also Dekens *et al*., 2003; Lyman-Gingerich *et al*., 2005; Pelegri *et al*., 1999). Typical survival and yield values are presented in Table IV.

To estimate the number of genomes screened using an EP-based method, one needs to keep in mind that the number of genomes screened depends on the level of EP-induced homozygosity (F_m), which in turn is inversely related to the centromere to locus distance. Thus, the number of genomes screened will differ according to the position of genes with respect to the centromere. To estimate such values, one can first

Table IV
Statistics in an F_3 gynogenesis-based screen

% F_2 clutches grown to adulthood[a]	45%
% F_2 clutches with fertile adult females[b]	20%
# screened haploid genomes/# F_2 clutches with fertile females[c]	0.45 (proximal) – 0.27 (average)
Maternal-effect mutants identified/# haploid genomes screened[d]	0.11–0.19
% of candidate mutations that are recovered[e]	44%

[a] EP-derived F_2 clutches with at least six viable fish on day 6 of development.
[b] Fertile females are defined as those that produce normal eggs, which upon activation exhibit the wild-type translucent appearance and can be fertilized to exhibit either a normal or a characteristic abnormality in the early cleavage pattern.
[c] The number of genomes screened depends not only on the number of females tested but also on the average distance of the loci to the centromere (Section II.B.1). The values presented are derived from the number of females tested for each family, so that mutations present in the family have a 0.97 chance of being identified, assuming F_m values of 0.50 and 0.23, respectively, for centromere-linked loci and loci at an average distance to the centromere. F_2 families are considered to carry a candidate mutation when they contain females that produce a phenotype in 100% of the F_3 offspring, and the phenotype appears in more than one independent F_3 clutch.
[d] The range given is estimated by assuming that the isolated mutations are all either proximal (0.11) or at an average distance to the centromere (0.19).
[e] Mutations in some lines are not recovered due to a variety of reasons: false positives in the original tests, inability to recover the line due to insufficient fish to perform recovery crosses, and variability in the penetrance of the mutation.

estimate the critical number of F_2 females that need to be screened in a given family to result in a 90% probability of detecting a newly induced mutation present in that family. For example, with an average F_m value of 0.23 (see Section II.B.1), nine F_2 tested females per clutch would be needed to detect a newly induced mutation with a 90% probability. This estimate corresponds to an average locus and varies greatly depending on the centromere–locus frequency: To reach a similar frequency of detection for centromere-linked (F_m: 0.5) and distal (F_m: 0.05) loci, the critical number of F_2 females tested per clutch is 4 and 44, respectively. For specific F_m values, each family where the number of tested F_2 females is equal to or larger than the critical number of F_2 females contributes one screened haploid genome. In cases where the number of F_2 females tested per family is less than the critical value for a given F_m, the number of tested F_2 females can be pooled to find a combined number of screened genomes contributed by that pool. The latter value is calculated by dividing the pooled number of tested F_2 females by the critical value needed to screen one haploid genome at a 90% certainty. The total number of haploid genomes screened is the sum of all families with more than the critical number of tested F_2 females and the combined number of genomes calculated from the pooled number of tested F_2 females.

C. Screening

Adults homozygous for newly induced mutations, either in the F_3 generation in an extended-family screen or the F_2 generation in an EP-based screen, can be screened

for adult or parental-effect phenotypes. In practice, both of these types of traits may be screened in tandem within the same screen, first for any number of adult phenotypes and subsequently, upon crossing and creating a subsequent generation, for parental-effect genes in the progeny. This is only applicable, however, when the adult phenotype assays are non-invasive and are not expected to interfere with each other, gametogenesis or mating behavior.

When adult screen assays are not compatible with each other (e.g., when adult assays are invasive or involve treatments that may affect another assay), multiple assays can still be combined albeit utilizing different sets of screened fish. However, the screens as normally carried out generate a relatively low number of homozygous mutants: for an F3 extended family approach 1/16 of a typical size family of 60, or about 4 (or 2 if the phenotype is specific to one of the sexes); for an F2 EP-based gynogenesis approach 1–12 (0–6 if sex-specific) of a clutch of 24 reaching adulthood. Considering that the identification of multiple individuals within the same family can be crucial to distinguish genetically transmitted traits from possible syndromes with a non-genetic basis, screens as typically carried out are therefore reaching the limits of detection. Therefore, in cases of non-compatible assays it is advisable to generate a larger number of adults (i.e., an increase in number proportional to the number of non-compatible assays being used). This allows for sufficient screenable adults for reliable detection of new phenotypes while taking advantage of the already generated mutagenized P0 males and heterozygous carrier F1 individuals.

1. Screening for Adult Traits

Adults potentially homozygous for newly induced mutations can be screened for any desired adult phenotype (see, for example, Andreeva *et al*., 2011; Bauer and Goetz, 2001; Haffter *et al*., 1996; Johnson *et al*., 1995; Lang *et al*., 2009; Tschopp *et al*., 2010). Given the limited number of expected adult homozygous mutants in a given family or clutch (see above), adult trait screens create a particular concern in that the screenable assays need to be robust. One aspect of assay robustness is whether the mutation exhibits the phenotype with a high penetrance when present in homozygous mutant individuals. Ideally, this can be confirmed by the existence of previously isolated, highly-penetrant mutations with a similar phenotype. In the absence of such preexisting mutations, a pilot screen may determine whether highly penetrant mutations affecting the trait can be isolated.

Another important variable to assay robustness is the incidence of false positives caused by non-genetic deficiencies and syndromes. This can be determined in practice by determining, prior to initiating the screen, the fraction of individuals that appear to test positive (defective) in a large number of unmutagenized, wild-type adult fish of the same genetic background. Because of sporadic false positive effects, it is highly advisable that, when possible, any putative homozygous mutant individual be tested at least a second time with the primary assay, or tested with a different yet related assay. Such retests or secondary screening tests can eliminate most sporadic effects. Acceptable rates of false positives are therefore determined by

the fraction of screened individuals that can be feasibly retested/secondarily tested during the screening process. For example, false positive incidence rates that result in having to retest up to 10% of all screen families may be acceptable, provided that these subsequent tests can be readily incorporated into the ongoing screen.

Even when a putative mutant individual retests with the same phenotype as in the primary screen, it remains a possibility that the cause is syndromic (non-genetic). In this case, secondary, more precise screens, which may be more involved but have a lower rate of false positive incidence, may allow one to distinguish between non-genetic syndromes and newly identified mutations. Ultimately, however, propagation of the mutation to a subsequent generation in a mendelian manner is the only fool-proof method to verify the genetic basis for an observed phenotype. Because such a propagation step is labor intensive, the assay(s) used in the screen should attempt to eliminate false positives as much as possible, thus minimizing the fraction of lines that need to be tested for mendelian segregation through additional generations.

2. Screening Embryos for Parental-effects

Once adults that may be homozygous for a parental-effect mutation are produced, embryos from those adults are screened for potential defects. For parental-effects, the uncertainty caused by sporadic non-genetic syndromes is lessened by the fact that parental-effect phenotypes caused by homozygosity in the parent are observed in multiple progeny individuals (ideally most if not all of potentially hundreds of embryos), each of them providing support for the existence of a true genetic mutation. This essentially amplifies the signal from the screen and frequently allows one to infer the presence of a new mutation from a single identified mutant individual (a notable exception to this is the case of axis-deficient or "ventralized" phenotypes, which occur sporadically as a maternal effect at a relatively high frequency in wild-type females; in this case retesting of the female or identification of multiple mutant affected sibling females with the same phenotype is essential).

To screen for parental-effect phenotypes, embryos are collected as early as possible or within 2 hpf. As mentioned above (see Section II.B), in EP-based strategies for maternal-effect genes, clutches can be derived by IVF. This allows one both to observe and synchronize their development immediately after fertilization. On the other hand, if the screened embryos are produced through natural crosses, early observation and synchronization of the clutches can be approached by taking advantage of the propensity of zebrafish to lay eggs during the early hours of their daylight cycle. In practice, this is done by setting up crosses toward the end of the light cycle and collecting embryos during the early hours of the following light cycle. Alternatively, pair matings can be set up during the first two hours of the light cycle, and embryos can be collected in the same day (this prevents the occurrence of eggs being laid during the previous night cycle but has the disadvantage that egg laying is less synchronized, and continuous monitoring for egg laying is needed throughout the first half of the light cycle).

The early collection of eggs allows one to discard clutches that have undergone aberrant ovulation, which can occur in a small fraction of clutches from wild-type females and which if undetected would provide false positives in the screen. This early observation also allows one to determine whether the cleavage and cellularization pattern characteristic of wild-type embryos is normal. Because unfertilized embryos also exhibit a pattern of irregular cleavages (pseudocleavages, Kane and Kimmel, 1993), it is important that such early embryos are carefully observed to reveal potential differences between pseudocleavage formation and an abnormal early cellular pattern. Once the regular pattern of cellular cleavage characteristic of normal fertilized embryos is detected, fertilized embryos are sorted and transferred to a clean plate at low densities (40 embryos per 10.5 cm dia. plate). Embryos are subsequently screened for deviations from the wild-type developmental pattern (Kimmel *et al*., 1995, see also van Eeden *et al*., 1999, for a sample scoring chart).

This screening strategy relies on the incorporation of the sperm into the egg, which is necessary for patterns of cleavage distinct from those in unfertilized eggs. Parental-effect mutations acting at earlier steps in oogenesis and egg maturation may also be identified by screening clutches shortly after fertilization, for example maternal-effect mutations affecting the animal–vegetal axis of the egg manifest their defects in the directional movement of ooplasm during egg activation (Fig. 3B, compare to A; Dosch *et al*., 2004), and paternal-effect mutations affecting centriole duplication during spermatogenesis exhibit a one-cycle delay in cell division (Yabe *et al*., 2007). In addition, it is possible to identify mutations affecting gametogenesis by selecting for non-egg laying females or in the case of males, selecting for a failure to fertilize eggs and then screening these individuals for gonadal defects by dissection and/or sectioning (Bauer and Goetz, 2001; M.C.M. unpublished).

III. Selection of Lines for Genetic Screens

A. Selection for Lethal/Sterile-free Background Lines

An important characteristic desired in a genetic background is the absence of preexisting mutations, whether zygotic or parental. In any kind of screen, whether based on gynogenetic techniques or natural crosses, the use of lines free of preexisting mutations is important for two reasons. First, lines free of lethal mutations diminish unwanted background lethality, which reduces brood sizes and can preclude the isolation of new mutations linked to the background mutation. In addition, the use of lines free of preexisiting mutations eliminates the possibility of isolating multiple copies of a mutant allele already present in the genetic background. Selecting for the absence of lethal or sterile mutations can be carried out in two ways.

1. Continuous Inbreeding

Wild-type stocks free of zygotic lethal and sterile mutations can be obtained by inbreeding individuals for two generations and essentially screening the F_2

generation for lethality and sterility phenotypes. In a stock maintained by mixed breeding of many individuals (to maintain genetic diversity and prevent unhealthy, highly inbred stocks) two generations of inbreeding of several pairs of fish can reduce the likelihood of background mutations being present in the parental generation. Individual pairs of wild-type fish are intercrossed and their respective F1 progeny raised in separate tanks. The F_1 progeny are then intercrossed and screened for zygotic lethal mutations in the F_2 generation. A reliable indicator of zygotic lethal mutations is the lack of swim bladder inflation at 5 dpf, in addition to obvious defects at earlier stages in 25% of the brood. By examining F_2 embryos from at least 12 intercrosses from one F_1 family, a >95% probability exists that a mutation will be detected, if it exists in that particular family.

If lethality is not observed in any of the 12 crosses, then the F_2 fish from the 12 individual F_1 intercrosses are raised in a separate tank and screened for late lethal mutations, as well as maternal and paternal effect, and female- and male-sterile mutations. The total number of F_2 adults is counted and compared to the number of larvae initially raised. If a late lethal mutation exists, then 25% of the larvae will not survive to adulthood. Several crosses can be raised between unrelated individuals to control for non-genetic lethality associated with normal fish raising. Only lethality significantly beyond that of normal fish husbandry is then considered as a potential late lethal mutation. From each F_2 family 12 males and females are intercrossed and their embryos examined. If the F_3 embryos are normal, then the probability is >95% that maternal- and paternal-effect and sterile mutations do not exist in those F_2 fish. If two such F_2 lines are established from different F_1 fish, then the males can be mutagenized and then interbred to females from another F_2 line in the parental generation of the screen to prevent further inbreeding.

2. Whole Genome Homozygosity Through HS-induced Gynogenesis

The gynogenetic method of HS induces homozygosity at every locus (see Section II.B.1) and is thus particularly effective at selecting, in one single generation, for fish that lack *any* background mutations. We grew a large number of HS-derived gynogenetic clutches from our substrate line gol-mix, a hybrid line with both AB and TU genetic backgrounds that is robust and whose females can be easily stripped of eggs, and selected for adult fish free of lethal or sterile mutations. From our starting gol-mix population, we previously generated two lines, golFL-1 and golFL-2, from four different HS-derived individuals. These two lines were combined to create golFL-3 (golFL-2 was 100% male and could not be propagated as a pure stock). Similar strategies had been used previously to select for such lethal/sterile-free strains, which can be further propagated through EP to generate clonal lines (Streisinger *et al*., 1981). Selection of lines through HS and EP can also lead to stocks of higher viability under gynogenetic conditions, presumably by the reduction of background detrimental alleles (Streisinger *et al*., 1981).

B. Specific Requirements for Lines in EP-based Screens

1. Selection of Lines Amenable to Gynogenetic Methods

The majority of lines we have examined, including lines recently derived from the wild, tend to produce low yields of fertile gynogenotes (Pelegri and Schulte-Merker, 1999). Selection of appropriate lines is therefore important for an efficient gynogenetic-based maternal screen.

2. Selection of Lines that Produce a High Yield of Gynogenotes

The experimental induction of gynogenesis relies on the manipulation of *in vitro* fertilized eggs at very early stages. Therefore, it is necessary that females should, as a first requirement, readily yield eggs when manually stripped. Different fish strains differ greatly in their ability to be manually stripped of eggs (Pelegri and Schulte-Merker, 1999). The capacity to be stripped of eggs is distinct from being fertile and able to mate successfully by standard "natural" crosses in the laboratory (Eaton and Farley, 1974; Pelegri and Schulte-Merker, 1999). This may be related to the fact that, under natural conditions, release of mature oocytes from their follicles into the ovarian lumen requires hormonal stimulation (Selman *et al*., 1994), which may normally be triggered by vigorous chasing by the males (Eaton and Farley, 1974). Lines that can be most easily stripped of eggs appear to be those which have been propagated by artificial fertilization methods, which also involve stripping of eggs, such as those derived from the AB Oregon line (Streisinger *et al*., 1981). In contrast, lines from the wild or laboratory lines that have been propagated mostly by natural crosses tend not to be easily stripped of eggs. We found that females from the gol-mix line produce a high yield of gynogenotes. Thus, we chose to continue our selections and genetic screen schemes with this starting population. In addition, the fact that this line is marked with the recessive pigment marker *golden* allows the detection of unwanted products of incompletely inactivated sperm (isolated from *golden* + males) after the EP procedure.

3. Selection for Favorable Sex Ratios Under Gynogenetic Conditions

Sex determination in fishes varies from species with sex-determining chromosomes to multi-factor autosomal ones, and in some cases sex has been shown to be influenced by external factors (reviewed in Chan and Yeung, 1983). The mechanism of sex determination in zebrafish, although poorly understood, appears to fall into the latter category. Zebrafish lack a single sex chromosome and sex determination is sensitive to growing conditions. Most gynogenetic clutches, after grown to adulthood, exhibit sex ratios that are strongly biased toward males (86–88% males; see Fig. 4 in Pelegri and Schulte-Merker, 1999). The phenomenon of sex bias in gynogenetic clutches is likely related to the tendency of zebrafish and other teleosts to develop into males under suboptimal conditions, for example, in overcrowded conditions or in subviable genetic backgrounds (see Chan and Yeung, 1983; our

observations). Presumably, gynogenetic clutches, because of their high degree of inbreeding, also have a suboptimal genetic background that under normal circumstances produces males. Nevertheless, a small fraction of the gynogenetic clutches (5–10%) are composed of at least 50% females (Pelegri and Schulte-Merker, 1999).

The observation of rare gynogenetic clutches with a high female to male ratio indicates that it is possible to select for genetic backgrounds which produce a high proportion of females even under gynogenetic conditions. In fact, the gynogenetic procedure itself may act as a selection for female-rich genetic backgrounds, as exemplified by the fact that one out of two lethal/sterile-free lines that we derived from HS-derived adult gynogenotes consists of mostly females (92% females; Pelegri and Schulte-Merker, 1999). In this line, golFL-1, there exists a small fraction of males that can mate with wild-type females but tend to produce unfertilized eggs. Nevertheless, treatment of this line with testosterone for the first 14 dpf leads to the production of larger percentages of fertile males (72% males), and thus allows the production of males both for mutagenesis and for the propagation of the line through natural crosses.

C. A Hybrid/inbred Approach

Selection of lines can increase the frequency of certain desired traits and also lead to inbreeding, which often causes a reduction in overall robustness and fertility (Thorgard, 1983). Thus, the best lines for gynogenetic-based maternal screens might be hybrids between gynogenetically selected lethal/sterile-free lines. This approach would be essential for genetic screens that incorporate a simultaneous mapping strategy. In this case, both polymorphic lines can be selected independently for the characteristics desired in the screen.

IV. Recovery and Maintenance of Adult and Parental-effect Mutations

Once individuals are identified as exhibiting an adult and/or parental-effect phenotype, the mutation needs to be recovered. In the case of viable mutations (e.g., affecting pigmentation or adult morphology but not viability or fertility), recovery and maintenance of the mutation is straightforward, because new stocks can be initiated from the affected individuals and the mutation can be maintained as a visible marker or even a homozygous stock. Here we present strategies to recover and maintain mutations that are either not viable or not fertile when homozygous. In this case, the observed adult, infertile or parental-effect phenotypes are expected to be caused by homozygosity for recessive mutations, because dominant mutations are unlikely to be propagated through generations that occur prior to screening. Because homozygosity for recessive lethal, infertile, or parental-effect mutations leads to the

inability of an affected individual to reproduce, a genetic scheme has to allow the recovery of the mutations through genetically related individuals.

A. General Methods for Recovering Mutations

A mutation can be recovered by three general means: through known heterozygous carriers, through siblings of homozygous mutant individuals, and through rare survivor progeny derived from homozygous mutant individuals.

1. Recovery Through Known Heterozygous Carriers

Individuals that produce homozygous affected adults (F_2 parents in an F_3 adult/F_4 parental-effect natural crosses screen and F_1 individuals in a EP F_2 adult/F_3 parental effects screen) are heterozygous carriers of the mutations. Thus, such fish are stored separately until their progeny reach adulthood and are tested for potential phenotypes. After the 3–5 months that are required to grow up and test their progeny, we find that the majority of the separated individual fish are still alive and fertile, and thus can be used to recover the mutation.

2. Recovery Through Siblings of Homozygous Mutant Individuals

In some instances a known carrier for a mutation is not available or is not fertile. In these cases, maternal-effect mutations can be recovered by performing crosses between siblings of homozygous mutant individuals (F_3 siblings in an F_3 adult/F_4 parental-effect natural crosses screen and F_2 siblings in an F_2 adult/ F_3 parental-effect EP-based screen), a fraction of which are carriers of the mutation. If the mutation is mapped (see Section V.A), siblings can be selected that are either homozygous or heterozygous for the mutation by genotyping.

If the mutation has not yet been mapped, propagation is ensured by raising progeny from multiple sibling intercrosses. In an F_3-extended family strategy (where families are composed of siblings and cousins), ~50% of the sibling males and females are heterozygous carriers and 6.25% of the males and females are homozygous carriers. Thus, 25% of F_3 sibling/cousin intercrosses are between heterozygotes of the mutation and will yield F_4 homozygous mutant individuals. If the F_3 (or subsequent) generation is made between two heterozygous carriers, then 75% of all F_3 males and females are carriers (50% heterozygous and 25% homozygous for the mutation). In this case, 50% of the intercrosses should yield F_4 homozygous mutants.

In EP-based screens, it is preferable to generate outcrosses, rather than incrosses, for the recovery of mutations. This is because EP-derived fish do not mate as efficiently as wild-type fish. In addition, the background used in an EP-screen results in female-rich tanks, and such abnormal ratios interfere with subsequent propagation of the mutation. We have found that outcrossing to a line such as *leopard long fin* (also known as TL), which tends to have a slight bias toward maleness, results in

hybrid stocks that have normal sex ratios in subsequent generations. Outcrossing also improves the general robustness and fertility of the line.

In the case of maternal-effect mutations isolated in EP-based screens, outcrossing is preferably carried out through sibling F_2 males, rather than sibling females, because females homozygous for a maternal-effect mutation are sterile, whereas homozygous males should be fertile unless the mutated gene also affects male fertility. This is particularly true in the case of centromere-linked loci (F_m toward 0.5), when most *fertile* females are expected to be homozygous for the wild-type allele (see Section II.B.1 and below) and therefore cannot transmit the mutation. The reverse logic applies to paternal-effect mutations.

In EP-derived clutches, the frequency of heterozygotes and homozygotes for a given mutation varies depending on the centromere–locus distance. For centromere-linked loci (F_m close to 0.5), 50% of the siblings are homozygous for the mutation. As the centromere–locus distance increases, the fraction of homozygous siblings (F_m) decreases but the fraction of heterozygous siblings increases two times as rapidly. For a distal mutation of $F_m = 0.05$, for example, 5% of the F_2 siblings are homozygous mutant, and 90% are heterozygous carriers. Thus, the overall frequency of F_2 carrier siblings (homozygous or heterozygous) varies from 50% for centromere-linked loci to percentages approaching 100% for distal loci. Therefore, the recovery of mutations through F_2 siblings can also be an efficient strategy in EP screens. In large F_2 EP-derived clutches, F_m, and therefore the fraction of siblings that are heterozygous or homozygous carriers for the mutation, can be estimated by the proportion of F_2 individuals that exhibit the adult or parental-effect mutant phenotype.

3. Recovery Through Rare Survivors

In cases of mutations that are not 100% lethal or sterile, the mutation may also be recovered by crossing rare homozygous mutant survivors of a generally lethal larval or juvenile defect (F_3 individuals in a natural crosses screen and F_2 individuals in an EP screen), or by raising and crossing rare surviving progeny of a parental-effect or sterile mutant adult (F_4 clutches in a natural crosses screen and F_3 clutches in an EP screen). The presence of such "escapers" may be due to variability in the phenotype caused by residual function of a hypomorphic allele or some degree of redundancy in the affected pathways. Progeny from such subviable/subfertile individuals are expected to be heterozygous carriers for the mutation, and the mutation can be propagated by incrossing fish derived from them.

Of the above options, the schemes in (1) and (3) are the most efficient because they use individuals that are known carriers of the mutation.

B. Maintenance of Adult and Parental-effect Mutations

Whether a mutation has been mapped or not, it is tempting to maintain stocks carrying it either by repeated inbreeding or through "escaper" embryos (see Section

IV.A.3). However, repeated inbreeding eventually generates inbred stocks that are weak and have aberrant sex ratios that typically lead toward maleness (see Section III.B.2), thus interfering with future stock propagation and, in the case of maternal-effect mutations, the identification of homozygous mutant females. Moreover, maintenance of the mutation by repeated propagation through "escaper" embryos may select for genetic constellations that gradually weaken the mutant phenotype. Lastly, in an extended family, natural crosses approach, the high ENU dose used leads to induction of multiple lethal mutations. Such mutations are typically unlinked to the adult/parental-effect mutation but can nevertheless reduce the size of intercross families and are best crossed out of the mutant background. To address these issues, adult and parental-effect mutations can be routinely propagated through cycles of crosses to a wild-type stock (outcrosses), followed by crosses between siblings (incrosses). Typically, an outcross that is known to carry a mutation can be kept for a period of time and additional incrosses performed from the outcross fish to produce new families containing homozygous individuals. It works well to perform a cycle of one outcross, which can generate several incrosses over a period of a year or more, and then initiate a new cycle by carrying out an outcross from one of the more recent incrosses.

1. Maintenance of Recessive Lethal Adult or Parental-effect Mapped Mutations

If an adult mutation is mapped to a chromosomal position, genotyping is used to identify heterozygous carrier individuals. To propagate the mutation, such heterozygous carriers can be outcrossed to wild-type individuals carrying alleles, for example SSLP markers, flanking the wild-type allele that are polymorphic to those linked to the mutation. The resulting generation can then be incrossed to produce a new generation that carries homozygous individuals, and this cycle can be repeated. Multiple generations of incrossing can be performed prior to performing a new outcross generation.

In the case of female-sterile or maternal-effect genes, males homozygous for the mutation are selected from siblings of homozygous mutant females (25% of the male siblings are expected to be homozygotes). Outcrosses are then initiated from such homozygous males to wild-type females carrying DNA markers flanking the wild-type allele that are polymorphic to those of the mutant allele in the homozygous male. The progeny of this cross are all heterozygous carriers and can be interbred to produce a family that contains 25% homozygous mutant females. The reverse strategy can be followed (propagation through homozygous sibling females) to identify sibling carriers and propagate paternal-effect mutations. These strategies allow the unambiguous identification of heterozygous and homozygous carriers through the use of polymorphic markers flanking the mutation.

2. Maintenance of Recessive Lethal Adult or Parental-effect Unmapped Mutations

If the mutations are not mapped to a chromosomal location, a similar strategy is followed except through multiple, random crosses. In the case of adult phenotypes,

multiple crosses can be initiated through viable, fertile siblings, 66% of which are expected to be heterozygous carriers of the mutation (the expected mendelian frequency of 75% carriers is modified in this case because homozygous mutant individuals are inviable or not fertile). If the mutations are female-sterile or have a maternal effect, and similar to when recovering these mutations (see Section IV.A.2), it is more efficient to initiate the crosses through sibling males, 75% of which are expected to be either homozygous or heterozygous carriers for the mutation. The reverse logic would apply to male-sterile and paternal-effect mutations.

Multiple crosses ensure the propagation of the mutation. For example, in the case of a maternal-effect mutation, 75% of outcrosses derived from males that are siblings of homozygous mutant females consist of families of carrier individuals. Thus, five random outcrosses from sibling males of homozygous mutant females ensure a 99.9% probability of transmission of the allele to at least one of the outcrosses. Within such outcrossed-derived families, the percentage of heterozygotes is expected to be 100% or 50%, depending, respectively, on whether the original outcrossed male is homozygous or heterozygous for the mutation. Multiple incrosses from such families allow the recovery of homozygous mutant females in the next generation. For example, eight incrosses from a tank that consists of 50% heterozygous carriers lead to a 90% probability of finding homozygous females (at a 25% frequency) in at least one of the incrossed families.

A variation of these approaches is to outcross identified heterozygous or homozygous carrier individuals. These individuals are identified as carriers by interbreeding them with siblings, raising the progeny, and determining if their offspring yields mutant individuals. Parental individuals yielding progeny with mutant adult/parental-effect phenotypes are then outcrossed. These outcross progeny are then inbred to produce a new generation of homozygous mutant individuals, as discussed above. In this modified approach, two blind generations of intercrosses are avoided by first identifying the individual carriers prior to outcrossing them.

During the maintenance of mutations, individual carriers should be outcrossed to fish of the same strain. This avoids increasing the degree of polymorphism in the carrier line, which in turn facilitates the subsequent process of mapping (see Section V).

V. Mapping Adult and Parental-effect Mutations

Mapping a mutation to a chromosomal position can be carried out either simultaneously with an F_3 adult/F_4 parental-effect screen using natural crosses (see Section II.A), or at any time after the identification and recovery of the mutation. Specific details on mapping protocols have been previously described (Geisler, 2002; Talbot and Schier, 1999). Here we describe the modification of this approach for mapping of adult or parental-effect mutations. Briefly, the approach consists of outcrossing a carrier of the mutation to a polymorphic wild-type stock to yield F_1 hybrid families. Incrosses from the F_1 family in turn allow the production of F_2

adults, which can be tested for homozygosity of the mutation and analyzed for linkage to DNA markers throughout the genome.

A. Mapping Concomitant with F_3 Adult/F_4 Parental-effect Genetic Screens

A mapping cross can be integrated into an extended-family natural crosses screen strategy (see also Dosch *et al*., 2004). Two strains that are polymorphic to each other, e.g., TU and AB, are mutagenized. The mutagenized males are crossed with females of their respective strain to produce an F_1 generation. F_1 fish are then interbred between the two strains to make a hybrid F_2 generation. F_2 fish are intercrossed to make the mutations homozygous in the F_3 generation and F_3 mutant adults can be used to map the mutation. The F_1 grandparent DNA is crucial in examining linkage using bulk segregant analysis. Thus, the F_1 fish are frozen and kept for mapping purposes, should a mutation be found that one wants to map.

Intercrossing strains to make a map cross gives rise to very robust stocks, so-called hybrid vigor. As a consequence of interbreeding F_1 fish of different strains, we find the F_2 hybrid generation to be particularly healthy and prolific, with increased reproductive longevity, compared to either independent strain. This is advantageous in regenerating the identified mutations to produce additional individuals for mapping (see below). However, a drawback is that hybrid vigor leads to an increased propensity to produce females in the F_2 hybrid generation. Thus, typically 10–20% of F_2 families yield three or fewer males. We assess the sex ratio at about 2 months of age, and discard those with fewer than four males. In the F_3 generation, the sex ratio is not distorted and we rarely find such sex biased families. It is possible that future lines could be developed that do not exhibit the sex bias in the F_2 hybrid generation.

As with zygotic mutations (Geisler, 2002; Talbot and Schier, 1999), bulk segregant analysis is used to map adult and parental-effect mutations. However, it is considerably more effort to generate adult and, even more effort, parental-effect mutant individuals than embryos mutant for zygotic mutations as used in the original bulk-segregation protocol. We have found that for maternal-effect mutations, 14 mutant females are sufficient to map a mutation efficiently. In using an extended family screen (see Section II.A), mutant females represent 1/16 of the total, producing 1–5 mutant females in a tank of 60 fish. Even five females are insufficient to map a mutation efficiently. Thus in performing such a screen, we keep up to ten pairs of fish of the F_2 generation in a small 2-L tank. If we are interested in a mutation, we return to the F_2 fish and regenerate the mutant through F_2 pair-wise crosses. Each F_3 family is raised separately and the F_2 fish stored individually. Regenerated F_3 families are screened for mutant females (as described in Section IV.B.2); the F_2 parents of those that yield mutants can then be used to produce more mutant females. The F_3 mutant and non-mutant sibling females are then used for mapping the mutation. We routinely use as few as 14 mutant females to map a mutation, and have mapped mutations with just nine mutant individuals. In the latter cases, considerably more false positive linkages are detected in bulk segregant analysis. We have reliably regenerated >50 maternal-effect mutants and efficiently mapped most

of the mutations using this strategy (Bontems *et al*., 2009; Dosch *et al*., 2004; Gupta *et al*., 2010; Holloway *et al*., 2009; Marlow and Mullins, 2008; Mei *et al*., 2009; Wagner *et al*., 2004; Yabe *et al*., 2007, 2009; M.C.M., unpublished).

B. Mapping After Identification and Recovery of Mutations

In the case of viable and fertile adult mutations, homozygous mutant individuals can be directly identified by their phenotype, and can be used to generate mapping crosses by crossing to a polymorphic strain such as WIK. Incrossing of the resulting hybrids leads to a generation with homozygous mutants, which can be used for bulk segregant analysis to find linkage as described above.

In the case of lethal/sterile adult mutations or parental-effect mutations, genetic mapping can be initiated by a process similar to that carried out for the maintenance of unmapped mutations (see Section IV.B.2). Multiple (P) siblings of affected (presumed homozygous) individuals are outcrossed to a polymorphic strain. F_1 hybrid individuals from such outcrosses are incrossed at random to generate F_2 crosses, a fraction of which will yield homozygous mutants.

For adult lethal mutations, 66% of (P) siblings are expected to carry the mutation, and these can be used to initiate the mapping crosses. In the case of female-sterile and maternal-effect mutations, it is most efficient to initiate mapping crosses from homozygous mutant (P) sibling males, because 75% of males are expected to be carriers compared to 66% of sibling females. Among these males, 25% are expected to be homozygous for the mutation. Outcrosses of homozygotes yield F_1 families where all individuals are heterozygous carriers for the mutation, so that all random F_2 incrosses contain homozygous mutant females. Outcrosses of heterozygous carrier males, expected at a 50% frequency amongst the siblings of homozygous mutant females, produce F_1 families where 50% of individuals are heterozygous carriers, so in this case only 25% of random F_2 incrosses yield homozygous mutant F_2 females. Thus, it is more efficient to initiate the mapping strategy using homozygous males, because a much larger fraction of incrosses from the F_1 hybrids (100% compared to 25% if using heterozygous males) yield homozygous F_2 females. Parental males can be identified in advance as homozygotes through genetic crosses, and then the mapping cross initiated with such a male. Alternatively, the mapping strategy can be initiated with 8 sibling males of undetermined genotype, which are outcrossed to polymorphic females. This relatively large number of outcrosses increases the probability that at least one of the outcrosses originates from a homozygous male. After the F_1 hybrid fish are incrossed to make an F_2 generation, they are kept separately. Multiple incrosses from F_1 hybrids and testing of F_2 females for maternal effects allows inferring whether the original P male was homozygous for the mutation, if it was not identified in advance, and which pairs of F_1 hybrids are heterozygous carriers. Such pairs, now identified, can be crossed repeatedly to produce more mutant F_2 females for mapping. This general strategy is reversed in the case of male-sterile or paternal-effect mutations.

F_2 individuals from mapping crosses are separated into two phenotypic classes according to the specific adult/parental-effect assay: phenotypically mutant

individuals (i.e., individuals homozygous for the mutation) and phenotypically wild-type (i.e., siblings that are either heterozygous or homozygous for the wild-type allele). Identified mutant individuals may be retested with the same assay or a different, secondary assay to ensure that they produce the expected phenotype and to check for potential errors in handling during phenotypic identification.

After their classification into phenotypic classes, F_2 individuals are anesthesized, their tail fin clipped, and the tail fin DNA is isolated (the remaining part of the body is also frozen and serves as a backup in case additional DNA needs to be isolated). Individual tail fin DNA from 20 individuals of each phenotypic class is used to make two DNA pools. These pools are used to carry out a first-pass mapping (see Section II.A; Geisler, 2002; Talbot and Schier, 1999) to identify linkage. Once linkage has been found, DNA from single fish is analyzed separately with respect to polymorphisms to markers within the linked region (Geisler, 2002; Talbot and Schier, 1999).

C. Efficient Fine Mapping of Maternal-effect Mutations

Narrowing down the location of a mutation through fine mapping is necessary to identify the molecular nature of the mutated gene through either candidate gene or positional cloning approaches. For viable adult and maternal-effect mutations, a different strategy is used for fine mapping, which can be performed much more efficiently than the initial mapping. Once a viable adult or maternal-effect mutation is mapped, then homozygous male and heterozygous female carriers can be identified by polymorphic markers flanking the mutation. Such F_0 fish are intercrossed to map finely the position of the mutation in the F_1 progeny. In a cross of a heterozygous female to a homozygous male, mapping will be performed with recombinants generated only through meiosis in the F_0 female, because the male is homozygous for the mutation. Thus, each F_1 fish represents a single meiosis, rather than two meioses for progeny from heterozygote intercrosses. However, the loss of recombination events from the homozygous male is offset by the fact that meiotic recombination is suppressed in males compared to females, so that the vast majority of all recombinants generated in intercrosses of heterozygotes are from female, not male meioses (Singer *et al*., 2002; S.L. Johnson, personal communication; MCM unpublished). Thus, little is lost in crossing heterozygous females with homozygous males, and considerable gain is achieved using this strategy, as discussed below. Because this strategy depends on crosses from one homozygous mutant adult, this approach is not feasible for recessive mutations that result in adult lethality, or sterility in both sexes. Paternal effect and male-sterile mutations are also not effectively mapped by this strategy, given the reduced rate of recombination in male meioses.

Fine mapping of an adult viable or maternal-effect mutation through crosses between heterozygous females and homozygous males is similar to haploid mapping of zygotic mutations (Postlethwait and Talbot, 1997). In both cases, all fish can be examined for recombinants. For such mutations, both the phenotypically mutant and wild-type F_1 adults are examined for recombination between, respectively, the mutation and a wild-type-linked flanking marker or the wild-type allele of the gene

and a mutant-linked flanking marker. Thus, all the progeny (female progeny in the case of maternal-effect mutations) from a cross between a heterozygous female and a homozygous male are informative, in contrast to only 1/4 of the progeny in intercrosses of heterozygotes. This strategy saves considerable effort and tank space, which is important because it is significantly more effort to generate adult individuals to map adult or maternal-effect mutations, compared to generating embryos or young larvae to map finely zygotic mutations.

Depending on the complexity of the assay, fine mapping can be made even more efficient in a map cross between a heterozygous female and homozygous male by genotyping all F_1 progeny, rather than phenotyping them. This is the case for maternal-effect mutations, because this strategy avoids having to generate crosses to examine their progeny except for those rare cases where an informational recombination event has occurred. In this case, all F_1 females (and males, see below) are genotyped with the closest polymorphic markers flanking the mutation to determine whether they are non-recombinant mutants or heterozygotes, or recombinants within the interval of the flanking markers. Only the small subset of recombinants is phenotyped to determine whether they are mutant or wild-type females and thereby establish where the recombination occurred relative to the mutation. As the critical interval is narrowed and closer polymorphic markers are defined, fewer recombinants are identified and consequently fewer females phenotyped in crosses. We typically genotype individual fish at 2 months of age and maintain only the small fraction of recombinants until breeding age to determine their phenotype and the position of the recombination (the recombination "break point") relative to the mutation. The total number of F1 females genotyped is compared to the number of recombinants to determine the genetic distance between the mutation and the flanking markers.

Once the interval of the maternal-effect mutation is narrowed to a $\sim$0.5 centiMorgan region, we also find it worthwhile genotyping F_1 males from the intercross of the heterozygous female and homozygous male for recombination within the critical interval. The disadvantage of males is that a test cross must be performed between the recombinant male and a heterozygous female to determine whether the male is homozygous or heterozygous for the mutation. Female progeny carrying the recombinant male chromosome of this cross over a mutant chromosome must be tested to determine whether they exhibit a mutant or wild-type phenotype to assess the genotype of the recombinant male chromosome. Although this is considerable effort, identifying a recombinant that narrows down the interval can be so valuable that we have found it worth the effort.

VI. Solutions, Materials, and Protocols

A. Solutions

- MESAB stock solution: 0.2% ethyl-m-aminobenzoate methanesulfonate. Adjust to pH 7.0 with 1 M Tris pH 9.0. Keep at 4 °C.
- MESAB working solution: 7 mL stock solution per 100 mL fish water.

- Hank's Solutions: stock solutions 1, 2, 4, and 5 and premix can be stored at 4 °C. Stock Solution #6 is prepared fresh and added to the premix to form the final Hank's solution.

  Solution #1: 8.0 g NaCl, 0.4 g KCl in 100 mL double distilled (dd) H_2O.
  Solution #2: 0.358 g Na_2HPO_4 anhydrous, 0.60 g KH_2PO_4 in 100 mL ddH_2O.
  Solution #4: 0.72 g $CaCl_2$ in 50 mL ddH_2O.
  Solution #5: 1.23 g $MgSO_4.7H_2O$ in 50 mL ddH_2O.

Hank's Premix:	combine the following in order: 10.0 mL solution #1 1.0 mL solution #2 1.0 mL solution #4 86.0 mL ddH_2O 1.0 mL solution #5

Solution #6 (prepare fresh): 0.35 g $NaHCO_3$ in 10 mL ddH_2O

Hank's (Final):	990 μL Hank's premix 10 μL solution #6

- E2 saline (used specially during the testosterone treatment because of its higher buffering properties): 15 mM NaCl, 0.5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.15 mM KH_2PO4, 0.05 mM Na_2HPO_4, 0.7 mM $NaHCO_3$.
- E3 saline (a simpler version of E2 used for routine embryo raising): 5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 10^{-5}% methylene blue.
- Testosterone stock: 150 mg testosterone in 50 mL absolute ethanol. Store in aliquots at −20 °C.
- Testosterone working solution: while stirring, add 10 μL of stock solution per 600 mL of (a) E2 saline, for babies before d6, or (b) fish water supplemented with 3 g/L Red Sea salt (Read Sea Fish pHarm, Israel), for larvae between d6 and d15. In our hands E2 (instead of E3) and Red Sea salt in the fish water improve the survival of testosterone-treated larvae. Stir 10 min.

B. Other Materials

- UV lamp: Sylvania 18 in. 15 W germicidal lamp.
- French Press Cell, 40 mL (SLM-Aminco)
- French Pressure Cell Press (SLM-Aminco) or Hydraulic Laboratory Press (Fisher)
- Heat Shock baskets: these can be made by cutting off the bottom of Beckman Ultraclear centrifuge tubes, and heat-sealing a fine wire mesh to the bottom edge of the tube.

- EP vials: disposable glass scintillation vials, with plastic caps (3.2 cm height and 2.2 cm diameter, Wheaton) or similar vials. The plastic caps are perforated several times with a needle to better allow exposure to the hydrostatic pressure. Only two vials of these types can fit at once in a pressure cell. To fit four vials in one cell, we have custom-built shorter plastic vials (1.8 cm height, including cap, 2.5 cm diameter, 0.3 mm wall thickness) which fit the plastic caps from the scintillation vials.

C. Protocols

1. Sperm Collection (Adapted from D. Ransom)

- A sperm solution can be made with testes dissected from ten males for each 1 mL of Hank's solution. Keep the isolated testes and Hank's solution on ice. Shear the testes with a small spatula and by pipetting up and down with a 1000 μL Pipetteman. Allow debris to settle and transfer supernatant to a new tube. Sperm solution on ice is effective for about 2 h. Sperm still remaining inside the sheared, settled testes can be further collected by adding 300 μL of fresh Hank's and letting the mixture rest for 1/2 h or longer. For more details see Ransom and Zon, 1999.

2. UV-inactivation of Sperm

- Transfer the sperm solution to a watch glass. Avoid pieces of debris, because they may shield sperm from the UV light. Place the watchglass on ice at a distance of 38 cm (15 in.) directly under the UV lamp. Irradiate for 2 min with gentle stirring every 30 s. Transfer to a new Eppendorf tube with a clean pipette tip. UV-treated sperm solution on ice is effective for about 2 h.

3. Stripping of Eggs

- Our observations suggest that females are more amenable to manual stripping if removed from their tank and placed in a clean tank (1–10 females per 2-L tank) the evening before stripping. Best stripping and egg clutch quality are obtained during the first 4 h after the start of the light cycle of the first day after the separation of the females. The presence of males together with the separated females does not significantly affect the ability of gol-mix females to be stripped (our observations), although it may have an effect when working with other fish lines (Eaton and Farley, 1974).
- Anesthetize females in MESAB working solution until they reduce their gill movements (2–4 min, MESAB solution may need to be boosted through time with more stock solution in 0.5–1 ml increments). Overexposure to MESAB will impede the recovery of the female, and fish should be placed in fresh water if they are not going to be used within 1 or 2 min after they stop their movements.
- With the aid of a spoon, rinse a female in fish water and place her on several paper towels to remove excess moisture.

- Place the female on the bottom half of a Petri plate. With a soft tissue, dry further the anal fin area. Excess water may prematurely activate the eggs.
- Slightly moisten the index fingers of both hands (dry hands will stick to the skin of the fish). With one finger support the back of the female, and with the other gently press her belly. Females which can be stripped will release their eggs upon gentle pressure. Healthy eggs have a translucent, yellowish appearance. Separate the eggs from the female with a small, dry spatula. Females can be placed separately in boxes and identifying tags can be attached to the box with the female and the corresponding egg clutch. If necessary, clutches can wait for several minutes before being activated. In this case, we cover the clutches with the Petri plate lid to reduce drying of the clutch. Fertilization can occur after even longer delays (in our hands, up to 6 min), although not in a consistent manner. Egg activation can be delayed for periods of 1.5 h or more with ovarian fluid from the rainbow trout or coho salmon (Corley-Smith *et al*., 1995), or with Hank's saline buffer supplemented with 0.5% BSA (Sakai *et al*., 1997), although we have not tested these methods in combination with gynogenesis.

4. *In vitro* Fertilization

- Add 25 μL of untreated or UV-irradiated sperm to the egg clutch. Mix the sperm and eggs by moving the pipette tip without lifting it from the Petri plate (to minimize damage to the eggs). If desired, proceed at this point to Heat Shock or Early Pressure Protocols. If not, add 1 mL of E3 saline to activate the eggs, and, after 1 min, fill the Petri plate with E3. Incubate at 27–29 °C.

5. Heat Shock

- After IVF with UV-treated sperm, add 1 mL of E3 saline to activate the eggs and start the timer.
- Add more E3 after 30 s. Transfer the eggs to a Heat Shock basket. Immerse the basket in a water bath with stirring and E3 saline at 28.5 °C.
- At 13.0 min, blot briefly the bottom of the basket onto a stack of paper towels and transfer the basket to a water bath with stirring and E3 saline at 41.4 °C.
- At 15.0 min, blot briefly the bottom of the basket and transfer the basket back to the 28.5 °C E3 bath.
- Allow the embryos to rest for about 45 min and transfer to a Petri plate. Allow embryos to develop in a 27–29 °C incubator (see note under "Early Pressure").

6. Early Pressure

To maximize the number of clutches produced, we work on cycles in which we include up to four clutches in separate vials within the Pressure Cell. For this, we typically anesthetize 6–12 females. Once four healthy-looking clutches are obtained,

the females that have not yet been stripped of eggs are transferred to fresh fish water until they completely recuperate. It works well to begin to anesthetize females for the next EP cycle at around minute four within a current cycle.

- After mixing eggs with UV-treated sperm (see *in vitro* fertilization), activate up to four clutches simultaneously by adding 1 mL of E3 saline to each clutch and start the timer (at least two people are required to timely manipulate four clutches).
- After 12 s, add more E3. A squirt to the side of the Petri plate will make the fertilized eggs collect in the middle of the plate.
- With a plastic pipette, transfer the fertilized eggs to an EP vial. Fill the vial with E3 and cap it with the perforated plastic lid. Avoid large air bubbles. Place the vials inside the pressure cell, ensuring that no air remains trapped inside it. Record the relative position of the clutch within the pressure cell by placing the tags in the corresponding order on a dry surface. Fill the pressure cell with E3 and close it allowing excess E3 to be released from the side valve. Close the side valve without overtightening. Insert entire assembly on the French Press apparatus and apply pressure to 8000 lb/sp. in. by time 1 min 20 s after activation. For different strains and/or presses, different pressure values may be optimal (see Gestl *et al*., 1997).
- At 6.0 min, release the pressure and remove the pressure cell from the French Press apparatus. Maintaining the relative order of the vials, remove the vials from the pressure cell, dry them with a towel, and label them with their corresponding number tags. Place the vial in a 27–29 °C incubator.
- After all EP cycles have been completed, allow the embryos to rest in the vial for at least 45 min but not more than 4 h. Transfer embryos with their corresponding tags to Petri plates. Let embryos develop in a 27–29 °C incubator.

Note: Due to the large amount of embryonic lethality induced by the HS and EP procedures, we incubate the embryos at a low density of 80 embryos maximum per 94 mm Petri plate (this is particularly important for the first 24 h of development).

7. Testosterone Treatment

- Before embryos reach 24 hpf, remove the chorions from the embryos, remove as much E3 as possible, and replace with testosterone/E2 working solution.
- Each consecutive day, replace 1/2 of the testosterone/E2 with fresh testosterone/E2.
- On day 6, transfer the embryos to mouse cages with 1 L of testosterone solution in fish water supplemented with 3 g/L Coral Reef Salt. Start feeding as normally. Continue replacing 1/2 of the solution every day by carefully aspirating the solution and refilling with fresh testosterone solution.
- On day 15, remove the testosterone by aspirating most of the solution and refilling with fresh fish water. Rinse again by repeating this procedure. Embryos can now be connected to the water system.

VII. Conclusions

We describe methodologies to identify, recover, maintain, and map adult and parental-effect mutations. Two main genetic screening strategies are described: an F_3/F_4 extended-family screen based solely on natural crosses and an F_2/F_3 EP-based screen. Each of these strategies has advantages and disadvantages. Extended-family screens based solely on natural crosses are technically relatively simple, and allow simultaneous mapping as well as the identification of maternal-zygotic mutations. However, such an approach requires larger amounts of space, generation time, and labor. On the other hand, EP screens require substantial selection of specialized lines amenable to the procedure but can be carried out using less generation time and space and are more amenable after *in vitro* fertilization to the observation of events immediately after fertilization. Both of these methods, however, have allowed the unbiased identification of many maternal-effect mutants (Dosch *et al*., 2004; Pelegri and Schulte-Merker, 1999; Pelegri *et al*., 2004; Wagner *et al*., 2004) and can be used for the identification of mutations in juvenile and adult traits (Andreeva *et al*., 2011; Bauer and Goetz, 2001; Haffter *et al*., 1996; Harris *et al*., 2008; Johnson *et al*., 1995; Johnson *et al*., 1995; Lang *et al*., 2009; Parichy and Turner, 2003; Rohner *et al*., 2009; Tschopp *et al*., 2010; Wagner *et al*., 2004). These efforts continue to pave the way toward the genetic analysis of early development in this vertebrate species and form the basis for the future genetic dissection of adult traits of biomedical relevance.

Acknowledgments

F.P. is grateful to Stefan Schulte-Merker, Marcus Dekens, Hans-Martin Maischein, and Catrin Weiler for their participation in the original EP screen, as well as Christiane Nüsslein-Volhard for support and advice at the Max-Planck Institut für Entwicklungsbiologie, Tübingen. Current support in the laboratory of F.P at the University of Wisconsin – Madison is provided by an NIH grant RO1 GM65303. M.C.M. thanks Daniel Wagner, Florence Marlow, Elliott Abrams, Hong Zhang, Tripti Gupta, Lee Kapp, Roland Dosch, and Keith Mintzer for their contributions in performing the F4 natural crosses screen described here and numerous discussions; Florence Marlow for contributing Fig. 2A, B; M.C.M. is supported by NIH grants R01 HD065600 and R21HD062952.

References

Abrams, E. W., and Mullins, M. C. (2009). Early zebrafish development: it's in the maternal genes. *Curr. Opin. Genet. Dev.* **19**, 396–403.

Andreeva, V., Connolly, M. H., Stewart-Swift, C., Fraher, D., Burt, J., Cardarelli, J., and Yelick, P. C. (2011). Identification of adult mineralized tissue zebrafish mutants. *Genesis* **49**, 360–366 Epub ahead of print.

Bauer, M. P., and Goetz, F. W. (2001). Isolation of gonadal mutations in adult zebrafish from a chemical mutagenesis screen. *Biol. Reprod.* **64**, 548–554.

Bontems, F., Stein, A., Marlow, F., Lyautey, J., Gupta, T., Mullins, M. C., and Dosch, R. (2009). Bucky ball organizes germ plasm assembly in zebrafish. *Curr. Biol.* **19**, 414–422.

Chan, S. T. H., and Yeung, W. S. B. (1983). Sex control and sex reversal in fish under natural conditions. *In* "Fish Physiology," (W. S. Hoar, D. J. Randall, and E. M. Donaldson, eds.), Vol IX, Part B, pp. 171–222. Academic Press, New York.

Corley-Smith, G. E., Lim, C. J., and Brandhorst, B. P. (1995). Delayed *in vitro* fertilization using coho salmon ovarian fluid. *In* "The zebrafish book," online version. University of Oregon Press, Eugene.

Dekens, M. P., Pelegri, F. J., Maischein, H. M., and Nüsslein-Volhard, C. (2003). The maternal-effect gene futile cycle is essential for pronuclear congression and mitotic spindle assembly in the zebrafish zygote. *Development* **130**, 3907–3916.

Dosch, R., Wagner, D. S., Mintzer, K. A., Runke, G., Wiemelt, A. P., and Mullins, M. C. (2004). Maternal control of vertebrate development before the Midblastula transition: mutants from the zebrafish I. *Dev. Cell* **6**, 771–780.

Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C. F., Malicki, J., Stemple, D. L., Stainier, D. Y. R., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., and Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46.

Eaton, R. C., and Farley, R. D. (1974). Spawning cycle and egg production of zebrafish, *Brachydanio rerio*, in the laboratory. *Copeia* **1**, 195–204.

Geisler, R. (2002). Mapping and cloning. *In* "Zebrafish," (C. Nüsslein-Volhard, and R. Dahm, eds.), Vol. 261, pp. 175–212. Oxford University Press, Oxford Practical Approach Series.

Gestl, E. E., Kauffman, E. J., Moore, J. L., and Cheng, K. C. (1997). New conditions for generation of gynogenetic half-tetrad embryos in the zebrafish (*Danio rerio*). *J. Heredity* **88**, 76–79.

Gupta, T., Marlow, F. L., Ferriola, D., Mackiewicz, K., Dapprich, J., Monos, D., and Mullins, M. C. (2010). Microtubule actin crosslinking factor 1 regulates the Balbiani body and animal-vegetal polarity of the zebrafish oocyte. *PLoS Genet.* **6**, e1001073.

Gupta, T., and Mullins, M. C. (2010). Dissection of organs from the adult zebrafish. *J. Vis. Exp.* **4,**(37), pii: 1717 doi: 10.3791/1717.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hamerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J. M., Jiang, Y. -J., Heisenberg, C. -P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., and Nüsslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36.

Harris, M. P., Rohner, N., Schwarz, H., Perathoner, S., Konstantinidis, P., and Nüsslein-Volhard, C. (2008). Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. *PLoS Genet.* **4**, e1000206.

Holloway, B. A., Gomez de la Torre Canny, S., Ye, Y., Slusarski, D. C., Freisinger, C. M., Dosch, R., Chou, M. M., Wagner, D. S., and Mullins, M. C. (2009). A novel role for MAPKAPK2 in morphogenesis during zebrafish development. *PLoS Genet.* **5**, e1000413.

Johnson, S. L., Africa, D., Walker, C., and Weston, J. A. (1995). Genetic control of adult pigment stripe development in zebrafish. *Dev. Biol.* **167**, 27–33.

Kane, D. A., and Kimmel, C. B. (1993). The zebrafish midblastula transition. *Development* **119**, 447–456.

Kemphues, K. J., Kusch, M., and Wolf, N. (1988). Maternal-effect lethal mutations on linkage group II of *Caenorhabditis elegans*. *Genetics* **120**, 977–986.

Kemphues, K. J., and Strome, S. (1997). Fertilization and establishment of polarity in the embryo. *In* "C. elegans II," (D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, eds.), pp. 335–359. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Kimmel, C., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). Stages of embryonic development in the zebrafish. *Dev. Dyn.* **203**, 253–310.

Kishimoto, Y., Koshida, S., Furutani-Seiki, M., and Kondoh, H. (2004). Zebrafish maternal-effect mutations causing cytokinesis defects without affecting mitosis or equatorial vasa deposition. *Mech. Dev.* **121**, 79–89.

Lang, M. R., Patterson, L. B., Gordon, T. N., Johnson, S. L., and Parichy, D. M. (2009). Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. *PLoS Genet.* **5**, e1000744.

Lehmann, R., and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in Drosophila. *Cell* **47**, 141–152.

Lindeman, R. E., and Pelegri, F. (2010). Vertebrate maternal-effect genes: insights into fertilization, early cleavage divisions, and germ cell determinant localization from studies in the zebrafish. *Mol. Reprod. Dev.* **77**, 299–313.

Lyman-Gingerich, J., Lindeman, R., Putiri, E., Stolzmann, K., and Pelegri, F. (2006). Analysis of axis induction mutant embryos reveals morphogenetic events associated with zebrafish yolk extension formation. *Dev. Dyn.* **235**, 2749–2760.

Lyman-Gingerich, J., Westfall, T. A., Slusarski, D. C., and Pelegri, F. (2005). *hecate*, a zebrafish maternal effect gene, affects dorsal organizer induction and intracellular calcium transient frequency. *Dev. Biol.* **286**, 427–439.

Marlow, F. L., and Mullins, M. C. (2008). Bucky ball functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. *Dev. Biol.* **321**, 40–50.

Mei, W., Lee, K. W., Marlow, F., Miller, A. L., and Mullins, M. C. (2009). *hnRNP*/*brom bones* is required for Ca -mediated egg activation in the zebrafish. *Development* **136**, 3007–3017.

Mullins, M. C., Hammerschmidt, M., Haffter, P., and Nüsslein-Volhard, C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* **4**, 189–202.

Neuhauss, S. (1996). Craniofacial development in zebrafish (*Danio rerio*): mutational analysis, genetic characterization, and genomic mapping. Ph.D. Thesis. Fakultät für Biologie, Eberhard-Karl-Universität Tübingen.

Newport, J., and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686.

Newport, J., and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696.

Parichy, D. M., Elizondo, M. R., Mills, M. G., Gordon, T. N., and Engeszer, R. E. (2009). Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev. Dyn.* **238**, 2975–3015.

Parichy, D. M., and Turner, J. M. (2003). Zebrafish *puma* mutant decouples pigment pattern and somatic metamorphosis. *Dev. Biol.* **256**, 242–257.

Pelegri, F., Dekens, M. P. S., Schulte-Merker, S., Maischein, H. -M., Weiler, C., and Nüsslein-Volhard, C. (2004). Identification of recessive maternal-effect mutations in the zebrafish using a gynogenesis-based method. *Dev. Dyn.* **231**, 324–335.

Pelegri, F., Knaut, H., Maischein, H. M., Schulte-Merker, S., and Nüsslein-Volhard, C. (1999). A mutation in the zebrafish maternal-effect gene nebel affects furrow formation and vasa RNA localization. *Curr. Biol.* **9**, 1431–1440.

Pelegri, F., and Schulte-Merker, S. (1999). A gynogenesis-based screen for maternal-effect genes in the zebrafish, *Danio rerio*. *In:* "The zebrafish: Genetics and Genomics," (Detrich, W., Zon, L.I., and Westerfield, M., eds.). *Meth. Cell Biol.* 60, pp. 1–20.

Poss, K. D., Nechiporuk, A., Stringer, K. F., Lee, C., and Keating, M. T. (2004). Germ cell aneuploidy in zebrafish with mutations in the mitotic checkpoint gene *mps1*. *Genes Dev.* **18**, 1527–1532.

Postlethwait, J. H., and Talbot, W. S. (1997). Zebrafish genomics: from mutants to genes. *Trends Genet.* **13**, 183–190.

Putiri, E., and Pelegri, F. (2008). Passing the torch at the oocyte to embryo transition. *Trends Dev. Biol.* **3**, 45–60.

Ransom, D.G., and Zon, L.I. (1999). Collection, storage, and use of zebrafish sperm. *In:* "The zebrafish: Genetics and Genomics," (Detrich, W., Zon, L.I., and Westerfield, M., eds.). *Meth. Cell Biol.* 60, pp. 365–372.

Rohner, N., Bercsényi, M., Orbán, L., Kolanczyk, M. E., Linke, D., Brand, M., Nüsslein-Volhard, C., and Harris, M. P. (2009). Duplication of fgfr1 permits Fgf signaling to serve as a target for selection during domestication. *Curr. Biol.* **19**, 1642–1647.

Sakai, N., Burgess, S., and Hopkins, N. (1997). Delayed *in vitro* fertilization of zebrafish eggs in Hank's saline containing bovine serum albumin. *Mol. Mar. Biotechnol.* **6**, 84–87.

Schnabel, R., and Priess, J. R. (1997). Specification of cell fates in the early embryo. *In* "C. elegans II," (D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, eds.), pp. 361–382. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Selman, K., Petrino, T. R., and Wallace, R. A. (1994). Experimental conditions for oocyte maturation in the zebrafish, *Brachydanio rerio*. *J. Exp. Zool.* **269**, 538–550.

Signoret, J., and Lefresne, J. (1971). Contribution a l'etude de la segmentation de l'oeuf d'axolotl: I. Definition de la transition blastuleenne. *Ann. Embryol. Morphog.* **4**, 113.

Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B., and Postlethwait, J. (2002). Sex-specific recombination rates in zebrafish (Danio rerio). *Genetics* **160**, 649–657.

Solnica-Krezel, L., Schier, A. F., and Driever, W. (1994). Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**, 1401–1420.

St. Johnston, D., and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**, 201–219.

Streisinger, G., Singer, F., Walker, C., Knauber, D., and Dower, N. (1986). Segregation analyses and gene-centromere distances in zebrafish. *Genetics* **112**, 311–319.

Streisinger, G., Walker, C., Dower, N., Knauber, D., and Singer, F. (1981). Production of clones of homozygous diploid zebra fish (*Brachydanio rerioI*). *Nature* **291**, 293–296.

Talbot, S., and Schier, A. F. (1999). Positional cloning of mutated zebrafish genes. *In* "The Zebrafish: Genetics and Genomics, Methods in Cell Biology," (H. W. Detrich, M. Westerfield, and L. I. Zon, eds.), Vol. 60, pp. 259–286. Academic Press, San Diego.

Thorgard, G. H. (1983). Chromosome set manipulation and sex control in fish. *In* "Fish Physiology," (W. S. Hoar, D. J. Randall, and E. M. Donaldson, eds.), Vol. IX, Part B,, pp. 405–434. Academic Press, New York.

Tschopp, M., Takamiya, M., Cerveny, K. L., Gestri, G., Biehlmaier, O., Wilson, S. W., Strähle, U., and Neuhauss, S. C. (2010). Fundoscopy in adult zebrafish and its application to isolate mutant strains with ocular defects. *PLoS ONE* **5**, e15427.

van Eeden, F. J. M., Granato, M., Odenthal, J., and Haffter, P. (1999). Developmental mutant screens in the zebrafish. *In* "The Zebrafish: Genetics and Genomics," (H. W. Detrich, M. Westerfield, and L. I. Zon, eds.), Vol. 60, pp. 21–41. Academic Press, San Diego Methods in Cell Biology.

Wagner, D. S., Dosch, R., Mintzer, K. A., Wiemelt, A. P., and Mullins, M. C. (2004). Maternal control of development at the Mid-Blastula transition and beyond: mutants from the Zebrafish II. *Dev. Cell* **6**, 781–790.

Wakimoto, B. T., Lindsley, D. L., and Herrera, C. (2004). Toward a comprehensive genetic analysis of male fertility in *Drosophila melanogaster*. *Genetics* **167**, 207–216.

Yabe, T., Ge, X., and Pelegri, F. (2007). The zebrafish maternal-effect gene *cellular atoll* encodes the centriolar component sas-6 and defects in its paternal function promote whole genome duplication. *Dev. Biol.* **312**, 44–60.

Yabe, T., Ge, X., Lindeman, R., Nair, S., Runke, G., Mullins, M. C., and Pelegri, F. (2009). The maternal-effect gene *cellular island* encodes Aurora B kinase and is essential for furrow formation in the early zebrafish embryo. *PLoS Genet.* **5**, e1000518.

CHAPTER 6

High-Throughput Target-Selected Gene Inactivation in Zebrafish

Ross N.W. Kettleborough[*], Ewart de Bruijn[†], Freek van Eeden[‡], Edwin Cuppen[†] and Derek L. Stemple[*]

[*]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[†]Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

[‡]MRC-CDBG Department of Biomedical Science, The University of Sheffield, Western Bank, Sheffield, UK

Abstract

There is an increasing requirement for efficient reverse genetics in the zebrafish, Here we describe a method that takes advantage of conventional mutagenized libraries (identical to ones used in forward screens) and re-sequencing to identify ENU-induced mutations in genes of interest. The efficiency of TILLING (Targeting Induced Local Legions IN Genomes) depends on the rate of mutagenesis in the library being screened, the amount of base pairs screened, and the ability to effectively identify and retrieve mutations on interest. Here we show that by improving the mutagenesis protocol, using *in silico* methods to predict codon changes for target selection, efficient PCR and re-sequencing, and accurate mutation detection we can vastly improve current TILLING protocols. Importantly it is also possible to use this method for screening for splice and mis-sense mutations, and with even a relatively small library, there is a high chance of identifying mutations across any given gene.

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00006-9

I. Introduction

Reverse genetics in zebrafish research has suffered because of the lack of an efficient gene targeting technology. The existence of a high-quality zebrafish genome sequence has contributed to the effectiveness of a reverse genetic method known as Targeting Induced Local Lesions IN Genomes TILLING, first described in *Caenorhabditis elegans* and *Arabadopsis thaliana* (Jansen *et al*., 1997; McCallum *et al*., 2000), which can be used to select for disruptive mutation in specific zebrafish genes (Wienholds *et al*., 2003). This method traditionally uses cel1 nuclease digestion at heterozygous positions and fragment analysis to identify ENU induced mutations (Till *et al*., 2004). This process is, however, labor-intensive and has a high false negative rate. Here we show that by sequencing amplified DNA from mutation carriers we are able to increase the sensitivity and efficiency of the TILLING process in a cost-effective manner. Using the methods described here, it is possible to identify knockout (KO) alleles, given that you have a suitably sized mutagenized library. Importantly it is also possible to use this method for screening for splice and mis-sense mutations, and with even a relatively small library, there is a high chance of identifying mutations across any given gene to give an allelic series.

II. Mutagenesis and Library Creation

The strain of zebrafish used has shown to have an effect on the mutation rate achieved. We have found the Tübingen Long Fin (TL) strain to be excellent for generation of highly mutagenized libraries. The TL strain appears to be more robust, is able to take higher doses of the mutagen, and has the strength required to survive the treatments and retain good fertility. TL fish also have the benefit of having high levels of fecundity, making library generation by outcrossing mutagenized F0's straightforward. ENU was found to be the most efficient chemical mutagen for inducing point mutations that can be recovered in an F2 breeding scheme (Mullins *et al*., 1994; Solnica-Krezel *et al*., 1994). For this type of reverse screen, ENU mutagenesis of mitotically dividing spermatogonial cells is ideal, as it is straightforward and effective.

The methods used for mutagenesis are a slightly modified version of previously published approaches (van Eeden *et al*., 1999). The frequency of ENU-induced mutations is dosage dependent, and increasing the ENU concentration from 2.0 to 3.0 mM results in a significant increase in mutation efficiency (Solnica-Krezel *et al*., 1994). ENU doses that are significantly higher than 3.0 mM have been shown to result in high lethality. It has been shown, however, that by using two separate treatment concentrations (3.0 and 3.3 mM) it is possible to maximize mutation efficiency without high lethality. By carefully monitoring the fish during treatments, it is possible to alter the ENU concentration, depending on their health, thereby maximizing the mutation rate in a batch of fish. Mutagenesis is performed in mesh bottomed mating tanks, to allow efficient and relatively stress free fish transfer

between treatments and washes (compared to netting). It is crucial that the fish are kept in the dark without any disturbance during the ENU treatment. To obtain approximately 20 fertile individuals from six weekly treatments, it is possible to start with 50 males, as long as they are well fed and looked after between treatments. Following the last mutagenesis, mutagenized males are mated at weekly intervals. F1 progeny obtained after 3 weeks are non-mosaic for ENU-induced mutations and are raised to adulthood (Mullins *et al*., 1994; Solnica-Krezel *et al*., 1994). The mutagenesis protocol can reduce the fertility of the fish significantly, and can alter their behavior and ability to breed naturally. In cases where males are not able to efficiently fertilize the eggs of females during natural matings it can be useful to perform *in vitro* fertilization (IVF) to obtain offspring. This can be a very efficient way of recovering offspring from highly mutagenized individuals. We routinely determine the efficiency of mutagenesis by measuring induced mutations in 19 exons from 384 individual F1 fish from the library. The list of exons used in this study is shown in Table I. The use of this consistent set of exons allows us to efficiently compare different mutagenesis experiments. Exons are amplified and sequenced exactly as is described in the screening process, and yields ~3 Mb of sequence from the 384 individuals. Mutation rates using this protocol are consistently between 1 mutation per 175,000–250,000 bp.

Outcrossed fish from mutagenized F1 individuals are raised for 3 months, and then a sample of caudal fin is taken from each individual. We house these fish in groups of 24 corresponding to two rows of a 96 well plate so the fish can be identified post screening. DNA is isolated by proteinase K digestion of caudal fin samples, followed

Table I
Exons used in rate test

lrrn1	ENSDARG00000060115	ENSDARE00000613900
vasn	ENSDARG00000076215	ENSDARE00000861807
lrrc4	ENSDARG00000069402	ENSDARE00001013589
lrrc4c	ENSDARG00000016739	ENSDARE00000337236
islr2	ENSDARG00000051875	ENSDARE00000523703
unc5b	ENSDARG00000033327	ENSDARE00000836942
flrt1a	ENSDARG00000077556	ENSDARE00000790664
flrt3	ENSDARG00000076895	ENSDARE00000877140
lrrtm2	ENSDARG00000071374	ENSDARE00000975875
lrrtm1	ENSDARG00000052713	ENSDARE00000532567
lrrtm4	ENSDARG00000077562	ENSDARE00000869580
fgfrl1a	ENSDARG00000032617	ENSDARE00000430893
lrrtm4l1	ENSDARG00000080015	ENSDARE00001036367
lrit2	ENSDARG00000030626	ENSDARE00000897040
lrrn1	ENSDARG00000077112	ENSDARE00000799931
jam3b	ENSDARG00000061794	ENSDARE00000646607
jam3b	ENSDARG00000061794	ENSDARE00000646680
fkrp	ENSDARG00000062451	ENSDARE00000661096
gyg1	ENSDARG00000011934	ENSDARE00000080562

by isopropanol precipitation of DNA and centrifugation in 96 deep well blocks and re-arrayed into 384-well format at approximately 10 ng/μl for screening purposes.

III. PCR and Re-sequencing of Exons

An outline of the screening process can be seen in Fig. 1. For capillary sequencing using the ABI3730xl platform, we used nested PCR owing to the need for uniform,

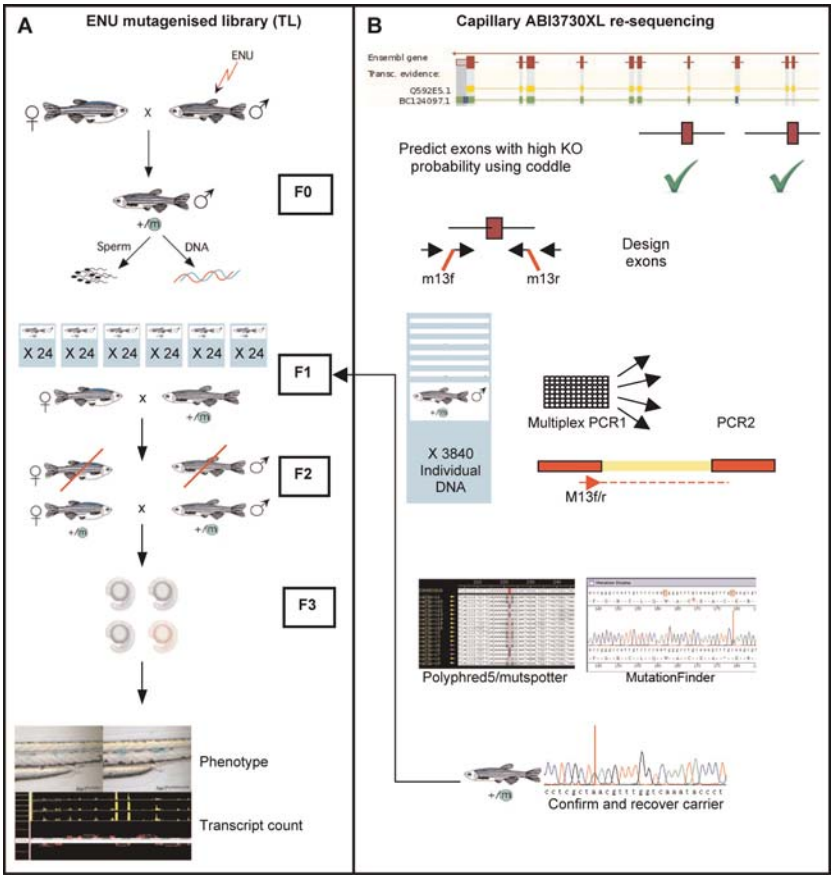


Fig. 1 Screening a mutagenized library by PCR and capillary sequencing. (A) A mutagenized library is created by ENU mutagenesis, mutagenized F0 males are outcrossed, and F1 fish are re-distributed into library tanks and fin clipped to make DNA for screening. After screening, tanks containing KO indviduals are fin clipped to identify the carrier of the mutation, carriers are outcrossed, F2's identified, and incrossed to check for a phenotype during the first 5 days of development (F3). (B) Exons with high probablilty of containing a nonsense allele are predicted with coddle, nested primers are designed around exons chosen, and PCR and sequencing is performed across every individual in the library. Sequences are screened with mutationfinder, and Polyphred5 and hits are recorded and recovered from the library. (See color plate.)

high-specificity PCR product for high-quality sequencing. Initially exons are screened for their suitability for mutation finding *in silico* (http://www.proweb.org/coddle/). For each exon to be screened, it is straightforward to predict the probability that each codon will produce a nonsense allele, according to the changes induced by ENU mutagenesis, and to give each exon a probability score. Again, this allows the user to optimize the probability of finding a nonsense mutation. Once suitable exons have been identified, nested primers are designed surrounding the exons (maximum 500 bp long), with standard m13 forward (m13f) and m13 reverse (m13r) tails being included on the internal primers to standardize the capillary sequencing. This also allows PCR1 to be multiplexed to save on time and reagents, and some of the components (primers and dNTPs) in PCR2 to be minimized to send the PCR reaction to completion, therefore standardizing PCR product concentration with low concentrations of residual primers.

DNA from each fish was amplified individually in a 384 well format for each amplicon by nested PCR. PCR1 samples contained 5 μl genomic DNA (approximately 50 ng) 0.2 μM forward primer (x4 amplicons), 0.2 μm reverse primer (x4 amplicons), 200 μM of each dNTP, 10 mM Tris-HCl (pH 8.5), 50 mM KCl, 2 mM $MgCl_2$, 0.1% Tween20 and 0.2 μl Taq polymerase. PCR1 was cycled using a touchdown program (94 °C for 180 s, 18 cycles of 92 °C for 20 s; 65 °C for 30 s [-0.5 °C per cycle]; 72 °C for 60 s, 25 cycles of 92 °C for 20 s; 56 for 30 s; 72 °C for 60 s, followed by one cycle of 72 °C for 180 s). PCR2 reactions were performed individually using template transferred from PCR1 with a 384-well disposable replicator (Genetix X5054). The reaction mix also contains 0.05 μM M13f tailed forward primer, 0.05 μM M13r tailed reverse primer, 50 μM of each dNTP, 10 mM Tris-HCl (pH 8.5), 50 mM KCl, 2 mM $MgCl_2$, 0.1% Tween20 and 0.2 μl Taq polymerase. Eight representative samples from each PCR2 were run on a 2% agarose gel, and successful PCRs were diluted with 30 μl double distilled H_2O. One 384 well plate was routinely sequenced with both m13f and m13r primers using standard ABI BigDye v3.1chemistry and run on an ABI3730XL DNA Analyzer. Sequences were manually checked and the best primer for sequencing was chosen to sequence the remainder of the library (frequent indels and polyT regions in exon flanking introns can adversely affect sequence quality and subsequent calling of heterozygous positions). Each PCR for any given exon can give varying results, and the amount of primer and dNTPs remaining in the reaction after amplification can adversely affect the sequencing reaction, creating noise in the background of the sequencing traces. This noise can be efficiently removed by treating the PCR products with small quantities of exonuclease I (to remove single-stranded primers) and shrimp alkaline phosphatase (to hydrolyze dNTPs) before sequencing.

IV. Analysis of Re-sequencing Data and Retrieval of Mutations

Traces from every individual covering each exon should be deposited in a convenient archive for analysis. Two programs are used for the analysis of sequencing

traces, MutationFinder (Goda *et al*., 2006) and PolyPhred (http://droog.gs.washington.edu/polyphred/) (Stephens *et al*., 2006). It is important that the capacity to analyze the data with these two programs be set up efficiently, as manual screening of sequences for heterozygous positions is laborious and often leads to mutations being missed. We found that using both programs for analysis, where possible, is the best solution for mutation detection, as they each use different algorithms producing both overlapping and non-overlapping valid hits. Both MutationFinder and PolyPhred compare sequence traces obtained from different individuals to identify heterozygous sites for single nucleotide substitutions. Each program runs an analysis package, and then uses a viewer (MutationViewer and Consed respectively) to present the results (Gordon *et al*., 1998; Goda *et al*., 2006). MutationViewer can view all traces from a given archive, whereas Consed allows 384 traces to be viewed efficiently at any one time. Each program marks the consensus trace with colors to identify heterozygous positions and allows the user to compare homozygous wild-type and heterozygous individuals at the position of the mutation. Positions in the sequence that contain a single individual with a heterozygous base at a given position can be marked and saved. MutationFinder also predicts the consequences of the mutations so only nonsense alleles (if these are the only mutations required) need to be saved, but both programs will also allow the identification of essential splice site and mis-sense mutations allowing an allelic series to be generated. Careful manual annotation of the heterozygous positions in the set of traces for each exon takes practice, but recognizing a real heterozygous position from background should become quickly obvious. Noisy traces (with background second peaks surrounding the "het" position) are a common reason for calling false positives, and all potential mutations should be confirmed before fin clipping any library fish.

Re-amplifying and re-sequencing the sample of DNA containing the detected mutation is used to independently confirm results from the MutationFinder and PolyPhred programs. DNA from the detected mutation, and surrounding DNA samples in the plate as a negative control, can be re-amplified and re-sequenced using the same primers. Once confirmed, heterozygous positions in traces can be translated into the corresponding tank ID from the library. Library tanks containing positive hits from the repeat PCR and sequencing can then be fin clipped, kept in single tanks, and re-sequenced to identify the founder carrying the mutation.

Each founder carrying a detected and confirmed mutation can then be kept in a single tank for outcrossing. Carriers should be well-treated and given extra food to ensure they perform well when being outcrossed. However, IVF can be performed if the individual fails to produce offspring by natural matings. Once outcrossed, a genotyping protocol for the line should be carefully designed to make genotyping efficient.

KASPar, an allele specific amplification genotyping assay provided by KBioscience (http://www.kbioscience.co.uk/) (Cuppen, 2007), allows carriers of heterozygous mutations to be identified in approximately 5 h (compared to 1 week for conventional PCR and re-sequencing) when combined with HotShot DNA preparation from fin clips (alkaline lysis of fin clips at 95 °C for 30 min in 50 μl

of 25 mM NaOH, 0.2 mM EDTA followed by neutralization by adding 50 µl 2 M Tris-HCl pH 5.5). KASPar uses 5' primers specific to each allele at a given heterozygous position, and a common 3' primer. The 5' primer has a tail sequence that is recognized by fluorescently labeled oligos of different colors, allowing the alleles to be measured by fluorescence using a standard plate reader. KASPar software is used to identify +/+, +/−, and −/− individuals from the data given from the plate reader. Being able to efficiently identify carriers for any given mutation saves time and effort and means fish do not need to be maintained in single tanks for long periods during genotyping. Once a family has been generated, in-crosses can be set up for phenotypic analysis.

Acknowledgments

The work was supported by the Wellcome Trust and ZF Models Sixth Framework Programme (EC Contract LSHG-CT-2003-503496). FvE was supported by MRC centre grant G0700091.

References

Cuppen, E. (2007). Genotyping by Allele-Specific Amplification (KASPar). *CSH Protoc 2007* pdb prot4841.

Goda, T., Abu-Daya, A., Carruthers, S., Clark, M. D., Stemple, D. L., Zimmerman, L. B. (2006). Genetic screens for mutations affecting development of Xenopus tropicalis. *PLoS Genet.* **2**(6), e91.

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* **8**(3), 195–202.

Jansen, G., Hazendonk, E., Thijssen, K. L., and Plasterk, R. H. (1997). Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat. Genet.* **17**(1), 119–121.

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeted screening for induced mutations. *Nat. Biotechnol.* **18**(4), 455–457.

Mullins, M. C., Hammerschmidt, M., Haffter, P., and Nusslein-Volhard, C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* **4**(3), 189–202.

Solnica-Krezel, L., Schier, A. F., and Driever, W. (1994). Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**(4), 1401–1420.

Stephens, M., Sloan, J. S., Robertson, P. D., Scheet, P., and Nickerson, D. A. (2006). Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nat. Genet.* **38**(3), 375–381.

Till, B. J., Burtner, C., Comai, L., and Henikoff, S. (2004). Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.* **32**(8), 2632–2641.

van Eeden, F. J., Granato, M., Odenthal, J., and Haffter, P. (1999). Developmental mutant screens in the zebrafish. *Methods Cell Biol.* **60**, 21–41.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R. H., Cuppen, E. (2003). Efficient target-selected mutagenesis in zebrafish. *Genome Res.* **13**(12), 2700–2707.

CHAPTER 7

Genetic Suppressor Screens in Haploids

Xiaoying Bai,[*,†,‡] Zhongan Yang,[¶] Hong Jiang,[¶] Shuo Lin[¶] and Leonard I. Zon[*,†,‡]

[*]Howard Hughes Medical Institute, Boston, Massachusetts, USA

[†]Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, USA

[‡]Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA

[¶]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California, USA

Abstract

As a vertebrate genetic model, the zebrafish has been well recognized for its strength in studying a variety of biological processes and human diseases. Traditional forward genetic screens in zebrafish have generated a large pool of mutants with interesting phenotypes resembling human diseases but the underlying mechanisms are not well understood. A powerful approach to elucidate the mechanisms of these mutants is the modifier screen, which identifies 2nd-site mutations that specifically enhance or block the phenotype of a given mutant. Here we described the first genetic suppressor screen in zebrafish, which identifies a novel transcriptional mechanism regulating erythropoiesis. In combination with the haploid genetics in zebrafish, we have shown the feasibility and strength of a modifier screen in zebrafish. This strategy will greatly broaden the utility of the zebrafish as a model for making original discoveries and establishing novel paradigms for understanding vertebrate biology.

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00007-0

I. Introduction and Rationale

The zebrafish has become a popular vertebrate genetic model organism for human diseases and pharmaceutical research due to the large number of embryos, the external embryonic development, and the optical transparency of the embryos. The traditional forward genetic screens have identified hundreds of mutants that affect evolutionarily conserved developmental pathways (Development 1996). An important approach for further dissecting these pathways is the modifier screen. By starting with a mutant of interest, forward genetics can be used to screen for additional mutations that can either enhance or suppress the original phenotype. These types of modifier genetic screens are a powerful and unbiased means to isolate interacting components of a genetic pathway because they do not rely on knowing the molecular nature or the underlying mechanism of the mutant gene. Extensively used in lower organism models, modifier screens have made a great contribution to our understanding of major genetic pathways that regulate developmental and biological processes (Jorgensen and Mango, 2002; St Johnston, 2002).

Traditionally, a modifier screen is performed in a sensitized genetic background in which the function of a particular pathway is only partially disrupted; therefore the viability of mutant animals is not affected. Such a genetic background can be achieved by using a heterozygous mutant with haploinsufficient phenotype or a homozygous mutant of a weak allele. In addition, modifier screens are often designed for isolation of dominant mutations in order to save time and spaces. However, most existing zebrafish mutants are not suitable for such modifier screens due to the embryonic lethality of these mutants and lack of phenotypes as heterozygous. In order to perform a modifier screen in an embryonic lethal mutant, we have developed a strategy to circumvent the lethality by incorporating a transgene-based rescue approach (Bai *et al*., 2010). Aimed to isolate both dominant and recessive mutations, we also took the advantage of the haploid genetics in the zebrafish.

II. Methods – an Example of Suppressor Screen in *Moonshine* Mutants

A. Rescue of *Moonshine* Mutants by a BAC Transgene

The zebrafish *moonshine* (*mon*) mutant is defective in a gene named TIF1γ, which belongs to the family of transcriptional intermediary factor 1 (TIF1) (Ransom *et al*., 2004). Loss of TIF1γ function in *moonshine* mutants causes profound anemia by blocking erythroid differentiation. Majority of homozygous mutants die at 7–10 days post fertilization. The rare escapers that survive to adulthood are also extremely anemic and never breed. To elucidate genetic factors that interact with TIF1γ in the regulation of erythropoiesis, a genetic suppressor screen was conducted to identify second-site mutations that could suppress *mon* phenotype by restoring

erythropoiesis in *mon* mutant (Bai *et al*., 2010). To generate viable *mon* mutants, we injected a modified zebrafish BAC construct into the one-cell stage embryos from *mon* heterozygous intercross (Fig. 1A). This BAC construct contains the full-length TIF1γ genomic locus. In addition, a GFP reporter cassette driven by the b-actin promoter was inserted closely to the TIF1γ locus by homologous recombination. After established germline transmissions, we confirmed the full rescue of *mon* mutants by the transgene and kept *mon* homozygous mutants in this transgenic background.

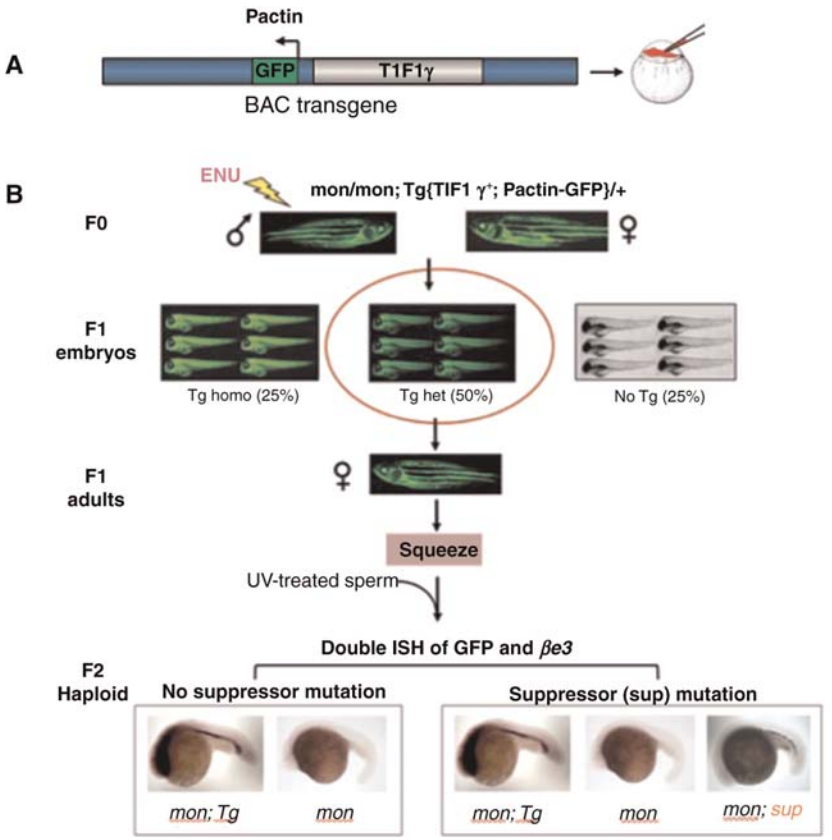


Fig. 1 Genetic suppressor screen in the *mon* mutant. A) A diagram of the BAC transgenic construct containing a wild-type TIF1γ locus. The transgene was injected into one-cell stage embryos to rescue the lethality of *mon* fish. A GFP maker driven by an actin promoter (Pactin) was inserted onto the backbone of BAC. B) Scheme of the suppressor screen. BAC transgenic fish are green fluorescent. In F1 generation, three groups of embryos were obtained: transgene homozygous (*mon*; *Tg/Tg*), transgene heterozygous (*mon*; *Tg/+*), and embryos with no transgene (*mon*). Only transgene heterozygous (in the red circled) were raised up to adults. *Double in situ hybridization* of GFP and *βe3* globin was performed on F2 haploid embryos. Note the GFP staining on *mon*; *Tg* haploids (strong in the head and weak throughout the body) but not on haploids lacking the transgene. "sup" indicates a suppressor mutation. (See color plate.)

B. The Strategy of Haploid Screen

Although this transgenic line would allow us to identify dominant suppressors in F1 generation, dominant suppressors are rare in general and may require a sensitized genetic background. On the other hand, a traditional F3 screen for recessive mutations is time consuming and requires large tank space. In order to identify recessive suppressors in a relatively faster way, we designed a haploid screen strategy (Fig. 1B). Haploid zebrafish embryos can survive up to 3 days with relatively normal development of most major organs including blood. At 24 hpf, *in situ* hybridization of *embryonic globin beta 3 (βe3)* reveals that primitive erythroid cells are present normally in wild-type haploid embryos but absent in mon haploids. We therefore chose *in situ* hybridization of *βe3* at 24 hpf as our screen readout.

1. ENU Mutagenesis and Generation of F1 Fish

We followed the mutagenesis protocol from Lila Solnica-Krezel's laboratory at Vanderbilt University. In brief, 4–5-month-old male fish that were homozygous for *mon* and heterozygous for the BAC transgene (*mon/mon*; *Tg/+*) were treated with ENU (N-ethyl-N-nitrosourea) at the final concentration of 3–3.5 mM for 1 h per treatment. After each treatment, the fish were recovered first in the recovery buffer (300 mg/L Instant Ocean, 100 mM sodium phosphate, PH = 6.5) containing 10 mg/L tricaine for 2–3 h, then in fresh recovery buffer (no tricaine) for overnight. Fish were treated once a week for 5 weeks. Forty-two male fish were used for ENU mutagenesis and 28 fish survived at the end of treatment. These F0 fish were used to cross with females that were of the same genotype (*mon/mon*; *Tg/+*) in the set-up of single-pair mating. In total, we obtained about 5600 embryos from mutagenized F0 males by breeding them once a week for 3 weeks. All of the F1 progeny should be *mon* homozygous and one-quarter would not carry the transgene; therefore were GFP^-. These embryos (*mon/mon*) were visually scored for blood cells in circulation at 3 dpf. *Mon* mutants have no circulating blood cells. If any circulating blood cells were found, it would indicate a dominant suppressor mutation. We did not detect any dominant suppressor mutations that could give rise to circulating blood cells in these F1 *mon/mon* progenies, which all died after 1 week. The rest embryos carried the transgene; therefore were GFP^+ and could survive to adulthood. In total, more than 4000 F1 progenies were raised up.

2. Haploid Screen for Recessive Suppressor Mutations

The survived F1 fish were genotyped and only the female fish that are heterozygous for the transgene (*mon/mon*; *Tg/+*) were kept for haploid screen. Each F1 female carried a different set of mutagenized genes derived from the independently mutagenized spermatogonial cells of their father. Generation of haploid embryos is as described previously (Shepard *et al*., 2005). In brief, individual F1 female was

squeezed to release eggs, which was then fertilized by UV-treated sperms from wild-type males. UV treatment destroys the paternal DNA and therefore will not contribute to the genome in the fertilized eggs. After 20 h post fertilization, survived embryos showed typical morphology of haploid embryos, such as shortened body length. These F2 haploid embryos were fixed at 22 hpf by 4% PFA for *in situ* hybridization. Because the F1 mom was heterozygous for the transgene, half embryos should carry the transgene; therefore were GFP^+, while the other half were GFP^- without the transgene. Theoretically, we only need to focus on the GFP^- embryos, which were true *mon* mutants. However in our case, all embryos were GFP^+ at 22 hpf due to the maternally deposited GFP protein in the eggs. This maternal, long half-life GFP protein made it impossible to distinguish embryos with and without the transgene based on GFP fluorescence. Luckily we found that the maternal GFP mRNA could not be detected at 22 hpf. We therefore decided to do a double *in situ* hybridization for both GFP and *βe3*. If an embryo carried the transgene, the whole embryo would be stained for GFP expression due to the ubiquitous *b-actin* promoter, whereas *βe3* should only be detected in the Intermediate Cell Mass (ICM) blood island. In order to do one-color staining for both probes, we decreased the concentration of the GFP probe to one-quarter of that of the *βe3* probe so that the *βe3*-staining could stand out of a GFP-stained background. As shown in Fig. 1B, in the absence of suppressor mutations, 50% haploid embryos were *mon* carrying the transgene (*mon*;*Tg*); therefore stained for both GFP and *βe3*, while the rest half were *mon* without the transgene (*mon* only); therefore no staining at all. If we detected a clutch of embryos with a third pattern, which was $\beta e3^+$ but GFP^-, it would indicate a presence of a suppressor mutation that could restore blood formation in *mon* mutants. If such a clutch were found, the same female would be squeezed at least one more time to make sure the rescue phenotype was repeatable. At the end we found three females that were potential hits (the putants). The numbers of fish used in the screen were summarized in Table I.

Table I
Summary of the number of fish and embryos used in the screen

No. of males used for ENU mutagenesis	42
No. of males survived after ENU mutagenesis (F0)	28
No. of F1 embryos generated from F0 males	5612
No. of GFP F1 embryos used to screen for dominant suppressors at 3 dpf	1400
No. of F1 females used for squeezing	2134
No. of F1 females that are squeezable	1310
No. of haploid clutches that had $\geq$10 embryos survived after 22 hpf and used for double ISH	815
No. of putant F1 females identified	3
No. of mutants verified in diploid	2

3. Recovery and Validation of Haploid Mutants

To validate putants, we first recovered suppressor mutations in the diploid background (Fig. 2). Each F1 putant female was crossed with a transgenic male (*mon/mon*; *Tg/+*) to generate the F2 family. The F2 progenies were raised up and half of them would carry the potential suppressor mutation (*sup/+*). Again, only the progeny with the genotype of *mon/mon*; *Tg/+* were selected and randomly paired with each other within the family. Double *in situ* hybridization for GFP and *βe3* was performed on the F3 embryos. In theory, if a suppressor mutation is fully penetrant and both parents are carriers, one out of 16 (6.25%) embryos in a given clutch from a single pair would be GFP^{-} and $\beta e3^{+}$, indicating a rescue. We were able to validate two out of three putants in the diploid background. At the same time, any morphological phenotype linked to the rescue phenotype would be very useful to identify mutants. In our case, both validated mutants (*sunshine* and *Eos*) had distinctive morphological phenotype and were recessive lethal. Because the suppressor mutants were identified

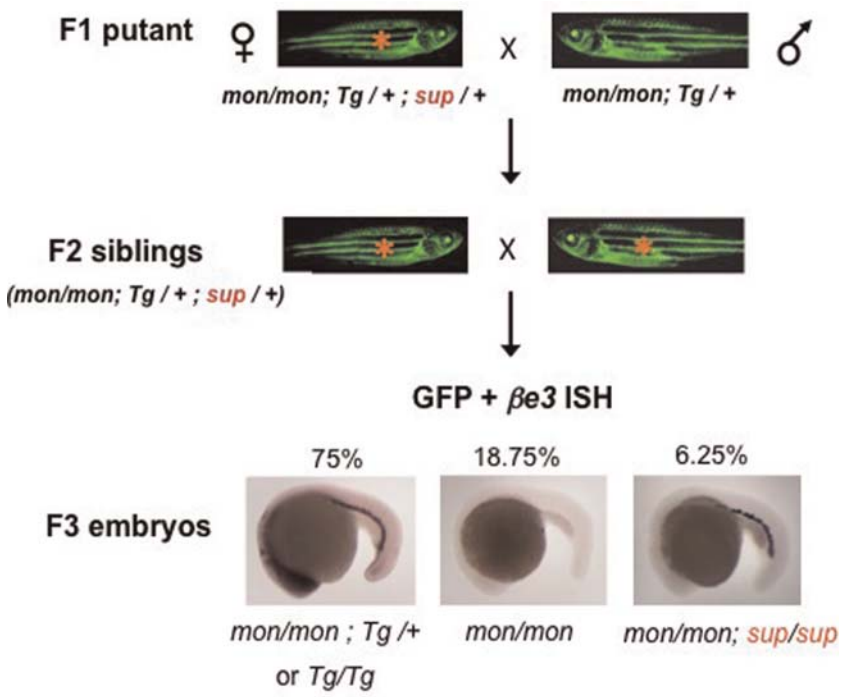


Fig. 2 Verify suppressor mutants in diploids. A putative suppressor mutant (F1 putant) female identified in the haploid screen (indicated by a red asterisk on the fish) was crossed with a transgenic male to give rise to an F2 family. F2 siblings were subjected to random pair mating. If both fish were heterozygous carriers of the suppressor mutation (sup), 6.25% embryos would be double mutant for *mon* and the suppressor mutation (*mon/mon*; *sup/sup*) and show GFP negative but *βe3* positive. (See color plate.)

in the transgenic *mon* mutants, it is important to verify them in the original *mon* mutants. Either mutant was therefore crossed with *mon*± fish to generate double mutant heterozygous, which were then intercrossed to give rise to double mutant embryos. *In situ* hybridization for *βe3* followed by genotyping confirmed that blood formation in *mon* was rescued by either repressor mutant. To clone the mutant gene, the suppressor mutant was outcrossed with wild-type fish for at least three generations to clean out additional mutations in the genome and then crossed with the mapping strain SJD for positional cloning.

III. Conclusions and Prospective

Here we described the first genetic suppressor screen in the zebrafish. Using a combination of transgene-mediated rescue and haploid genetics in the zebrafish, we identified two suppressor mutants that could rescue the blood formation in a bloodless mutant *mon*. By characterizing one of the mutants, *sunshine*, we have discovered a novel mechanism regulating erythroid differentiation (Bai *et al*., 2010). Our work not only demonstrates again that the zebrafish is a powerful model organism for the study of vertebrate genetics, but also provides a foundation for undertaking genetic modifier screens in the zebrafish. The transgenic rescue approach leads to large numbers of viable fish with a homozygous lethal mutation. Because the linkage of GFP on the transgene distinguishes the true mutants from the mutants rescued by the transgene, this method eliminates the need of genotyping and is more reliable than calculating the fraction of expected embryos based on the Mendelian ratio. This is particularly important for haploid screens, in which the average number of survived haploid embryos per clutch is usually very low (less than 30 embryos).

In our screen, we used a GFP marker driven by a strong, ubiquitous promoter, which led to a strong maternal effect that may cause a problem under certain circumstances. To avoid this problem, a weak ubiquitous promoter or a tissue-specific promoter will be better to use in the future. For instance, the recently published zebrafish ubiquitin promoter may be well suited to drive a marker gene expression on the transgene as it is expressed early and ubiquitous during embryonic development but much weaker than the actin promoter (Mosimann *et al*., 2011).

Compared to F3 screens for recessive mutations, haploid screen is much faster and can save a lot of tank space. However, the surviving rate of haploid embryos is poor, especially after 24 hpf, making it difficult to conduct screens on late-stage embryos. In addition, the morphology of haploid embryos can be extremely variable and therefore may not be suitable for morphology-based screens. In these cases, diploidization of the haploid embryos by early pressure (EP) or heat shock (HS) may be a better way. The artificially induced gynogenesis allows mutations in the maternal genome to become homozygous in up to 50% of the embryos (Streisinger *et al*., 1986). These diploid embryos develop with normal morphology and can survive to adulthood.

Although in theory a modifier screen can be conducted in virtually any mutants, understanding the molecular nature of the original mutant is important to predict

whether the screen is likely to work or not. In our case, the defective gene TIF1γ in *mon* mutant is a chromatin-associated factor and presumably regulates gene transcription through chromatin. Since the molecular interactions at the chromatin level are usually dynamic and reversible, it is possible to identify chromatin-related pathways that bypass the requirement of TIF1γ. Indeed, both of the suppressor mutants found in our screen are chromatin-related factors (Bai *et al*., 2010 and unpublished data). It is therefore reasonable to conduct a modifier screen on mutants that are defective in molecular pathways involving multiple components and dynamic interactions.

In summary, the suppressor screen described here can be applied to many existing embryonic lethal mutants in the zebrafish. Many of these mutants define critical steps of organogenesis and are excellent models of human diseases. Performing suppressor screens on these mutants will be very useful to identify interacting pathways that may have clinical implications.

Acknowledgements

We thank Jocelyn LeBlanc, Anna Sessa and Joseph Lee for technical assistance. This study was supported by NIH grants 5PO1HL32262-28 and 5R37 DK55381-010.L.I.Z. is a founder and stock holder of Fate, Inc. and a scientific advisor for Stemgent.

References

Bai, X., Kim, J., Yang, Z., Jurynec, M. J., Akie, T. E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., Zhou, Y., Grunwald, D. J., Lin, S., Cantor, A. B., Orkin, S. H., and Zon, L. I. (2010). TIF1γamma controls erythroid cell fate by regulating transcription elongation. *Cell* **142**, 133–143.

Jorgensen, E. M., and Mango, S. E. (2002). The art and design of genetic screens: caenorhabditis elegans. *Nat. Rev. Genet.* **3**, 356–369.

Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J., and Zon, L. I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177.

Ransom, D. G., Bahary, N., Niss, K., Traver, D., Burns, C., Trede, N. S., Paffett-Lugassy, N., Saganic, W. J., Lim, C. A., Hersey, C., Zhou, Y., Barut, B. A., Lin, S., Kingsley, P. D., Palis, J., Orkin, S. H., and Zon, L. I. (2004). The zebrafish moonshine gene encodes transcriptional intermediary factor 1gamma, an essential regulator of hematopoiesis. *PLoS Biol.* **2**, E237.

Shepard, J. L., Amatruda, J. F., Stern, H. M., Subramanian, A., Finkelstein, D., Ziai, J., Finley, K. R., Pfaff, K. L., Hersey, C., Zhou, Y., Barut, B., Freedman, M., Lee, C., Spitsbergen, J., Neuberg, D., Weber, G., Golub, T. R., Glickman, J. N., Kutok, J. L., Aster, J. C., and Zon, L. I. (2005). A zebrafish bmyb mutation causes genome instability and increased cancer susceptibility. *Proc. Natl Acad. Sci. U. S. A.* **102**, 13194–13199.

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. *Nat. Rev. Genet.* **3**, 176–188.

Streisinger, G., Singer, F., Walker, C., Knauber, D., and Dower, N. (1986). Segregation analyses and gene-centromere distances in zebrafish. *Genetics* **112**, 311–319.

CHAPTER 8

Transgenic Zebrafish Using Transposable Elements

Karl J. Clark, Mark D. Urban, Kimberly J. Skuster and Stephen C. Ekker

Mayo Clinic, Department of Biochemistry and Molecular Biology, Mayo Addiction Research Center, Rochester, Minnesota, USA

Abstract

DNA transposons are effective chromosomal engineering vehicles for making transgenic zebrafish. We describe both autonomous and non-autonomous transposable elements, and we compare and contrast popular transposon systems. The

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00008-2

Tol2 system is a robust gene transfer tool and has been selected as the primary transposon platform, facilitating the development of an array of reagents readily shared within the zebrafish community. We present common transposon and transposase vectors within the field based on the *Tol2* system. We describe methods with a high success rate of generating transgenic zebrafish using *Tol2* vectors, including key quality control steps during the transgenesis process. Together, these data should enable the ready generation of transgenic zebrafish for scientific inquiry.

I. Introduction

Stuart and colleagues reported the first germline transgenic zebrafish (*Danio rerio*) over 20 years ago (Stuart *et al*., 1988). Since then, increases in the efficiency of zebrafish genome engineering have dramatically amplified both the use and functionality of this vertebrate model organism. Beyond the revolution of fluorescent proteins, arguably no tool has aided zebrafish genome engineering more than active DNA transposons. Here we discuss how transposons work in general, the application of transposon systems to zebrafish, and share a method for routine transgenesis using the *Tol2* transposon system (Kawakami and Shima, 1999; Koga *et al*., 1996).

DNA or "cut and paste" transposons make up a significant portion of many genomes. When active, these "wild" transposons exist as both "autonomous" and "non-autonomous" versions. The autonomous version (Fig. 1A) encodes a complete and functional transposase that is capable of identifying, excising, and reinserting the DNA element defined by its inverted terminal repeats (ITR) or other elements with the same ITRs. Non-autonomous elements are DNA transposons that can still be moved by a functional transposase but are no longer able to produce their own transposase protein due to mutations or deletions within the coding region of the transposase (Fig. 1B). These elements are reliant on autonomous elements for mobilization. The ability of the transposon to work on non-autonomous elements easily permits the separation of transposase activity from the mobile element. Without this step, mobilization of the element could continue uncontrolled. This necessary step in the adaptation of wild transposons to "domesticated" genetic tools occurred rapidly once their mechanism was understood – the transposase coding sequence is simply replaced with an expression cassette of a gene of interest (Fig. 1C). As individual transposon elements were further understood, it was recognized that the inverted repeats could be modified to eliminate all unneeded sequences that might unknowingly affect transcription, splicing, polyadenylation, etc. Thus, current transposon tools are short – improving subcloning and transposition efficiencies (Fig 1D). The transposase can be supplied on a separate DNA expression cassette in *trans* (Fig 1E) or outside of the transposon cassette in *cis* (not shown). However, in zebrafish the most common way to supply the transposase is by co-injection of mRNA encoding the transposase (Fig. 1F). This

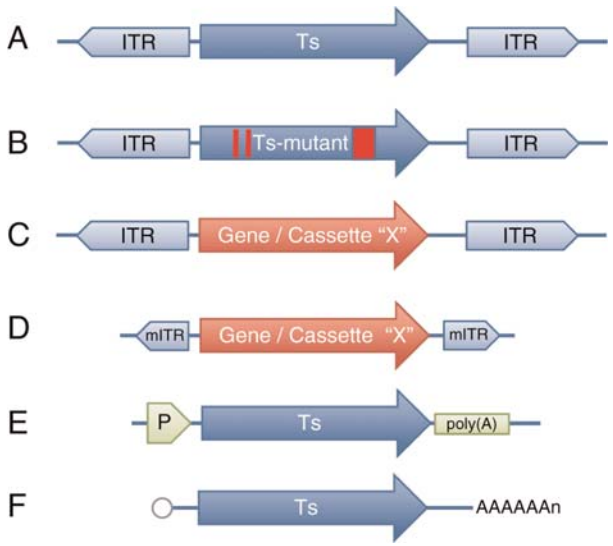


Fig. 1 Transposon structures. A) An autonomous transposon found within the genome in its "wild" configuration contains full inverted terminal repeats (ITR) capable of driving expression of an active transposase located within the ITRs. The transposase is transcribed and processed including polyadenylation from signals located within the ITRs. The transposase mRNA is translated into protein that can recognize sequences at the distal ends of the transposon and "cut" it from the genomic DNA and "paste" it into a new location within the genome. B) Mutations (red areas) in the transposase cause the transposon to become non-autonomous, meaning the transposon has lost its ability to produce functional transposase protein. However, the ITRs of a non-autonomous transposon can be recognized if functional transposase is provided from another source (e.g., an autonomous element elsewhere in the genome). C) An engineered or "domesticated" transposon can be made by replacing the transposase coding region with a different gene or expression cassette. D) As the transposon sequences required for mobilization are understood, a transposon can be made using so-called "minimal" ITR sequences (mITR). Doing so likely removes elements that are required for normal expression of the transposase (promoter, polyadneylation signals, etc.). E) A separate DNA expression cassette can be made by placing the transposase sequence between a promoter (P) and a polyadenylation signal (poly(A)). F) Alternatively, *in vitro* transcribed mRNA can be produced as a transient source of transposase. In zebrafish work, the most common genetic manipulations include the combination of a minimal transposon (D) and *in vitro* transcribed mRNA (F). (See color plate.)

insures that the transposase is only available transiently and that the transposon integrations will be stable following natural mRNA degradation.

The first DNA transposon widely adapted for genetic engineering was the P element in *Drosophila* (Rubin and Spradling, 1982; Spradling and Rubin, 1982). However, it would be years before a transposon system with significant activity in vertebrate cells was identified. In 1996, the first evidence of an active DNA transposon in the genome of the Medaka fish was reported (Koga *et al*., 1996). However, before the active *Tol2* transposase was characterized, another vertebrate DNA transposon, *Sleeping Beauty (SB)*, was reverse engineered from related, ancient, inactive

transposons found in several fish species (Ivics *et al*., 1997). In a series of deliberate steps, the *Tol2* element was isolated from a mutational insertion in the Medaka *tyrosinase* locus and shown to be an autonomous native mobile element (Kawakami *et al*., 1998; Koga *et al*., 1996). This work permitted the identification of functional *Tol2* mRNA, leading to germline transgenesis of zebrafish following injection of a non-autonomous element and *Tol2* mRNA into zebrafish embryos (Kawakami and Shima, 1999; Kawakami *et al*., 2000). Similarly, development of the *SB* transposon system continued with demonstration of remobilization in zebrafish cells followed by germline transgenesis (Davidson *et al*., 2003; Izsvák *et al*., 2000). Since the initial documentation of *SB* and *Tol2* transposon activity in zebrafish, many subtle changes in design and use have been implemented. Significant efforts to engineer *SB* for improved function have led to improvement of the terminal repeats (Cui *et al*., 2002) and successive generations of "hyperactive" transposases (Baus *et al*., 2005; Geurts *et al*., 2003; Mátés *et al*., 2009; Yant *et al*., 2004; Zayed *et al*., 2004). Despite these advances in function, the *SB* transposon system failed to make a widespread splash within the zebrafish research community. Direct comparisons between *Tol2* and *SB* in zebrafish demonstrated that transposition of *Tol2* occurred more rapidly and with substantially higher overall activity in somatic tissues (Balciunas *et al*., 2006). The latter activity permitted efficient analysis of enhancer function in F0 (injected embryos) for the first time (Fisher *et al*., 2006). In addition, and unlike *SB*, *Tol2* transposition efficiency does not significantly drop with transposon size (up to 10 kb) nor is it as sensitive to overexpression inhibition (Balciunas *et al*., 2006). Although other transposons have been subsequently used for germline transgenesis of zebrafish (Emelyanov *et al*., 2006; Koga *et al*., 2008), the best-characterized observations derive from the *Tol2* system. Collectively, these realizations have made *Tol2* the primary choice for generating transgenic zebrafish.

Since the initial characterization of *Tol2*, independent cloning of the *Tol2* element from Medaka (Parinov *et al*., 2004) and the subsequent production of several cloning vectors have facilitated the use of this element in zebrafish. Some have significantly reduced the length of the inverted terminal repeats, making expression vectors smaller and easier to handle while simultaneously removing potential undefined regulatory elements such as promoter, splicing, or polyadenylation signals within the terminal repeats (Balciunas *et al*., 2006; Urasaki *et al*., 2006). The availability of multiple sources of cloning vectors, including mini inverted repeat transposons and Tol2 transposase transcription vectors (Fig 1D,E), facilitated creation of the *Tol2* toolbox for zebrafish. Additional collections of *Tol2* transposon vectors made use of Gateway cloning vectors (Kwan *et al*., 2007; Villefranc *et al*., 2007). In particular, the distribution of the convenient "Tol2-kit" established a transformational series of versatile vectors for use by the scientific community (Kwan *et al*., 2007). The combination of multiple useful plasmids, the completeness of the kit for combining fluorescence reporters for promoter and fusion protein production, and an online "user community" (http://tol2kit.blogspot.com) helped to make the Tol2 transposon system a routine genetic engineering tool that is used widely by the zebrafish community.

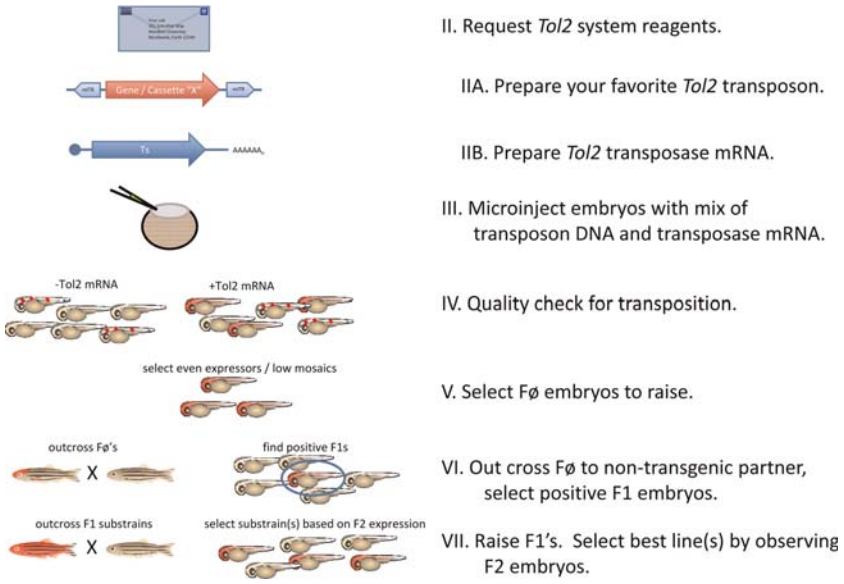


Fig. 2 Overview of Tol2 transgenesis method. The basic approach to producing a transgenic zebrafish using the Tol2 transposon system is diagramed and listed. See text for details. (See color plate.)

Methods: State of the Art *Tol2* Transposon Tools for Zebrafish Transgenesis. Fig. 2 presents an overview of the basic process for production of a transgenic zebrafish using Tol2 transposase.

II. Requesting and Assembling Tol2 System

A. Tol2 Transposon DNA

The first step of the process involves cloning the desired expression cassette between *Tol2* ITR. Many cloning vectors are available within the research community. Fig. 3 documents some of the basic Tol2 cloning vectors and their sources. pTol2Dest2pA (Fig. 3A) and pTol2Dest2pA2 (Fig. 3B) are two minimal Gateway Cloning® multisite destination vectors that are compatible with many available entry vectors as well as other destination vectors available from the Lawson and Chien (Tol2kit) labs, respectively (Kwan *et al*., 2007; Villefranc *et al*., 2007). Both of these Gateway Destination vectors use ITRs that are shorter than the original *Tol2* transposon vectors and are very similar, with the exception that the A2 version has removed about 2 kb of exogenous DNA from the *O. latipes tyrosinase* gene (the locus in which *Tol2* was original identified). Three conventional cloning vectors, pT2AL200R150G (Fig. 3C), pminiTol2 (Fig. 3D), and pKTol2-SE (Fig. 3E) are available from the Kawakami, Ekker, and Clark labs, respectively (Balciunas *et al*., 2006;

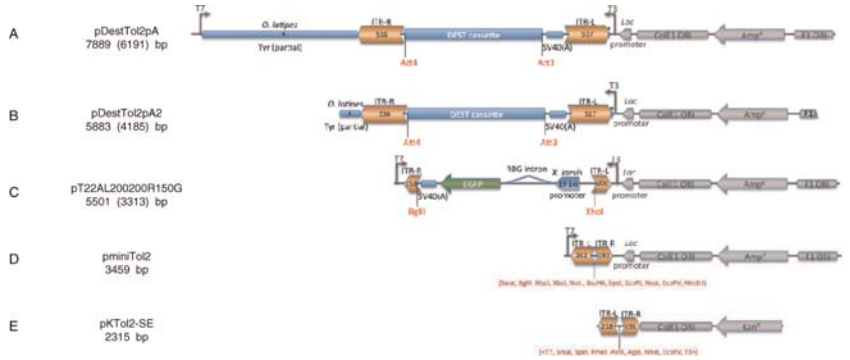


Fig. 3 Available Tol2 transposon vectors. A and B) pDestTol2pA and pDestTol2pA2 are multisite Gateway cloning vectors available from the Lawson lab (Villefranc *et al*., 2007) (http://lawsonlab.umassmed.edu/gateway.html) and the Chien lab (Kwan *et al*., 2007) (http://chien.neuro.utah.edu/tol2kitwiki/index.php/Main_Page), respectively. They are largely the same although the A2 version removes a large chunk of the *O. latipes tyrosinase* (tyr) gene that was cloned along with the original *Tol2* transposon ITRs. C) pT2AL200R150G is a minimal *Tol2* vector available from the Kawakami lab (Urasaki *et al*., 2006) (http://kawakami.lab.nig.ac.jp/). It includes a simple GFP expression cassette that is removed when cloning a gene between the *Bgl*II and *Xho*I restriction endonuclease sites. D) pminiTol2 is a minimal Tol2 transposon vector, available from the Ekker lab (Balciunas *et al*., 2006) (http://zfishbook.org), that shares the same multiple cloning sites as many SB vectors derived from pT/HB or pT/BH (Geurts *et al*., 2003). E) pKTol2-SE is a minimal Tol2 vector with a simplified vector backbone available from the Clark laboratory (http://zfishbook.org). pKTol2-SE shares a multiple cloning site with other vertebrate transposon cloning vectors: pKT2-SE, pPBT-SE, and pPPTn-SE for *Sleeping Beauty* (Clark *et al*., 2007; Ivics *et al*., 1997), *piggyBac* (Clark *et al*., 2007; Fraser *et al*., 1996), and *Passport* (Clark *et al*., 2009), respectively. The size of each plasmid (A–E) is noted. In cases where some of the plasmid is removed in the cloning process (A–C), the length of remaining elements is shown in parentheses. Abbreviations: ITR-R (Tol2 inverted terminal repeat right or 3′), ITR-L (Tol2 inverted terminal repeat left or 5′), DEST cassette (required for Gateway cloning, replaced with contents of entry vector), EF1α (elongation factor 1 alpha promoter), RBG intron (rabbit beta-globin intron), EGFP (enhanced green fluorescent protein), SV40(A) (SV40 polyadenylation signal), T7 (T7 polymerase binding site), T3 (T3 polymerase binding site), Lac Promoter (Lac operon promoter), ColE1 ORI (plasmid origin of replication), AmpR (ampicillin resistance gene), KanR (Kanamycin resistance gene), and F1 ORI (single-stranded phagemid origin). (See color plate.)

Urasaki *et al*., 2006). Each of these uses minimized ITRs, which help decrease the size of the transposon and plasmid. These *Tol2* cloning vectors have been used to produce transgenic zebrafish with good efficiency. The choice will come down to practicality and ease in generating the desired molecular biology cassette for use in zebrafish transgenesis experiments.

A key component of many zebrafish transgenesis vectors is the inclusion of a dominant reporter cassette for quality control during the transgenesis process and for downstream work with these lines. The role of the reporter is to provide an easy method of subsequent genotyping. Verification that transposition is occurring in the injected embryos is vital to successful transgenic fish production. This quality control check can be done quite easily when the expression cassette contains a dominant marker, like a fluorescent protein. If the primary transgene cassette will

not produce a dominant marker directly in zebrafish larvae (i.e., there is no dominant marker, the marker is targeted for adult expression, or the marker is inducible), it is often beneficial to include a small selectable cassette, such as the gamma-crystalin (γ-cryst) promoter or the cardiac myosin light chain (cmlc) driving a fluorescent protein in the lens or heart, respectively (Davidson *et al*., 2003; Huang *et al*., 2003). Not only will these expression cassettes help in the production of the desired transgenic fish line, but also they aid in husbandry by allowing easy selection of transgenic carriers as larvae.

B. Tol2 mRNA

Three widely available transcription vectors for making Tol2 transposase-encoding mRNA are shown in Fig. 4: pCS-TP (Kawakami *et al*., 2004) (Fig. 4A), pCS2-transposase (Kwan *et al*., 2007) (Fig. 4B), and pT3TS-Tol2 (Balciunas *et al*., 2006) (Fig. 4C). All three encode identical Tol2 transposase open reading frames. Both CS vectors, pCS2-TP and pCS-transposase, use the SP6 polymerase, share common linearization sites, and can potentially be used as an expression cassette with a CMV promoter and a functional SV40 poly(A) signal. However, the endogenous Tol2 UTRs (both 5′ and 3′) have been removed from pCS-transposase (Fig. 4B). The pT3TS-Tol2 vector differs from the CS vectors by using the T3 polymerase and incorporating UTRs from the *Xenopus* beta globin gene, sequences that have been shown to increase mRNA stability compared to CS-based vectors (Hyatt and Ekker, 1999). Ideally, synthetic, 5′-guanosine-capped mRNA should be produced using a kit such as Ambion's mMessage Machine®. *Tol2* transposition is robust and has a wide tolerance of transposon to transposase ratio. *Tol2* mRNA fidelity should be checked on an agarose gel (see Appendix C in Qiagen's mRNeasy kit for a convenient

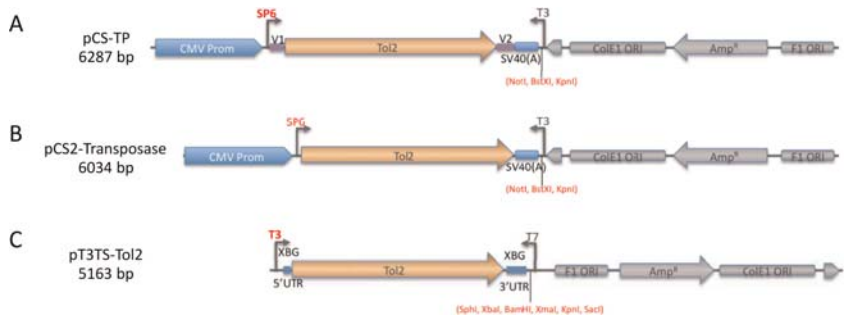


Fig. 4 Available Tol2 transposase transcription vectors. A and B) pCS-TP and pCS2-Transposase are very similar transcription vectors available from the Kawakami lab or Chien lab, respectively. Both are based on a pCS vector backbone and use the bacteriophage SP6 polymerase. pCS2-transposase lacks the V1 and V2 regions that correspond to cDNA from the native Tol2 transposase untranslated regions. C) pT3TS-Tol2 is available from the Ekker lab. It uses T3 polymerase to produce mRNA and includes untranslated regions from the *Xenopus* beta globin gene. The restriction sites available for linearization of the plasmid are shown in red. (See color plate.)

protocol, http://www.qiagen.com/hb/rneasymini), and care should be taken to avoid contamination with RNases. For these reasons, it is advisable to aliquot *Tol2* mRNA into single-use aliquots (for example 1 μL at 200 ng/μL, stored at −80 °C).

C. Positive Control Transposon

To assist in validating a newly constructed *Tol2* transposon, a functionally verified, positive-control transposon driving a fluorescent reporter can be requested from any of the sources of *Tol2* cloning vectors.

III. Microinjection

A. Injection Setup

An example set-up for microinjection in zebrafish is fully described (Bill *et al.*, 2009). Key components are the ability to visualize the procedure through the dissecting microscope, the micromanipulator control of the injection needle, and a volumetric regulator of the transposon DNA/transposase RNA solution during the injection process.

B. Final Preparation of Reagents

1. Transposon DNA: Transposase mRNA Injection Mix

a. On the morning of the injection, mix *Tol2* transposon DNA with an aliquot of *Tol2* mRNA. A recommended concentration is 12.5 ng/μL of both DNA and mRNA, diluted with RNase-free water as required. (It is important to use RNase-free reagents when isolating transposon plasmids initially and in the preparation of the DNA:RNA injection mix.)
b. The DNA:RNA injection mix is complete and should be handled with gloves and kept on ice for the duration of the injection session.

2. DNA:H_2O Control Injection Mix

A second injection mix is prepared that lacks *Tol2* mRNA (DNA:H_2O) to serve as an injection control to permit visualization of transposase activity (see quality control below). Prepare a DNA only solution using the same concentration of DNA used in the DNA:RNA injection mix (12.5 ng/μL).

C. Injecting

1. Calibrate the injection volume (such as described in Bill *et al*. (2009) or other suitable method) to inject 1–2 nL of DNA:RNA injection mix. Separately,

calibrate and inject DNA:H_2O injection mix into different sibling embryos (keep separate for comparison).
2. Enhanced rates of transgenesis are seen in embryos injected at the one cell stage. Higher volumes are tolerated in the yolk; however, transgenesis injections should be targeted to the cell/yolk interface or directly into the cell body for the highest efficiencies.
3. Injected embryos are transferred to Petri dishes and stored between 28 and 30 °C.
4. At the end of injection day, remove any dead or unhealthy embryos.

IV. Quality Control

The initial quality control step comes from expression of the marker that has been introduced into the transposon or cassette. Detection of the marker confirms the success of injection. Second, comparison of the DNA:RNA injections to control (DNA:H_2O) injections is also indicative of injection success. When performing control injections (DNA:H_2O), the number of fluorescently tagged embryos as well as the distribution (mosaicism) of fluorescence within the embryos will be dramatically reduced relative to co-injection of DNA with functional *Tol2* transposase mRNA.

A. Quality Control Fails – Troubleshooting

1. If no difference is noted between injections with or without transposase mRNA (i.e., both injections show low levels of fluorescence), then something has likely occurred to impair the transposition reaction. This is most often due to degradation of the *Tol2* mRNA, because it is the most sensitive component of the injection mix.
2. If marker expression is not observed in either injection condition (+/– *Tol2* mRNA), the integrity or identity of the transposon DNA is the likely first point for troubleshooting. The integrity of the DNA (or RNA) from the exact injection mix can be verified on an agarose gel if there is sufficient quantity left after the procedure. If the identity of the DNA transposon is in question, more thorough restriction digest analysis and/or sequencing of the transposon may be needed to determine the issue.

B. Quality Control Passed

1. Once *Tol2* mRNA-enhanced expression of the dominant marker is observed, the next step is to select and raise F0 embryos to produce transgenic fish.
2. Remove any embryos with obvious morphological defects as they are unlikely to produce healthy, fertile adults.

C. No Dominant Marker Used

If the choice was to forgo including a dominant marker within the transposon, then check the quality of the *Tol2* mRNA by co-injecting with a functionally verified positive control transposon obtained from one of the *Tol2* source labs or a colleague. Subsequent transgenesis will need to be scored by PCR or some other detection method.

V. Select Injected Fish to Raise as Founder Generation F0

A. Select 50–100 embryos that express the dominant marker in the appropriate context, e.g., heart expression from a cmlc promoter. If using a more ubiquitous promoter, most embryos will have some expression of a fluorescent protein. In this case, select the embryos with the most uniform (low mosaic) expression of the fluorescent protein, as they represent the embryos that have likely had early integration of a transposon. This increases the chances that there has been germline integration in these embryos.

B. Raise these selected embryos to produce F0 adults according to standard rearing protocols. Ideally, raise a minimum of 30 or so fish to adulthood to provide enough F0 adults to establish the desired transgenic fish line. However, while these fish are developing into adults, we recommend injecting another set within 2–4 weeks as backup in case something goes wrong with the first set (especially during rearing) or in case the injection was not as successful as first evaluated.

VI. Production of F1 Generation

A. Outcross F0 Fish

1. When F0 fish become sexually mature, usually between 2.5 and 4 months depending on rearing conditions, screen for germline transgenesis. Because the efficiency of transgenesis is generally high, it is desirable to outcross these fish with non-transgenic brood stock. Use fish that are easily distinguishable from F0 fish to prevent cross-contamination – leo^{tq270}/+ is one choice of a dominant visible marker with an altered pigmentation pattern in adult zebrafish (Watanabe *et al*., 2006), and there are many recessive pigment markers that can be used.
2. After obtaining eggs from an outcross of an F0, keep the adult fish separate for the week while determining whether or not it is transmitting the transposon through its germline.

B. Screen Embryos for Germline Transmission

1. Examine the embryos for dominant marker expression at the suitable time in development. In general, germline transgenesis is mosaic, meaning that expression will not be Mendelian at this generation. If the F0 is transmitting the transposon

through its germline, it is common to see about 10–15% of the embryos expressing the dominant marker. However, the actual transmittance can vary from a single embryo in a clutch to more than 50%; such higher expression frequencies are due to multiple transposon integrations in the founder fish germline.

2. Independent integration of transposons from one F0 to another often results in subtle (sometimes dramatic) differences in expression of the transgenic cassette used. Therefore, maintain expressing F1 embryos from different F0 fish as separate substrains. These siblings may represent a collection of multiple integration events within the mosaic F0 germline. The transposon integrations within these F1 embryos are now stable, and subsequent generations will inherit these in a Mendelian fashion. At this point, select a manageable number of substrains of the transposon (4–6) and trim this down in subsequent generations based on signal to noise and proper expression domains in the F2 embryos.
3. F0s that produce 40 or more viable embryos that do not show any expression of the dominant marker are considered negative and retired from the screen.

VII. Select the Best Substrain(s) Based on Observation of F2 Generation

The F2 embryos from F1 adults will inherit stable transposons in a Mendelian fashion. This permits evaluation of many embryos for proper expression based on the promoter used within the injected transposon. In addition to spatio-temporal control there will likely be differences in signal-to-noise ratio of the transgene. Select one or more lines to maintain.

VIII. Discussion

The use of transposons in zebrafish makes this animal perhaps the most readily modifiable organism within a biological scientists' genetic toolbox. Most injected animals pass at least one transgenic chromosome to their offspring. The *Tol2* system is robust and has been selected as the primary transposon platform, facilitating the development of an array of reagents readily shared within the zebrafish community.

Acknowledgments

This work was supported by the Mayo Foundation and the NIH (grants DA14546, GM63904, and DK84567).

References

Balciunas, D., Wangensteen, K. J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P. B., Largaespada, D. A., and McIvor, R. S., *et al*. (2006). Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet.* **2**, e169.

Baus, J., Liu, L., Heggestad, A. D., Sanz, S., and Fletcher, B. S. (2005). Hyperactive transposase mutants of the *Sleeping Beauty* transposon. *Mol. Ther.* **12**, 1148–1156.

Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. A., and Ekker, S. C. (2009). A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69–77.

Clark, K. J., Carlson, D. F., Foster, L. K., Kong, B., Foster, D. N., and Fahrenkrug, S. C. (2007). Enzymatic engineering of the porcine genome with transposons and recombinases. *BMC Biotechnol.* **7**, 42.

Clark, K. J., Carlson, D. F., Leaver, M. J., Foster, L. K., and Fahrenkrug, S. C. (2009). *Passport*, a native *Tc1* transposon from flatfish, is functionally active in vertebrate cells. *Nucleic Acids Res.* **37**, 1239–1247.

Cui, Z., Geurts, A. M., Liu, G., Kaufman, C. D., and Hackett, P. B. (2002). Structure-function analysis of the inverted terminal repeats of the *Sleeping Beauty* transposon. *J. Mol. Biol.* **318**, 1221–1235.

Davidson, A. E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., Sivasubbu, S., Cliff, M. P., Hackett, P. B., and Ekker, S. C. (2003). Efficient gene delivery and gene expression in zebrafish using the *Sleeping Beauty* transposon. *Dev. Biol.* **263**, 191–202.

Emelyanov, A., Gao, Y., Naqvi, N. I., and Parinov, S. (2006). Trans-kingdom transposition of the maize dissociation element. *Genetics* **174**, 1095–1104.

Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., Urasaki, A., Kawakami, K., and McCallion, A. S. (2006). Evaluating the biological relevance of putative enhancers using *Tol2* transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305.

Fraser, M. J., Ciszczon, T., Elick, T., and Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of *Lepidoptera*. *Insect. Mol. Biol.* **5**, 141–151.

Geurts, A. M., Yang, Y., Clark, K. J., Liu, G., Cui, Z., Dupuy, A. J., Bell, J. B., Largaespada, D. A., and Hackett, P. B. (2003). Gene transfer into genomes of human cells by the *Sleeping Beauty* transposon system. *Mol. Ther.* **8**, 108–117.

Huang, C., Tu, C., Hsiao, C., Hsieh, F., and Tsai, H. (2003). Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**, 30–40.

Hyatt, T. M., and Ekker, S. C. (1999). Vectors and techniques for ectopic gene expression in zebrafish. *Methods Cell Biol.* **59**, 117–126.

Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvák, Z. (1997). Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510.

Izsvák, Z., Ivics, Z., and Plasterk, R. H. (2000). *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* **302**, 93–102.

Kawakami, K., Koga, A., Hori, H., and Shima, A. (1998). Excision of the *Tol2* transposable element of the medaka fish, *Oryzias latipes*, in zebrafish, *Danio rerio*. *Gene* **225**, 17–22.

Kawakami, K., Shima, A., and Kawakami, N. (2000). Identification of a functional transposase of the *Tol2* element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. U. S. A.* **97**, 11403–11408.

Kawakami, K., and Shima, A. (1999). Identification of the Tol2 transposase of the medaka fish *Oryzias latipes* that catalyzes excision of a nonautonomous *Tol2* element in zebrafish *Danio rerio*. *Gene* **240**, 239–244.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144.

Koga, A., Cheah, F. S. H., Hamaguchi, S., Yeo, G. H., and Chong, S. S. (2008). Germline transgenesis of zebrafish using the medaka *Tol1* transposon system. *Dev. Dyn.* **237**, 2466–2474.

Koga, A., Suzuki, M., Inagaki, H., Bessho, Y., and Hori, H. (1996). Transposable element in fish. *Nature* **383**, 30.

Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P., and Chien, C. (2007). The Tol2kit: a multisite gateway-based construction kit for *Tol2* transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099.

Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., and Matrai, J., *et al*. (2009). Molecular evolution of a novel hyperactive *Sleeping Beauty* transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41**, 753–761.

Parinov, S., Kondrichin, I., Korzh, V., and Emelyanov, A. (2004). Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes *in vivo*. *Dev. Dyn.* **231**, 449–459.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353.

Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science* **218**, 341–347.

Stuart, G. W., McMurray, J. V., and Westerfield, M. (1988). Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**, 403–412.

Urasaki, A., Morvan, G., and Kawakami, K. (2006). Functional dissection of the *Tol2* transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649.

Villefranc, J. A., Amigo, J., and Lawson, N. D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev. Dyn.* **236**, 3077–3087.

Watanabe, M., Iwashita, M., Ishii, M., Kurachi, Y., Kawakami, A., Kondo, S., and Okada, N. (2006). Spot pattern of leopard *Danio* is caused by mutation in the zebrafish *connexin41.8* gene. *EMBO Rep.* **7**, 893.

Yant, S., Park, J., Huang, Y., Mikkelsen, J., and Kay, M. (2004). Mutational analysis of the N-terminal DNA-binding domain of *Sleeping Beauty* transposase: critical residues for DNA binding and hyper-activity in mammalian cells. *Mol. Cell. Biol.* **24**, 9239.

Zayed, H., Izsvák, Z., Walisko, O., and Ivics, Z. (2004). Development of hyperactive *Sleeping Beauty* transposon vectors by mutational analysis. *Mol. Ther.* **9**, 292–304.

CHAPTER 9

Spatiotemporal Control of Embryonic Gene Expression Using Caged Morpholinos

Ilya A. Shestopalov and James K. Chen

Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California, USA

Abstract

Embryonic development depends on spatial and temporal control of gene function, and deciphering the molecular mechanisms that underlie pattern formation requires methods for perturbing gene expression with similar precision. Emerging chemical technologies can enable such perturbations, as exemplified by the use of

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00009-4

caged morpholino (cMO) oligonucleotides to photo-inactivate genes in zebrafish embryos with spatiotemporal control. This chapter describes general principles for cMO design and methods for cMO assembly in three steps from commercially available reagents. Experimental techniques for the microinjection and photoactivation of these reagents are described in detail, as well as the preparation and application of caged fluorescein dextran (cFD) for labeling irradiated cells. Using these protocols, cMOs can be effective tools for functional genomic studies in zebrafish and other model organisms.

I. Introduction

During embryonic patterning, genetic programs are precisely coordinated to create complex tissues and organs. Genome sequencing and forward genetic screens have revealed an extensive list of patterning genes, many of which are expressed in a tissue-specific manner at discrete time points in embryo development. One of the remaining challenges in developmental biology is to understand how these genes act in space and time to coordinate embryogenesis by stereotypically modulating cellular functions. Toward that goal, several chemical, optical, and/or genetic approaches for conditional gene regulation have been developed, and such technologies have provided key insights into the molecular mechanisms that underlie tissue patterning and function (Gradinaru *et al.*, 2010; Ouyang and Chen, 2010; Shestopalov and Chen, 2008).

As evident in this issue of Methods in Cell Biology, the zebrafish is ideally suited for visualizing vertebrate ontogeny, since its embryos and larvae are optically transparent and develop rapidly *ex utero.* However, methods for regulating endogenous gene function in zebrafish are underdeveloped relative to other model organisms; inducible RNA interference technologies have exhibited limited efficacy thus far (Blidner *et al.*, 2008; Zhao *et al.*, 2008), and targeted genetic knockouts are limited to zinc-finger nucleases (Remy *et al.*, 2010). In lieu of these approaches, synthetic oligonucleotides such as morpholinos (MOs) have been employed as antisense reagents in zebrafish embryos (Shestopalov and Chen, 2010). MO oligomers, typically 25 nucleotides in length, display DNA bases from a morpholine ring system and are connected by a phosphorodiamidate backbone (Summerton, 1999). Due to this non-natural structure, MOs are resistant to nucleases and persist in zebrafish embryos for up to four days (Bill *et al.*, 2009; Nasevicius and Ekker, 2000). When MOs are injected into zebrafish prior to the eight-cell stage, they become uniformly distributed throughout the embryo and constitutively block either RNA splicing or translation, depending on the targeted RNA sequence.

While MOs are widely used to interrogate gene function in zebrafish, the utility of these reagents is limited by their constitutive activity, as gene expression is inhibited immediately after microinjection. Conventional MOs therefore are less effective for

studying genes that are required for embryonic survival and/or have pleiotropic functions during embryonic patterning. To overcome this limitation, we recently developed caged morpholino (cMOs) that can be activated by 360-nm light, taking advantage of the transparency of zebrafish embryos (Shestopalov *et al*., 2007). This was achieved by tethering a complementary MO-derived inhibitor to the 25-base targeting sequence through a dimethoxynitrobenzyl (DMNB) group-containing photocleavable linker (Fig. 1). The resulting intramolecular hairpin suppresses hybridization of the targeting sequence to its complementary RNA, and cMOs are significantly less active *in vivo* than their 25-base counterparts. Linker cleavage with 360-nm light converts the cMO hairpin into an intermolecular MO duplex that is energetically favored to dissociate, allowing the targeting MO to alter RNA splicing or translation. Similar caging strategies have been previously described for regulating RNA function *in vivo* (Tang *et al*., 2007, 2008), and this general approach differs from other oligonucleotide-caging technologies that target individual nucleoside bases (Young *et al*., 2008), the phosphate backbone (Ando *et al*., 2001), the oligonucleotide termini (Nguyen *et al*., 2006), or use an excess of complementary caged RNA oligomer (Tomasini *et al*., 2009). The cMO technology is advantageous to other methods for light-controlled antisense as it uses the well-established MO antisense scaffold, a single photo-cleavable group, and a stoichiometric amount of MO-regulating oligonucleotide (Shestopalov and Chen, 2010). Cells containing photoactivated cMOs can be traced with a variety of light-inducible fluorescent proteins (Lippincott-Schwartz and Patterson, 2008) or dextran-conjugated caged fluorophores, such as caged fluorescein dextran (cFD) (Kozlowski and Weinberg, 2000) or caged rhodamine dextran (Gee *et al*., 2001).

While our initial published results established the general principle of using caged oligonucleotides to conditionally regulate *in vivo* gene expression (Shestopalov *et al*., 2007), there were limitations to generalizable cMO implementation. First, our initial studies required preparation of the inhibitory MO oligomer and its appendant photocleavable linker through solid-phase chemistry, since conventional MOs amenable to functionalization at the 5' end were not commercially available at that time. These procedures are laborious and time-consuming,

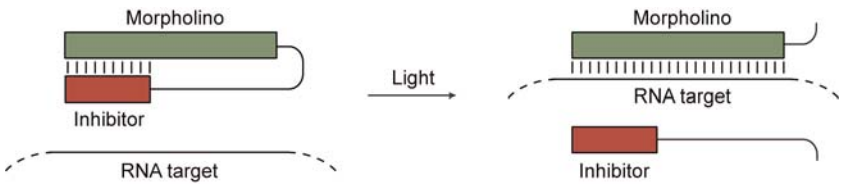


Fig. 1 Schematic representation of the cMO hairpin technology. MO/RNA hybridization can be abrogated by tethering a complementary inhibitor MO via a photocleavable linker. Irradiation with 360-nm light leads to inhibitor dissociation, allowing the targeting MO to base pair with RNA containing a complementary sequence. Adapted with permission (Ouyang *et al*., 2009; Copyright 2009, American Chemical Society). (See color plate.)

encumbering the synthesis and evaluation of other cMOs. Second, generalizable guidelines for the design of hairpin cMOs were not evident, as only one inhibitory sequence and structural configuration was tested. To resolve these challenges we synthesized a DMNB-based bifunctional linker that can be used to conjugate the targeting MO and its complementary inhibitor in only three steps, starting with appropriately functionalized, commercially available MO oligomers (Ouyang *et al*., 2009). We have utilized this optimized synthetic route to systematically analyze the *in vitro* thermodynamics and *in vivo* efficacy of various cMO hairpin structures and established design criteria for optimizing cMO activity profiles. These advances have enabled us to prepare photoactivatable reagents targeting the zebrafish genes *no tail a* (*ntla*) (Halpern *et al*., 1993), *heart of glass* (*heg*) (Mably *et al*., 2003), *floating head* (*flh*) (Talbot *et al*., 1995), *ets variant gene 2* (*etv2*) (Sumanas and Lin, 2006), *SRY-box containing gene 32* (*sox32*) (Alexander *et al*., 1999), and $\alpha - catenin$ (*ctnna*) (Cerda *et al*., 1999). We further demonstrated the versatility of our caging approach by replacing the DMNB chromophore with a bromohydroxyquinoline (BHQ) group, which has a significantly greater cross section for two-photon excitation (Fedoryak and Dore, 2002; Ouyang *et al*., 2009).

In this chapter, we describe detailed methods for the design, synthesis, and utilization of cMO hairpins based on our most recent publication (Ouyang *et al*., 2009). We also describe synthesis of cFD from commercially available precursors for co-injection with cMOs to enable tracking of the irradiated cells.

II. Design and Synthesis of cMOs

A. Design of cMOs

As described above, cMOs are composed of a 25-base, RNA-targeting MO tethered to a complementary MO oligonucleotide, thereby generating a stem-loop structure through intramolecular base pairing that abrogates its binding to RNA targets (Fig. 1). Linker photolysis converts the cMO hairpin into a MO duplex with significantly weaker, intermolecular base-pairing interaction that permits duplex dissociation and MO/RNA hybridization. For optimum performance, the cMO hairpin should therefore exhibit certain biophysical properties. There must be enough intramolecular binding energy within the hairpin to minimize strand exchange and intermolecular binding to RNA. However, cMO hairpins with too much intramolecular binding energy, once photolyzed, will form MO duplexes with significant intermolecular binding energy and strand exchange with RNA will be inhibited. This delicate balance of binding energy also underscores the requirement for an inhibitory oligomer with a MO backbone, as cMOs containing DNA or RNA inhibitors can exhibit high basal gene-silencing activities, presumably due to hydrolysis of the phosphate backbone *in vivo* (S. Sinha and J. K. Chen, unpublished observations).

To establish general principles for cMO design, we systematically investigated the thermodynamic properties of MO/MO and MO/RNA base pairing, both *in vitro* and *in*

vivo. The intra- and intermolecular interactions of MO duplexes were compared, and as a result of these studies, Eq. 1 was developed as a general formula for calculating the melting temperature (T_m) for intermolecular MO duplexes (Ouyang *et al*., 2009).

$$T_m = 1.9 \times (\#\text{ of A/T base pairs}) + 5.7 \times (\#\text{ of G/C base pairs}) \quad (1)$$

We find that this simple formula holds true for MO duplexes containing up to 16 base pairs, and longer duplexes may need further analysis of nearest-neighbor effects for accurate T_m estimation. We also determined that cMO hairpins exhibit optimum efficacy *in vivo* when they are based on MO duplexes with a T_m of 41–44 °C and adopt blunt-ended rather than staggered stem-loops. For example, we found that an optimum cMO hairpin against *ntla* should be made from the following MOs:

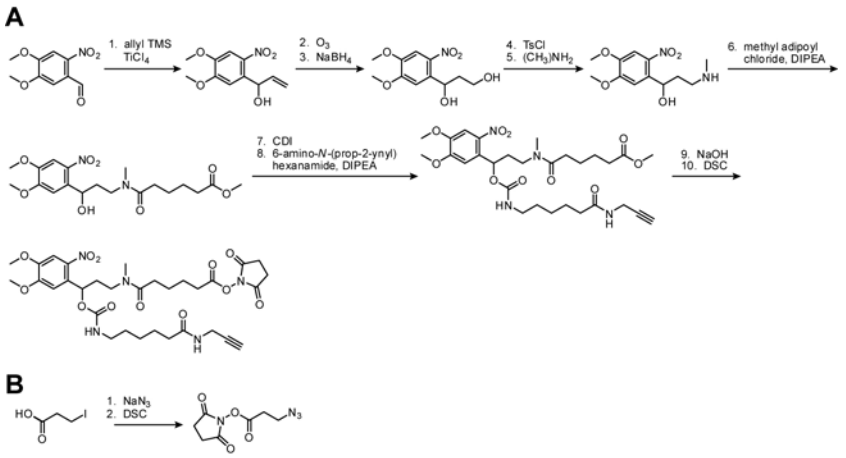
ntla MO: 5'-GACTTGAGGCAGACATATTTCCGAT-3'
Inhibitor MO: 5'-GCCTCAAGTC-3'

$$T_m = 1.9 \times 4 + 5.7 \times 6 = 41.8\,^{\circ}\text{C}$$

For other MO sequences, bases can be added or removed from the 5' end of the inhibitor MO to achieve an MO duplex T_m of 41–44 °C, along with the removal of the 3'-most base in the inhibitor sequence, if necessary. Typical inhibitor MOs are 10–13 bases in length.

B. Synthesis of a Propargyl-Functionalized, Photocleavable Crosslinker

The bifunctional, photocleavable crosslinker can be synthesized from commercially available 6-nitrovertaldehyde in ten steps (Scheme 1A). Detailed procedures



Scheme 1 Crosslinker syntheses. Synthetic procedures for preparing the propargyl-functionalized, photocleavable crosslinker (A) and azide-functionalized crosslinker (B). TMS = trimethylsilane, Ts = tosyl, DIPEA = diisopropyl ethyl amine, CDI = carbonyl diimidizole, DSC = disuccinimidyl carbonate. Adapted with permission (Ouyang *et al*., 2009; Copyright 2009, American Chemical Society).

for steps 1–5 (Shestopalov *et al*., 2007) and 6–10 (Ouyang *et al*., 2009) have been published, and the synthetic yield for step 5 can be further improved by performing the reaction in a pressure tube at 80 °C. Although the *cis* and *trans* forms of the crosslinker carbamate (compounds after step 7) are separable by silica gel chromatography, it is not necessary to isolate the individual stereoisomers since they will slowly interconvert. The final product was column-purified to remove all trace impurities and stored as a lyophilized solid in single-reaction aliquots (1.1 mg for reactions on 100 nmol scale) at −20 °C to prevent NHS ester hydrolysis during freeze-thaw cycles. All reactions and synthetic manipulations were done under ambient light without noticeable decomposition, but the photolabile carbamates were shielded from light and kept at −20 °C during long-term storage.

C. Synthesis and Purification of cMOs

1. Synthesis of an Azide-Functionalized Crosslinker

Preparation of an azide- and NHS ester-containing crosslinker can be accomplished in two steps from commercially available 3-iodopropionic acid (Scheme 1B), and detailed synthetic procedures have been published (Shestopalov *et al*., 2007). The final product was stored dry in single-use aliquots (0.5 mg for 100-nmol-scale reactions) at −20 °C to minimize hydrolysis of the activated ester.

2. NHS Ester Coupling with Amine-Modified MO Oligomers

To construct cMO hairpins, MO oligomers with a 3' amine modification and inhibitory MO oligomers with a 5' amine modification were obtained from GeneTools, LLC in 300-nmol quantities and acylated with the appropriate NHS-containing linkers (Scheme 2). Each MO was dissolved in 300 μL of water. In general, we recommend using a water bath sonicator to help dissolve MOs. Actual oligomer concentration was determined according to the manufacturer's instructions by diluting 2 μL of oligomer solution 40-fold in 0.1 M HCl, heating to 100 °C for 1 min to dissociate aggregates, cooling to ambient temperature, and measuring absorbance at 265 nm using a NanoDrop™ UV-VIS spectrometer. MO solutions prepared in this manner are typically in the 0.7–0.9 mM concentration range.

All MO conjugation reactions and manipulations were carried out in standard 1.5-mL polypropylene microcentrifuge tubes. Prior to acylation, the amine-functionalized MOs were lyophilized to dryness, dissolved in 0.1 M $Na_2B_4O_7$ (pH 8.5), heated to 100 °C for 1 min to dissociate aggregates, and allowed to cool to ambient temperature. To prepare the 3' azide 25-base targeting MO, one 0.5-mg aliquot of NHS ester-functionalized azide was dissolved in 20 μL of DMSO and added to the aqueous solution containing 100 nmol of amine-modified MO in 0.1 M $Na_2B_4O_7$ buffer. The 5' amine-modified inhibitor MO was functionalized with the DMNB-containing crosslinker in a similar manner; one 1.1-mg aliquot of the propargyl-functionalized NHS ester was dissolved in 15 μL

Scheme 2 cMO synthesis. (i) Azide-functionalized crosslinker, 0.1 M sodium borate (pH 8.5), DMSO, 80–90% after purification. (ii) Propargyl-functionalized, photocleavable crosslinker, 0.1 M sodium borate (pH 8.5), DMSO, 70–90% after purification. (iii) sodium ascorbate, TBTA, CuI, 0.1 M potassium phosphate (pH 8.0), DMSO, 10–25% after HPLC purification. Adapted with permission (Ouyang *et al*., 2009; Copyright 2009, American Chemical Society). (See color plate.)

of DMSO and added to the buffered solution containing 100 nmol of the inhibitor MO. In this latter case, a cloudy suspension initially forms and later disappears as the reaction proceeds. Coupling reactions were vortexed overnight in foil-wrapped microcentrifuge tubes, resulting in quantitative conversion to the acylated MOs.

3. Purification of Functionalized MO Oligomers

The 3' azide-functionalized targeting MO was purified from unreacted NHS ester and its hydrolyzed byproduct by passing the reaction mixture through a disposable size-exclusion column. One illustra NAP^{TM}-10 column (GE Life Sciences 17-0854) was typically used for each 100 nmol-scale reaction mixture, according to the manufacturer's instructions. Columns were equilibrated with water and the reaction mixture was diluted to 1 mL with water prior to loading onto the column. The MO-containing eluent was lyophilized to dryness and dissolved in 100 μL water, and the final MO concentration was determined as described in Section C2. This purification method typically yielded at least 50 nmol of azide-functionalized MO.

The inhibitor MO-crosslinker conjugate was more difficult to purify as the relatively large DMNB-containing linker is not completely removed by size-exclusion column chromatography. Therefore, following product elution from the NAP^{TM}-10 column and lyophilization (as described above for the 25-base targeting MO), the

inhibitor MO was dissolved in 400 μL of water and acidified with 4 μL of glacial acetic acid to protonate any hydrolyzed linker. The aqueous solution was washed three times with chloroform (400 μL) and twice with ethyl acetate (400 μL) to remove any unreacted or hydrolyzed crosslinker. These washes were performed in standard microcentrifuge tubes by vortexing the aqueous/organic mixture for 30 s followed by centrifugation (16,000 *g*, 1 min) and disposal of the organic layer. The aqueous/organic interface, often containing cloudy precipitate, was carried through to increase purification yield. Care should be taken not to dispose of the MO-containing aqueous layer during the ethyl acetate washes, as unlike chloroform, ethyl acetate is less dense then water. Following the organic washes, the aqueous layer acidity was quenched by adding 20 μL of 10% ammonium hydroxide, and MO solution alkalinity was confirmed by spotting 0.5 μL of the aqueous solution on pH paper. Failure to basify the MO solution results in oligonucleotide degradation during lyophilization. The mixture of MO, ammonium acetate, and water was lyophilized to dryness, and residual ammonium acetate was removed by two cycles of MO solubilization in 200 μL of water and lyophilization to dryness. The purified inhibitory MO was then dissolved in 400 μL of 0.1 M KH_2PO_4 (pH 8), and oligomer concentration was spectroscopically quantified as described in Section C2. This purification method typically yielded at least 40 nmol of DMNB-functionalized inhibitor MO.

Derivatization of MOs with the azide- or propargyl-functionalized linkers should be confirmed by mass spectrometry prior to click-chemistry conjugation. We used a Waters MicroMass ZQ liquid chromatography-mass spectrometry (LC-MS) system with a Zorbax^{TM} SB-C18 2.1 × 30 mm column (Agilent 823700-902) and a two-solvent gradient system (solvent A: water with 0.1% formic acid; solvent B: CH_3CN with 0.1% formic acid; 2–95% B in 20 min; flow rate of 0.3 mL/min). The inhibitor MO modified with the propargyl-containing, photocleavable linker typically had two masses from the same LC peak due to sample photolysis as it passed through the UV flow cell prior to mass analysis. Accurate mass spectral data could be obtained with as little as 0.5 nmol of MO oligomer.

4. Assembly of cMOs via Azide–Alkyne Huisgen Cycloaddition

Azide-alkyne Huisgen cycloaddition, commonly referred to as "click chemistry", was used to assemble cMO hairpins (Scheme 2) (Rostovtsev *et al*., 2002). For this reaction, the azide-functionalized 25-base MO was lyophilized to dryness, to which an equimolar amount of the propargyl-functionalized inhibitor MO dissolved in KH_2PO_4 buffer (pH 8) was added. The pH of the resulting solution was checked after mixing, and 1–2 μL of 1 M KOH was added to maintain a pH of 8 if necessary. The MO mixture was heated to 100 °C for 1 min to dissociate oligonucleotide aggregates and then cooled to room temperature. Solutions of 20 mM CuI in DMSO and 20 mM sodium ascorbate in 0.1 M KH_2PO_4 (pH 8) were prepared immediately prior to start of the reaction, whereas a stock solution of 20 mM tris [(1-benzyl-1H-1,2,3-triazol-4-yl) methyl]amine (TBTA) in DMSO was prepared

ahead of time and stored at −20 °C between uses. The azide-alkyne coupling reaction was initiated by the addition of sodium ascorbate, TBTA, and CuI, in that order (25 μL of each 20 mM stock solution per 50 nmol-scale reaction), to the 1.5 mL microcentrifuge tube containing the solution of the two MOs. The resulting reaction mixture was briefly sonicated, protected from light with aluminum foil, and then stirred overnight with a small magnetic stir bar.

5. Purification of cMOs

After completion of the azide-alkyne coupling reaction, the stir bar was removed and consumed, insoluble catalyst was pelleted by centrifugation (16,000 *g*, 1 min). The supernatant was then desalted by passage through a size-exclusion column in analogy to the procedure described in Section C3. For this step we typically used one GE Life Sciences illustra NAP^{TM}-5 column per 25 nmol of MO reaction mixture according to the manufacturer's instructions. The columns were equilibrated with water and the reaction mixture was diluted to 0.5 mL with water prior to sample loading onto the column. The crude cMO-containing eluent was lyophilized to dryness and dissolved in 100 μL of 0.02 M NaOH. The pH of the resulting solution was checked, and 1–3 μL of 1 M NaOH was added if necessary to achieve a pH greater than 12.

Our inability to resolve MO mixtures by reverse-phase and size-exclusion high-pressure liquid chromatography (HPLC) led us to develop a method for the efficient separation of cMO hairpins from unconjugated MOs using ion-exchange HPLC. Though MOs are uncharged oligomers, adenine and thymine morpholino bases become deprotonated (and negatively charged) at pH 11 and higher, allowing for HPLC separation with a DNAPac^{TM} PA100 22 × 250 mm column (Dionex 043011) and a two-solvent gradient system (solvent A: 0.02 M NaOH, 1% CH_3CN; solvent B: 0.02 M NaOH, 0.375 M $NaClO_4$, 1% CH_3CN). The CH_3CN co-solvent was used to decrease hydrophobic interactions between MOs and the column media, while the large perchlorate anion was used to effectively compete with ionic MO/media interactions.

To purify the reaction mixture, 2 μL of the crude cMO product was first resolved on the DNAPac^{TM} PA100 column by running a 27-min 7–50% solvent B gradient with a flow rate of 4 mL/min. As expected, MOs eluted in order of A/T base content with the cMO eluting last (Fig. 2A). MO absorbance was detected at 260 nm, and the photocleavable chromophore was detected at 347 nm. Elution time for the last unconjugated MO peak (Point X, Fig. 2A) and the beginning of the cMO hairpin peak (Point Y, Fig. 2A) was noted and converted to gradient composition (percentage of solvent B) at X and Y. An empirically derived correction factor was then subtracted from both X and Y to compensate for the solvent volume between gradient mixer and UV flow cell (7% solvent B for our HPLC system), to give the values X* and Y* (Fig. 2B). An elution method was then programmed consisting of: 7 to X*% solvent B in 5 min, X* to Y*% B in 10 min, Y* to 50% B in 1 min, and 50% B for 10 min (Fig. 2B). The 10-min pause between gradient points X* and Y* is intended

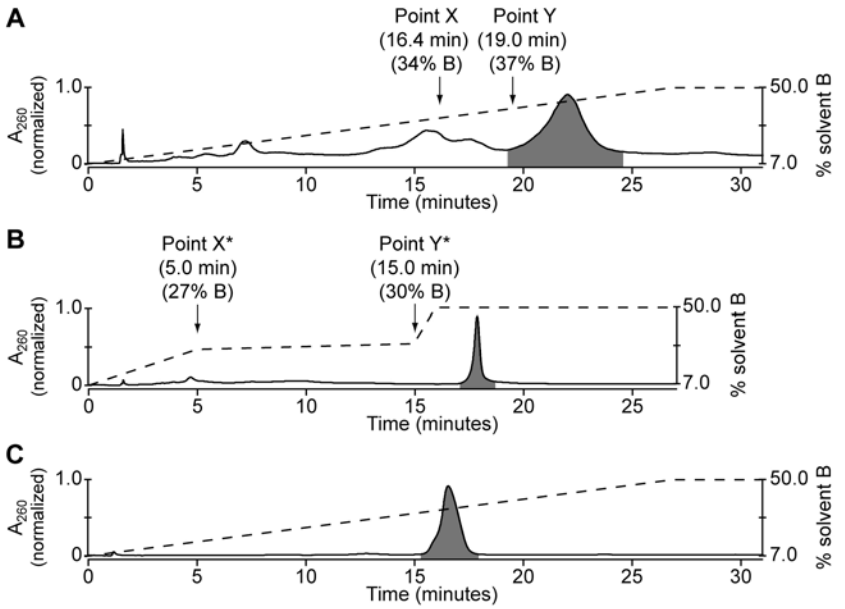


Fig. 2 Purification of a cMO hairpin. (A) Chromatogram of a crude cMO mixture separated by preparative ion-exchange HPLC using a continuous solvent gradient. MO oligomers eluted in order of increasing adenine and thymine content, with the cMO eluting last (highlighted in gray). The dashed line represents gradient composition (% solvent B), and the retention time and gradient composition associated with points X and Y are noted. (B) Chromatogram of the same crude cMO mixture separated by preparative ion-exchange HPLC using a step-wise solvent gradient and empirically derived transition points X* and Y*. The cMO eluted as a symmetric peak with a half-height temporal width of approximately 20 s. (C) Chromatogram of the cMO isolated by the step-wise solvent gradient in (B) and analyzed by analytical ion-exchange HPLC using a continuous solvent gradient. The cMO purity was determined to be 98.8% and its molecular mass was confirmed by LCMS.

to allow the elution of all unconjugated MOs, while the sharp gradient ramp from point Y* to 50% B should elute the cMO in one narrow peak. The composed method was verified using another 2 μL of the crude MO product (Fig. 2B). The unconjugated MO peak should be completely eluted from the column before the cMO hairpin is obtained as one symmetrical peak with a half-height temporal width of 30 s or less. Gradient parameters X* and Y* were fine-tuned to meet these criteria, if needed.

Once a separation method was developed, up to 50 nmol of the MO mixture was loaded onto the column using a 100-μL injection loop. It is important to turn off the UV-VIS detection system during this purification step as UV light readily photolyzes the cMO hairpin. Fractions were collected every 15 s starting at the 16-min point in the HPLC run, with cMOs typically eluting in 2–3 fractions. cMO-containing fractions were identified by UV absorbance and buffered by addition of 40 μL of 1 M NH_4OAc (pH 5). After each fraction was confirmed to have a final pH of less than 7,

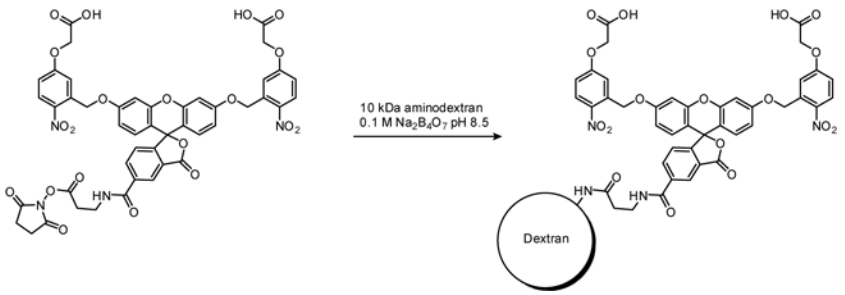
they were combined and lyophilized to dryness. Failure to buffer these fractions resulted in linker cleavage via base-catalyzed elimination of the triazole group.

Purified cMOs were redissolved in 500 μL of water and desalted using one illustra NAP^{TM}-5 column (GE Life Sciences 17-0853) according to the manufacturer's instructions. The columns were equilibrated with water prior to sample loading. The cMO-containing eluent was lyophilized to dryness, dissolved in 50 μL of water, and centrifuged (16,000 *g*, 2 min) to remove any particulates derived from size-exclusion medium. The cMO was then precipitated with 400 μL of acetone, vortexed, and centrifuged (16,000 *g*, 20 min, 4 °C). The precipitated cMO usually appeared as an oil in the microcentrifuge tubes that was difficult to see, and care was taken to remove most of the acetone-containing supernatants without touching the walls of the microcentrifuge tube. The cMO precipitate was then washed with 100 μL of CH_3CN, briefly sonicated, and centrifuged (16,000 *g*, 5 min, 4 °C). The CH_3CN was aspirated and the off-white cMO pellet was lyophilized for 10 min to remove trace solvent. The cMO pellet was then dissolved in 100 μL of water, heated at 100 °C for 30 s to dissociate aggregates, cooled to room temperature, and cMO concentration was determined by the 265-nm absorbance of an aliquot dissolved in 0.1 M HCl. This purification typically yielded 5–10 nmol of pure cMO. The cMO solution should be centrifuged again (16,000 *g*, 20 min) at room temperature to pellet insoluble particulates that can clog microinjection needles. Supernatant containing cMO was then transferred into an O-ring-capped 1.5 mL microcentrifuge tube, reduced in volume to 20 μL to facilitate the preparation of injection solutions, and stored in the dark at −20 °C. Brief exposures of the cMO solution to ambient light did not cause detectable photolysis of the cMO linker.

Prior to its use *in vivo*, cMO mass and purity should be confirmed by LC-MS and ion-exchange HPLC. LC-MS was performed as described in Section C3. Ion-exchange HPLC was performed using a DNAPac^{TM} PA200 4 × 250 mm column (Dionex 063000) using a gradient of 7–50% solvent B in 27 min with a flow rate of 1.2 mL/min, using the two-solvent system described in Section C5. At least 0.3 nmol of cMO should be used for HPLC analysis, and a greater than 98% purity should be observed (Fig. 2C).

III. Synthesis of cFD

Since cFD is no longer commercially available, we prepared this reagent using carboxymethylnitrobenzyl (CMNB)-caged fluorescein NHS ester (Invitrogen C20050, Invitrogen is headquartered in Carlsbad, CA) and 10-kDa aminodextran (Invitrogen D1860) (Scheme 3). Approximately 3.5–5 mg of the aminodextran dissolved in 500 μL 0.1 M $Na_2B_4O_7$ buffer (pH 8.5) was added to 1 mg of caged fluorescein in its Invitrogen-supplied, tinted microcentrifuge tube, and the reaction mixture was vortexed overnight. The resulting cFD was separated from unreacted caged fluorescein using a Zeba Desalt spin column (Pierce 89889) according to the manufacturer's instructions. The yellow-colored eluent was lyophilized to dryness,



Scheme 3 **cFD synthesis**. cFD was obtained in 68% yield after purification, with an average loading of 2.5 caged fluorescein molecules per 1 molecule of amminodextran.

weighed, dissolved in water to make a 1% (w/v) stock solution, and stored at −20 °C in 2-μL aliquots to avoid detrimental freeze-thaws. Spectroscopic analysis indicated an average loading of 2.5 caged fluorescein molecules per 1 molecule of dextran. The 1% cFD solution was typically diluted 10–20-fold for microinjection into zebrafish embryos.

IV. Microinjection of Caged Reagents

Injection solutions of caged reagents were prepared in Milli-Q water containing 0.1% phenol red and 100 mM KCl. cMO microinjection solutions (with or without cFD) were heated to 100 °C for 15 s to dissociate MO aggregates and centrifuged (16,000 *g*, 2 min, 4 °C). As with conventional MOs, cMOs and cFD were microinjected either into the animal cell or the yolk prior to eight-cell stage using previously described procedures (Ekker, 2004).

Accurate microinjection dosing is essential for the successful use of cMOs, as over-dosing will recapitulate weak morphant phenotypes from hybridization between the cMO hairpin and RNA. Therefore, O-ring capped microcentrifuge tubes should always be used for handling injection solutions to avoid changes in concentration due to evaporation. The microinjection volume should also be calibrated to deliver accurate cMO doses. We recommend quantifying injection volume by the mineral oil droplet method (Nüsslein-Volhard and Dahm, 2002). In this procedure, mineral oil is placed in a glass depression slide and droplets of reagent-containing injection solution are microinjected into the oil. A stereoscope equipped with a micrometer is then used to measure the diameter of each spherical droplet, allowing the injection volume to be calculated. We have not observed photolysis of cMOs or cFD from the incandescent light source on the microinjection stereoscope, though uncaging can occur with more sensitive photocleavable groups, such as the BHQ chromophore in our two-photon-compatible cMOs. To minimize compound photolysis during the microinjection procedure, the stereoscope stage can be covered with UV light-absorbing plastic, such as the 3M™ Sun Control window film.

V. Global Photoactivation of Caged Reagents

A. Embryo Mounting and Irradiation

Appropriate UV irradiation protocols are critical for reproducible photoactivation of caged reagents. The light source must emit photons of the correct wavelength to drive the photochemical reaction over the threshold activation energy. Light must also have sufficient irradiance (intensity) to attain a high probability of collision between photons and all photo-reactive molecules, thereby completing the photochemical reaction within seconds while minimizing UV damage. To achieve these conditions, we typically performed global irradiations on a Leica DM4500B upright compound epifluorescent microscope equipped with an HBO 100 W mercury short-arc lamp, an A4 filter cube (360 ± 40 nm; Leica 11513874) and a $20\times$ water-immersion objective (0.5 numerical aperture (NA); Leica 506147). We have found that irradiation with a fluorescent stereoscope (such as the Leica MZFLIII or M205FA) or a handheld UV lamp does not provide enough light intensity for efficient photoactivation. Uniform global photoactivation of caged molecules is also difficult to achieve on a confocal microscope due to the limited depth of field and short dwell time as the laser beam scans across the sample.

We have also observed that embryo placement and orientation significantly influences uncaging efficiencies. To facilitate zebrafish embryo irradiation, we typically immobilized chorionated embryos in agarose microinjection templates, orienting them on a stereoscope so that the animal cells face upward. The embryos were then irradiated one at a time on the Leica DM4500B microscope (Fig. 3A). We used microinjection templates with individual, evenly spaced wells as opposed to troughs, leaving an empty well between embryos to avoid inadvertent exposure to scattered UV light. When globally irradiating blastula-stage embryos, some rotated out of the desired animal pole orientation and had to be re-oriented on the stereoscope.

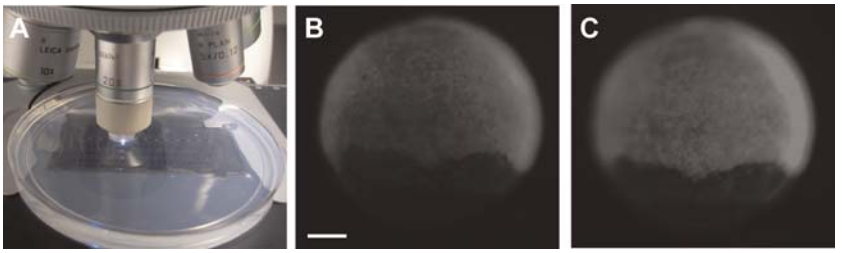


Fig. 3 Global irradiation of zebrafish embryos. (A) Whole-embryo irradiation of zebrafish mounted in an agarose template using a Leica DM4500B upright fluorescence microscope equipped with a $20\times/0.5$ NA water-immersion objective. (B) An embryo injected with mRNA encoding the photoconvertible protein Kaede and globally irradiated from the animal pole for 10 s at 7 hpf. Fluorescence imaging revealed more Kaede photoconversion at the animal pole then at the margin. (C) The same embryo globally irradiated laterally, as mounted in (B), for an additional 10 s, resulting in even distribution of photoconverted Kaede protein. Scale bar: 200 μm. (See color plate.)

Irradiating mis-oriented embryos may result in the bottom-most cells (those furthest away from the light beam) receiving lower levels of UV light due to shielding by the overlying cells. This effect was observed in embryos microinjected with mRNA encoding the Kaede protein, which undergoes a green-to-red fluorescence photoconversion upon exposure to 360-nm light (Ando *et al*., 2002). When such embryos were irradiated from the anterior at 7 hpf, the cells below the "equator" plane of the embryo had less photo-converted Kaede protein than cells above the equator (Fig. 3B). However, when the same embryo was further irradiated laterally, red photoconverted Kaede protein was present evenly throughout the embryo (Fig. 3C). Uniform irradiation is therefore paramount to achieving consistent global photoactivation.

B. Optimization of UV Light Duration Using cFD

Since the CMNB caging group in cFD is chemically similar to the DMNB moiety in the cMO linker (Goeldner and Givens, 2005; Kobayashi *et al*., 2007), cFD can be used to estimate the irradiation duration that most efficiently photoconverts the cMO on a epifluorescence microscope system equipped with a specific set of illumination systems, optical filters, and objectives. In such experiments, cFD-microinjected embryos were globally irradiated at approximately 2 hpf as described above. UV light intensity was kept at 100% while irradiation duration was incrementally increased until maximum fluorescence intensity is achieved (all fluorescein was uncaged). With our 20×/0.5 NA water-immersion objective, fluorescein signal intensity reached saturation after a 10-s irradiation. Experimental parameters for other objectives can be estimated using the following equation describing the relationship between uncaging rate, NA, and magnification (M) (Eq. 2):

$$\text{Uncaging rate} \propto \text{NA}^2\text{M}^2 \tag{2}$$

Therefore a 20× lens with a lower NA (such as 0.40) produces UV light with 36% less intensity and disproportionally longer exposures may be required, possibly up to 20 s. Higher magnification objectives can enable shorter exposure times, but the irradiated regions will be smaller in area. The calibration of UV irradiation duration should therefore be performed with each epifluorescence microscope prior to its use for cMO uncaging. Typical irradiation durations are in the 10–30 s range when the microscope is capable of producing UV light with sufficient intensity to readily uncage cFD and cMO reagents.

C. Minimization of UV Phototoxicity

UV light can be damaging to zebrafish embryos, especially during blastula stages. Once the sufficient UV light exposure for cFD photoactivation was determined (10 s with our microscope system), embryos microinjected with cFD were globally irradiated as above at 2.5 hpf and raised to 1 day post fertilization (dpf). This dose of UV light was adequately tolerated as 67% of the embryos appeared to develop normally

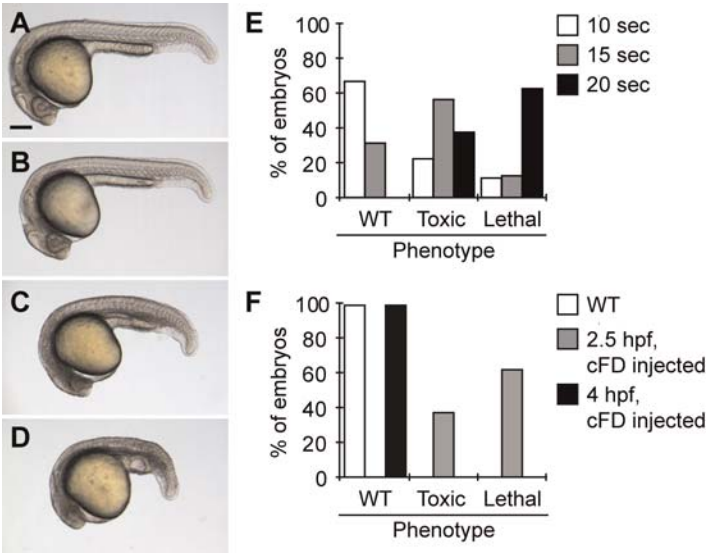


Fig. 4 UV light-induced toxicity in zebrafish embryos. Representative micrographs of embryos microinjected with 2 nL of a 0.05% cFD solution and irradiated for 0 (A), 10 (B), 15 (C), and 20 (D) s at 2.5 hpf. Embryos in (C) and (D) have morphological defects resulting from UV light damage. Scale bar: 200 μm. (E) Levels of embryonic toxicity and lethality by 1 dpf associated with different durations of UV irradiation at 2.5 hpf. (F) Levels of embryonic toxicity and lethality by 1 dpf associated with global, 20-s UV irradiation. Wild-type and gastrula-stage, cFD-injected embryos were less susceptible to UV light damage than blastula-stage, cFD-injected embryos. (See color plate.)

(Fig. 4A–B, E). However, embryos globally irradiated for 15 s or longer at 2.5 hpf had increasingly more gastrulation delays, necrotic tissues, and death by 1 dpf, indicating potential phototoxicity (Fig. 4C–E). In contrast, irradiation of wild-type embryos at 2.5 hpf or cFD-injected embryos at 4 hpf for 20 s did not produce any noticeable toxicity (Fig. 4F). It is therefore preferable to perform global irradiations at the gastrula stage or later as blastula-stage embryos are more susceptible to UV damage. Irradiation procedures that minimize phototoxicity should therefore be established empirically on each individual microscope system prior to using caged reagents for biological studies. These assessments should also be conducted with embryos microinjected with cFD and/or cMO, as the uncaging reaction itself can be associated with some cytotoxicity (Fig. 4F).

VI. Titration of cMO Dose for Optimum Dynamic Range

Since accurate reagent dosing is essential for cMO efficacy, it is important to perform a careful titration with the 25-base targeting MO to find the minimum effective dose. For example, in preparation for our studies with a *ntla* cMO, we systematically assess the embryonic phenotypes associated with different doses of

the conventional *ntla* MO. Embryos lacking the *ntla* T-box transcription factor fail to form the notochord, are posteriorly truncated, and exhibit U-shaped rather than V-shaped somites (Halpern *et al*., 1993; Schulte-Merker *et al*., 1994). Mutants and morphants lacking *ntla* function also exhibit ectopic medial floor plate, a ventral region of the developing spinal cord, and it is believed that *ntla* acts as a transcriptional switch between notochord and medial floor plate cell fates (Amacher *et al*., 2002; Halpern *et al*., 1997). Phenotypes resulting from loss of *ntla* function can be categorized into four classes according to their severity: class I = a fully penetrant *ntla* mutant phenotype characterized by no notochord, U-shaped somites, and a lack of posterior structures; class II = no notochord, U-shaped somites, and some posterior somites; class III = incompletely vacuolated notochord, V-shaped somites, and a shortened anterior–posterior axis; and class IV = wild-type phenotype (Ouyang *et al*., 2009). These phenotypes can be recapitulated by microinjecting varying doses of the conventional *ntla* MO (GACTTGAGGCAGACATATTTCCGAT), and class I phenotype was found to require a minimum dose of 115 fmol/embryo (Fig. 5).The minimum dose of the *ntla* MO that produced class I phenotype (115 fmol/embryo) served as a starting point for titration of the *ntla* cMO. Following microinjection of

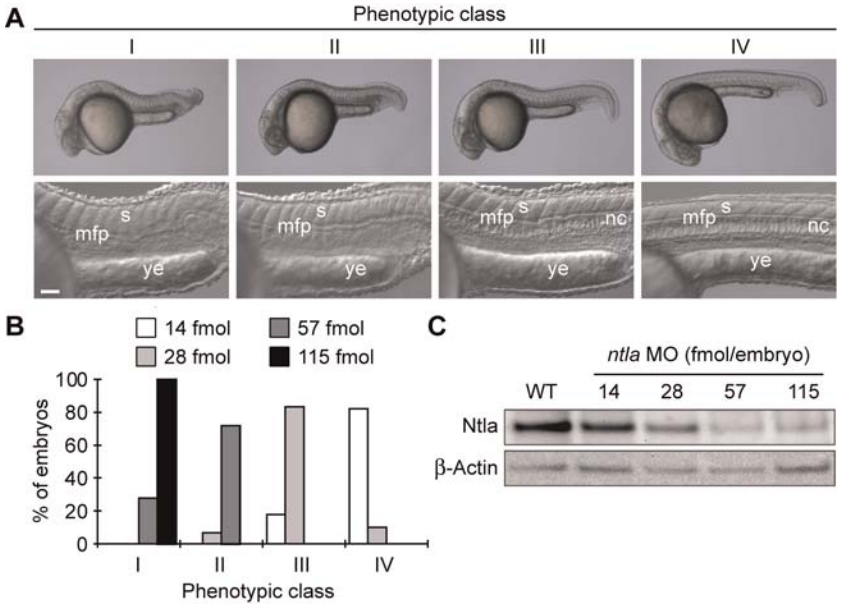


Fig. 5 Classification of*ntla* morphant phenotypes. (A) Morphology of phenotype classes I–IV at 1 dpf. Somitic (s), medial floor plate (mfp), notochord (nc), and yolk extension (ye) tissues are labeled. Scale bars: top panels, 200 μm; bottom panels, 50 μm. 115 fmol is equal to 1 ng of *ntla* MO. (B) Phenotypic distribution associated with different embryonic doses of the *ntla* MO. (C) Immunoblots showing Ntla protein levels in 10-hpf zebrafish embryos microinjected at the one-cell stage with various doses of the *ntla* MO. β-Actin levels are shown as a loading control. Adapted with permission (Ouyang *et al*., 2009; Copyright 2009, American Chemical Society). (See color plate.)

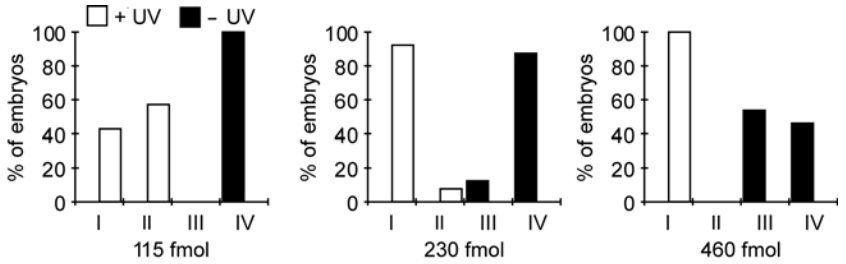


Fig. 6 Dose-dependent activity of the *ntla* cMO. Global irradiations were performed at 2.5 hpf and phenotypes were scored at 1 dpf according to the phenotype classes described in Fig. 5.

the *ntla* cMO, half of the embryos were globally irradiated for 10 s at 2 hpf as described in Section 5. At 1 dpf, the irradiated embryos had a mixture of class I and class II *ntla* phenotypes, whereas the unirradiated embryos appeared wildtype (class IV) (Fig. 6). The lack of fully restored MO activity upon uncaging was not surprising, as photochemical reactions do not typically go to completion due to competing, inactivating side reactions. When the *ntla* cMO dose was increased to 230 fmol/embryo, the desired activity profile was achieved, with global irradiation producing predominately class I embryos (Fig. 6). Higher doses of *ntla* cMO resulted in appearance of class III phenotypes in unirradiated embryos (Fig. 6). The 230 fmol/embryo dose was therefore used along with a 10-s irradiation protocol for all subsequent studies.

VII. Localized Photoactivation of Caged Reagents

A. Spatially Restricted UV Illumination Using a Photomask

Localized photoactivation can be performed on a compound epifluorescence microscope by reducing the size of the illumination field diaphragm. For example, our Leica DM4500B equipped with a 20× 0.5 NA water-immersion objective described in Section 5A can illuminate regions as small as a 100 μm-diameter circle or a 200 × 300 μm rectangle using its adjustable diaphragm. The spatial limits of this irradiation can also be visually confirmed using either cFD or the photoconvertible protein Kaede (Fig. 7A–B). Irradiation of smaller features or more complex patterns can be achieved using higher magnification objectives and/or micromirror array systems such as the Mosaic Digital Illumination System (Photonic Instruments).

Precise irradiation targeting is a prerequisite for reproducible photoactivation experiments, requiring methods for efficiently immobilizing of live embryos and targeting specific morphological features with UV light. To mount and orient gastrulation-stage zebrafish we used the same agarose microinjection template as that described in Section 5 for our global irradiation studies and kept the embryos in their chorions. Segmentation-stage embryos were dechorinated and placed in agarose templates cast with 0.018 inch-wide channels, whereas pharyngula-stage larva were

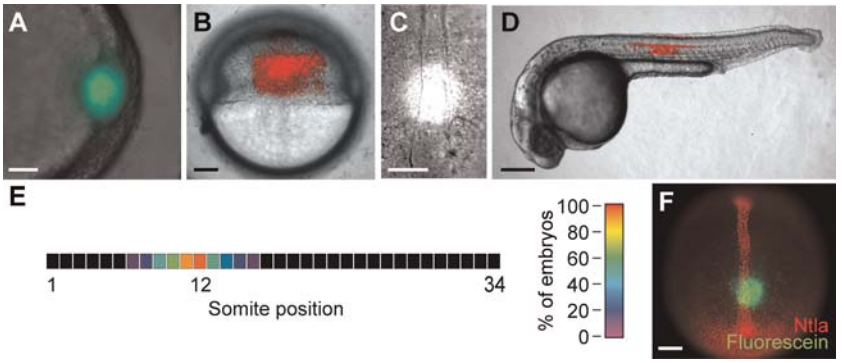


Fig. 7 Localized irradiation of zebrafish embryos. (A) Embryos injected with 2 nL of a 0.05% cFD solution were irradiated within the shield at 6 hpf for 10 s using a circular photomask (100-μm diameter). As expected, a circular region of green fluorescence was immediately apparent in the targeted region. (B) Embryos injected with 50 pg of *Kaede* mRNA and irradiated laterally at 6 hpf for 10 s using a rectangular photomask (200 × 300 μm). A rectangular region of red fluorescence was immediately observed in the targeted region. (C) Brightfield micrograph of a 10-hpf embryo undergoing UV irradiation through a circular, 100-μm-diameter photomask positioned 100 μm above the posterior end of the chordamesoderm. Grid overlays using Metamorph® software are not shown. (D) Embryos injected with 50 pg of *Kaede* mRNA and locally irradiated as in (C). A red fluorescent region of notochord and floor plate cells centered around the 12th somite was visible at 1 dpf. (E) Heat map demonstrating the precision with which zebrafish embryos can be locally irradiated as described in (C). The average location of red fluorescent notochord cells along the anterior–posterior axis resulting from the targeted irradiation of 10-hpf embryos is shown (n = 18 embryos). (F) Fluorescence micrograph of a 10-hpf embryo injected with cFD irradiate as described in (C), immediately fixed with paraformaldehyde, and immunostained with anti-Ntla and anti-fluorescein antibodies. A circular region of uncaged fluorescein was detected within the Ntla-expressing chordamesoderm, 100 μm anterior to the tailbud. Scale bars: A and C, 50 μm; B, 100 μm; D, 200 μm; F, 100 μm. (See color plate.)

immobilized in MegaMounts (https://wiki.med.harvard.edu/SysBio/Megason/MegaMounts) and treated with tricaine. Irradiating specific tissues in a consistent manner was then achieved by using the combined brightfield and fluorescence illumination mode on the Leica DM4500B microscope. For example, to irradiate a 100 μm-diameter circular region within the chordamesoderm in 10 hpf embryos, a 50 × 50 μm grid was digitally overlayed onto the live preview window using MetaMorph® software, grid lines were aligned to the center of the fluorescent mask, and immobilized embryos oriented using the grid lines to enable irradiation of the chordamesoderm lying 100 μm from the posterior boundary (Fig. 7C). When 18 Kaede-expressing embryos were irradiated in this manner, by 1 dpf they displayed a region of red-fluorescent notochord cells tightly clustered within the same position along the anterior–posterior axis (Fig. 7D–E).

B. Determination of Protein Levels in Targeted Tissues

Since cMOs inhibit the splicing or translation of their targeted RNAs, the rate by which the encoded protein is degraded will determine the time point when gene

function is actually lost after cMO photoactivation. For this reason, accurately interpreting cMO-induced phenotypes requires an assessment of protein levels for the targeted gene in the irradiated cells. This can be accomplished by immunodetection technologies, using antibodies that recognize the protein of interest and an anti-fluorescein antibody (Roche Applied Science, Indianapolis, IN 1426320) that specifically binds to uncaged cFD. For example, when embryos microinjected with cFD and irradiated as described in Section 7A were immediately fixed and immunostained, a 100 μm-diameter circular staining pattern was apparent, marking an irradiated region within the Ntla-expressing chordamesoderm (Fig. 7F). The immunostaining procedure used an anti-Ntla polyclonal antibody (1:1000 dilution) (Schulte-Merker *et al*., 1994) and a monoclonal anti-fluorescein antibody (1:200 dilution), as well as their corresponding secondary antibodies (anti-rabbit Alexa-Fluor 594 and anti-mouse Alexa-Fluor 488, each at a 1:200 dilution) according to standard whole-mount immunostaining protocols (Nüsslein-Volhard and Dahm, 2002). Identical irradiations were performed on embryos co-injected with the *ntla* cMO and cFD, followed by fixation and immunostaining at various time points. These collective experiments revealed the time frame in which Ntla levels are significantly diminished in the irradiated cells, allowing phenotypes associated with *ntla* cMO activation to be linked to loss of Ntla function.

VIII. Conclusion

The protocols for cMO design, synthesis, and application described in this chapter are intended to facilitate the use of these reverse-genetic tools in zebrafish and other organisms. The cMO technology is broadly applicable to the embryonic transcriptome, as illustrated by our development of cMOs targeting several patterning genes. In addition, the methods employed for global or spatially restricted cMO photoactivation could be generally applied to emerging optogenetic systems.

While cMOs exemplify the potential of synthetic reagents for *in vivo* studies, it is worth discussing some of the challenges that remain in the implementation of this technology. These include both practical and experimental limitations. For example, although the synthetic and purification procedures for cMO preparation are relatively straightforward and robust, they involve techniques and equipment that are not common in most biological laboratories. Further simplification of cMO design, synthesis, and purification procedures—or perhaps more ideally, commercialization of the cMO technology—would help promote the use of these reagents by the developmental biology community.

We have also experienced difficulty in caging certain MOs. In some cases this appears to be due to cytotoxicity associated with the cMO oligonucleotide, which is typically 10–13 bases longer than conventional MOs. For instance, we previously attempted to generate a *spt* cMO, using a 25-base targeting MO that exhibited a narrow range of effective doses; partial *spt* mutant phenotypes were obtained at a dose of 3 ng/embryo and embryonic toxicity was observed at higher MO

concentrations (Ouyang *et al*., 2009). Tethering the inhibitory oligomer to the *spt* MO increased rather than mitigated reagent toxicity, and photoactivation of the *spt* cMO did not fully recapitulate the developmental defects induced by the conventional MO. Consistent with these results, quantitative models of cMO activity suggest that as the MO dose required for gene silencing increases, the functionally equivalent dose of photoactivated cMO rises disproportionately, as the inhibitory MO released by cMO photolysis can begin to interfere with RNA hybridization by the targeting MO (Ouyang *et al*., 2009). The greater potential for off-target effects associated with cMO hairpins may also limit their utility in studying larval gene function, since relatively large amounts of MO are frequently microinjected in these experiments to compensate for reagent dilution during later developmental stages.

MOs that potentiate UV light-induced toxicity are also not amenable to our caging strategy, as demonstrated by our attempts to generate a cMO against *notch1a* (I. A. Shestopalov and J. K. Chen, unpublished observations). When we injected the *notch1a* cMO and globally irradiated the embryos at 2.5 hpf, we observed severe UV-dependent defects during blastula formation and gastrulation that were not found in conventional *notch1a* morphants. Interestingly, morphologically identical defects occurred in embryos injected with the conventional *notch1a* MO and globally irradiated at 2.5 hpf, indicating that the targeting MO itself can increase embryonic sensitivity to UV light damage.

New strategies for cMO design and synthesis could help overcome these challenges. Approaches that minimize the cMOs cytotoxicity and the potential for UV light damage would be important advances. In principle, this could be achieved by minimizing or eliminating the inhibitory MO portion of these reagents, replacing the nitrobenzyl-based photocleavable group with chromophores that undergo photochemical reactions with lower doses of UV light or at less damaging wavelength, devising methods that obviate the need to utilize cMOs at doses two-fold greater than that of targeting MO, and/or utilizing new oligonucleotide scaffolds as a less toxic substitute for MOs (Shestopalov and Chen, 2010). The utility of cMOs as functional genomic probes would also benefit from the development of new uncaging technologies. For example, enzymatically triggered cMOs could be used in combination with transgenic organisms to allow MO activation with a spatial precision and three-dimensional complexity that would be difficult to achieve with photomasks or even micromirror array-based illumination. Reversible control of MO function would similarly enable gene silencing with greater temporal dexterity, facilitating studies of how genes dynamically regulate embryonic patterning.

Achieving these advances will require a collaborative effort by chemists and developmental biologists, as well as scientists trained in both disciplines. Given the amenability of zebrafish to optical technologies, transgenesis, and chemical perturbations and the burgeoning use of molecular probes in zebrafish studies, we anticipate that the zebrafish community will play a leading role in these scientific explorations. Just as synthetic reagents such as MOs have transformed how we study zebrafish biology, zebrafish biology can inspire new ways in which chemistry can provide insights into biological processes.

Acknowledgments

We gratefully acknowledge financial support from the NIH Director's Pioneer Award (DP1 OD003792), the NIH/NIGMS (R01 GM072600), the March of Dimes Foundation (1-FY-08-433), and the California Institute for Regenerative Medicine (T1-0001). Some figures and schemes in this chapter have been adapted from the authors' previous publication (Ouyang *et al*., 2009) with permission from the American Chemical Society.

References

Alexander, J., Rothenberg, M., Henry, G. L., and Stainier, D. Y. (1999). Casanova plays an early and essential role in endoderm formation in zebrafish. *Dev. Biol.* **215**, 343–357.

Amacher, S. L., Draper, B. W., Summers, B. R., and Kimmel, C. B. (2002). The zebrafish T-box genes no tail and spadetail are required for development of trunk and tail mesoderm and medial floor plate. *Development* **129**, 3311–3323.

Ando, H., Furuta, T., Tsien, R. Y., and Okamoto, H. (2001). Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat. Genet.* **28**, 317–325.

Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002). An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12651–12656.

Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. A., and Ekker, S. C. (2009). A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69–77.

Blidner, R. A., Svoboda, K. R., Hammer, R. P., and Monroe, W. T. (2008). Photoinduced RNA interference using DMNPE-caged 2'-deoxy-2'-fluoro substituted nucleic acids in vitro and in vivo. *Mol. Biosyst.* **4**, 431–440.

Cerda, J., Reidenbach, S., Pratzel, S., and Franke, W. W. (1999). Cadherin-catenin complexes during zebrafish oogenesis: heterotypic junctions between oocytes and follicle cells. *Biol. Reprod.* **61**, 692–704.

Ekker, S. C. (2004). Nonconventional antisense in zebrafish for functional genomics applications. *Methods Cell Biol.* **77**, 121–136.

Fedoryak, O. D., and Dore, T. M. (2002). Brominated hydroxyquinoline as a photolabile protecting group with sensitivity to multiphoton excitation. *Org. Lett.* **4**, 3419–3422.

Gee, K. R., Weinberg, E. S., and Kozlowski, D. J. (2001). Caged Q-rhodamine dextran: a new photo-activated fluorescent tracer. *Bioorg. Med. Chem. Lett.* **11**, 2181–2183.

Goeldner, M., and Givens, R. (2005). *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules.* Wiley-VCH, Weinheim.

Gradinaru, V., Zhang, F., Ramakrishnan, C., Mattis, J., Prakash, R., Diester, I., Goshen, I., Thompson, K. R., and Deisseroth, K. (2010). Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165.

Halpern, M. E., Hatta, K., Amacher, S. L., Talbot, W. S., Yan, Y. L., Thisse, B., Thisse, C., Postlethwait, J. H., and Kimmel, C. B. (1997). Genetic interactions in zebrafish midline development. *Dev. Biol.* **187**, 154–170.

Halpern, M. E., Ho, R. K., Walker, C., and Kimmel, C. B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* **75**, 99–111.

Kobayashi, T., Urano, Y., Kamiya, M., Ueno, T., Kojima, H., and Nagano, T. (2007). Highly activatable and rapidly releasable caged fluorescein derivatives. *J. Am. Chem. Soc.* **129**, 6696–6697.

Kozlowski, D. J., and Weinberg, E. S. (2000). Photoactivatable (caged) fluorescein as a cell tracer for fate mapping in the zebrafish embryo. *Methods Mol. Biol.* **135**, 349–355.

Lippincott-Schwartz, J., and Patterson, G. H. (2008). Fluorescent proteins for photoactivation experiments. *Methods Cell Biol.* **85**, 45–61.

Mably, J. D., Mohideen, M. A., Burns, C. G., Chen, J. N., and Fishman, M. C. (2003). heart of glass regulates the concentric growth of the heart in zebrafish. *Curr. Biol.* **13**, 2138–2147.

Nasevicius, A., and Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**, 216–220.

Nguyen, Q. N., Chavli, R. V., Marques, J. T., Conrad Jr., P. G., Wang, D., He, W., Belisle, B. E., Zhang, A., Pastor, L. M., Witney, F. R., Morris, M., Heitz, F., Divita, G., Williams, B. R., and McMaster, G. K. (2006). Light controllable siRNAs regulate gene suppression and phenotypes in cells. *Biochim. Biophys. Acta.* **1758**, 394–403.

Nüsslein-Volhard, C., and Dahm, R. (2002). *Zebrafish, A Practical Approach.* Oxford University Press, Oxford.

Ouyang, X., and Chen, J. K. (2010). Synthetic strategies for studying embryonic development. *Chem. Biol.* **17**, 590–606.

Ouyang, X., Shestopalov, I. A., Sinha, S., Zheng, G., Pitt, C. L., Li, W. H., Olson, A. J., and Chen, J. K. (2009). Versatile synthesis and rational design of caged morpholinos. *J. Am. Chem. Soc.* **131**, 13255–13269.

Remy, S., Tesson, L., Menoret, S., Usal, C., Scharenberg, A. M., and Anegon, I. (2010). Zinc-finger nucleases: a powerful tool for genetic engineering of animals. *Transgenic Res.* **19**, 363–371.

Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002). A stepwise huisgen cycloaddition process: copper (I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem. Int. Ed. Engl.* **41**, 2596–2599.

Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B., and Nusslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009–1015.

Shestopalov, I. A., and Chen, J. K. (2008). Chemical technologies for probing embryonic development. *Chem. Soc. Rev.* **37**, 1294–1307.

Shestopalov, I. A., and Chen, J. K. (2010). Oligonucleotide-based tools for studying zebrafish development. *Zebrafish* **7**, 31–40.

Shestopalov, I. A., Sinha, S., and Chen, J. K. (2007). Light-controlled gene silencing in zebrafish embryos. *Nat. Chem. Biol.* **3**, 650–651.

Sumanas, S., and Lin, S. (2006). Ets1-related protein is a key regulator of vasculogenesis in zebrafish. *PLoS Biol.* **4**, e10.

Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim. Biophys. Acta.* **1489**, 141–158.

Talbot, W. S., Trevarrow, B., Halpern, M. E., Melby, A. E., Farr, G., Postlethwait, J. H., Jowett, T., Kimmel, C. B., and Kimelman, D. (1995). A homeobox gene essential for zebrafish notochord development. *Nature* **378**, 150–157.

Tang, X., Maegawa, S., Weinberg, E. S., and Dmochowski, I. J. (2007). Regulating gene expression in zebrafish embryos using light-activated, negatively charged peptide nucleic acids. *J. Am. Chem. Soc.* **129**, 11000–11001.

Tang, X., Swaminathan, J., Gewirtz, A. M., and Dmochowski, I. J. (2008). Regulating gene expression in human leukemia cells using light-activated oligodeoxynucleotides. *Nucleic Acids Res.* **36**, 559–569.

Tomasini, A. J., Schuler, A. D., Zebala, J. A., and Mayer, A. N. (2009). PhotoMorphs: a novel light-activated reagent for controlling gene expression in zebrafish. *Genesis* **47**, 736–843.

Young, D. D., Lusic, H., Lively, M. O., Yoder, J. A., and Deiters, A. (2008). Gene silencing in mammalian cells with light-activated antisense agents. *Chembiochem.* **9**, 2937–2940.

Zhao, X. F., Fjose, A., Larsen, N., Helvik, J. V., and Drivenes, O. (2008). Treatment with small interfering RNA affects the microRNA pathway and causes unspecific defects in zebrafish embryos. *FEBS J.* **275**, 2177–2184.

PART 2

Transgenesis

CHAPTER 10

Advanced Zebrafish Transgenesis with Tol2 and Application for Cre/*lox* Recombination Experiments

Christian Mosimann[*,†,‡] and Leonard I. Zon[*,†,‡]

[*]Howard Hughes Medical Institute, Boston, Massachusetts, USA

[†]Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, USA

[‡]Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00010-0

Abstract

Spatio-temporal transgene regulation by transgenic DNA recombinases is a central tool for genetic research in multicellular organisms, with excellent applications for misexpression and lineage tracing experiments. Cre recombinase-controlled *lox* site recombination is a cornerstone of contemporary mouse genetics, and Cre/*lox* techniques therefore attract increasing interest in the zebrafish field. Tol2-mediated zebrafish transgenesis now provides a stable platform for *lox* cassette transgenes, while the ease of drug treatments in zebrafish makes the model an ideal candidate for Tamoxifen/4-hydroxytamoxifen-inducible CreER^{T2} experiments. In this chapter, we will first introduce the basics of Cre/*lox* methodology, CreER^{T2} regulation by Tamoxifen/4-hydroxytamoxifen, as well as the benefits of Tol2 transgenesis for Cre/*lox* experiments. We will then in detail outline practical experimental steps for Tol2 transgenesis toward the creation of single-insertion transgenes. Lastly, we will introduce protocols for 4-hydroxytamoxifen-mediated CreER^{T2} induction to perform spatio-temporal *lox* transgene regulation experiments in zebrafish embryos.

I. Introduction

A. Transgenesis

A transgene is an experimentally constructed piece of DNA that has integrated into the genome of a recipient organism. Once integrated into germ cells, subsequent generations inherit the transgene, referred to as stable transgene transmission. Engineered transgenes allow testing of molecular mechanisms, such as the function of a specific RNA or protein, by expressing a desired gene at different developmental times and in different cell types. Transgenes can further express an experimentally detectable marker in particular cell types or cellular structures for detailed, high-resolution observations.

Experimentally controlling spatio-temporal transgene expression by regulated DNA recombination has become a central method in model organism research. In the simplest case, a transgenic DNA recombinase enzyme 1) recognizes two genomic DNA target sequences in *cis* or in *trans*, and 2) recombines the target DNA sequences, resulting in a recombined DNA product based on the initial relative orientation of the target sequences. When acting in *cis*, the recombinase removes or inverts a piece of transgenic DNA flanked by target sequences, consequently eliminating or functionally reconstituting (‘‘switching’’) an engineered transgene. In mice and *Drosophila*, cassette excision is a classic application wherein a

recombinase excises a target sequence-flanked transgene cassette. Recently, also the zebrafish field has ventured into this territory.

B. The Cre/*lox* System, Inducible CreER^{T2}, and Their Utility in Zebrafish

In vertebrate molecular genetics, the current workhorse is the bacteriophage P1-derived cyclic recombinase Cre, which catalyzes the site-specific recombination between 34 base pair (bp) inverted repeat *lox* sites (from *locus of X-ing over*) (Sauer, 1987; Sauer and Henderson, 1988) without additional co-factors. Cre-mediated recombination of two *lox* sites with opposite orientation inverts the intervening DNA cassette, while two *lox* sites in direct orientation cause circularization, excision, and subsequent loss of the cassette, leaving one functional *lox* site at the transgenic locus. Several variations of *lox* sites have been created by mutagenesis (such as *loxP*, *lox2272*, *and loxN*) (Livet *et al*., 2007) that can be used together in the same transgene to create mutually exclusive, predictable cassette recombination patterns.

Cre/*lox* recombination finds widespread application in genetic two-component transgene systems (Fig. 1): in the default state, a transgene containing a promoter drives expression of a gene cassette flanked by *lox* sites. Cre expressed from an independent transgene excises the *lox* cassette and terminates its expression.

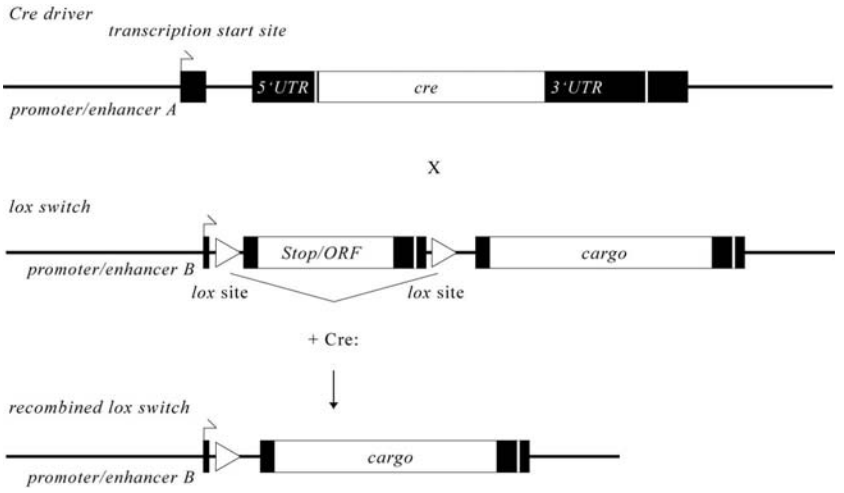


Fig. 1 Schematics of Cre/*lox* experiment using transgenes (also compare to Fig. 2). The Cre driver is a transgene expressing the Cre-encoding ORF from a specific genetic control element (promoter/enhancer *A*). The *lox* switch expresses an ORF framed by *lox* recombination sites (34 bp), or a transcriptional stop cassette, from promoter/enhancer *B* in the default state. Cre recombinase can recombine the *lox* sites in cells where the promoter/enhancer *A* is active and results in the excision of the *lox* cassette, leaving a single *lox* site in place. Promoter/enhancer *B* now expresses the cargo ORF once active.

A *lox*-flanked Stop cassette can also be followed directly by a second gene of interest, referred to as cargo. Cre-mediated *lox* site recombination excises the cassette and brings the cargo gene under direct control of the promoter, thus switching expression from the cassette template or a stopped condition to cargo transcription (Fig. 1). Such modular transgenes, which we refer to as *lox* switch constructs, are today's gold standard for lineage tracing as well as conditional loss- and gain-of-function experiments in mice (Bailey *et al*., 2009; Sauer, 1998).

In contrast to classic transgenic reporters, a *lox* switch controlled by an enhancer/promoter distinct from that driving Cre expression uncouples reporter labeling from the cell type-specific promoter activity of the Cre driver. The most universally applicable *lox* switches are controlled by ubiquitous promoters and can be flexibly crossed to tissue-specific Cre drivers, while the promoter driving the Cre transgene dictates spatio-temporal recombinase activity and thus onset of *lox* switch cargo expression. The first attempts at controllable Cre transgene activity in zebrafish used the heat-shock inducible promoter of the *hsp70* gene (*hsp70:cre*) (Le *et al*., 2007; Thummel *et al*., 2005) Other pioneering efforts employed *rag2* or *β-actin* promoter-controlled *loxP* cassettes followed by *c-MYC* or $kRAS^{V12}$ oncogene cargo; switching to cargo expression by injection Cre-encoding mRNA into transgenic embryos or by heat-induced expression of an *hsp70:cre* driver transgene gave rise to inducible models of tumorigenesis in zebrafish (Feng *et al*., 2007; Langenau *et al*., 2005; Le *et al*., 2007).

However, *hsp70* promoter-driven transgenes are problematic because they are ''leaky'' and may constitutively express at low levels (presumably due to *hsp70* sensitivity to various stresses and transgene locus-dependent effects). Furthermore, *hsp70* transgenes express during oogenesis, and *hsp70:cre* features maternally contributed constitutive Cre activity in the embryo (Hans *et al*., 2009, 2010). Tightly regulated temporal post-translational control of Cre activity has been elegantly achieved by fusing Cre with the modular ligand-sensing domain of the Estrogen Receptor (ER) (Feil *et al*., 1996). Heat-shock protein 90 (Hsp90) binds incompletely folded ER and traps the receptor in the cytoplasm. Binding of estrogen or structurally related steroids induces a conformational change in the ligand-binding domain of ER, which subsequently sheds Hsp90 and translocates to the nucleus. By this mechanism, a CreER fusion protein is constitutively sequestered in the cytoplasm and cannot reach nuclear *lox* sites until it binds estrogen. High expression of the CreER protein can, however, overwhelm endogenous Hsp90 levels and cause ''leaky'' *lox* cassette recombination by nuclear translocation of escaped CreER (see also below).

To circumvent potential CreER activation by endogenous estrogen, mutant versions of the ER domain have been developed that have high affinity for estrogen derivatives not found in metazoans. The most commonly used version for spatio-temporal lineage tracing experiments in transgenic mice is $CreER^{T2}$, which is insensitive to natural estrogen but strongly binds 4-hydroxytamoxifen (4-OHT) (Feil *et al*., 1996, 1997). 4-OHT is a labile metabolite of the prodrug Tamoxifen (TAM) with a 30–100-fold higher affinity for the ligand-binding ER domain than unmodified TAM (Katzenellenbogen *et al*., 1984). $CreER^{T2}$ transgenes have been successfully established in zebrafish (Hans *et al*., 2009; Mosimann *et al*., 2011), and

recently, *cmlc2:creER^{T2}* and *gata4:creER^{T2}* transgenic lines have been used for lineage tracing during heart regeneration (Jopling *et al*., 2010; Kikuchi *et al*., 2010).

In mice, body size poses an obstacle to TAM prodrug administration; once metabolized into 4-OHT after injection or feeding, the unstable compound enters CreER^{T2}-expressing tissues with metabolism- and tissue-dependent variation in concentration and distribution, which causes variation in *lox* recombination efficiency and lag time after induction. Furthermore, 4-OHT in the induced mouse cannot be removed post-induction. In contrast, zebrafish embryos are small, allow direct 4-OHT addition to the embryo medium, and rapidly take up the drug, which translates into fast CreER^{T2} responses (Hans *et al*., 2009; Mosimann *et al*., 2011). We have further observed that conformational induction of active CreER^{T2} by 4-OHT for 15 min, followed by wash out of free drug, triggers high *lox* recombination of a given transgene when performed in the first 24 hpf (Mosimann *et al*., 2011) (see also below). This suggests that CreER^{T2} induction in zebrafish can be controlled by discrete 4-OHT pulses, thus confining cell lineage labeling or gene misexpression to restricted developmental time windows.

C. Cre/*lox* and *Tol2* Zebrafish Transgenesis: Current Status

The growing collection of characterized enhancer/promoter elements in zebrafish provides a promising basis for Cre/*lox* transgenes. The zebrafish appears to be well-suited for CreER^{T2} applications, yet limitations in transgenesis technology restricted the efficient creation of the necessary transgenic lines until recently. The first efforts to establish Cre/*lox* transgenes relied on rare integration of concatemers of linearized plasmid DNA (Iyengar *et al*., 1996; Stuart *et al*., 1988). Concatemeric insertions result in tandem repeats of *lox* cassette transgenes. As Cre can act on any two *lox* sites in the multi-copy concatemer, individual embryonic cells may retain different copy numbers of the recombined transgene or experience unproductive recombination. Thus, it is not surprising that reported recombination efficiencies were low (Feng *et al*., 2007; Hans *et al*., 2009; Le *et al*., 2007). Furthermore, high-copy number concatemeric transgenes are prone to silencing and to DNA repair-mediated recombination, leading to variable transgene expression or loss over successive generations.

Single-insertion transgenesis via Tol2 transposons resolves these issues (Hans *et al*., 2009; Kawakami, 2005, 2007), and has fueled a notable increase in Cre/*lox* applications in zebrafish. The major remaining obstacles to widespread application in zebrafish are the small numbers of established tissue-specific Cre or CreER^{T2} lines and reliable *lox* reporter transgenes. Both components are required for synergistic development of the technology. Screening for effective *lox* switch insertions requires a collection of characterized Cre transgenes, while a strong, ubiquitously expressed and universally applicable *lox* switch facilitates screening for novel functional Cre/CreER^{T2} transgenics.

Several reports have described *lox* switch lines with *Xef1α* or *β-actin* promoter-controlled *lox* cassettes containing fluorescent protein cargo genes (Bertrand and Traver, 2009; Boniface *et al*., 2009; Hans *et al*., 2009, 2010; Langenau *et al*., 2005;

Yoshikawa *et al*., 2008). These promoters, however, do not drive ubiquitous transgene expression in embryonic or adult tissues, in contrast to *Rosa26* transgenes in mice (Soriano, 1999; Zambrowicz *et al*., 1997). We recently isolated the *ubiquitin* (*ubi*) promoter, which is highly active at all stages of zebrafish development and in adulthood, and used it to create the Cre-sensitive *ubi:loxP-EGFP-loxP_mCherry* (*ubi:Switch*) transgene (Mosimann *et al*., 2011). Besides serving as a sensitive lineage tracing tool, *ubi:Switch* facilitates screening of new tissue-specific CreER^{T2} drivers and assessment of their temporal 4-OHT responsiveness. This transgene also reveals leaky recombination activity due to excessively high CreER^{T2} expression. We further engineered *ubi:creER^{T2}* as a zebrafish line for ubiquitous expression of 4-OHT-inducible CreER^{T2} to simplify screening for functional *lox* switch transgenes (Mosimann *et al*., 2011).

The demand for Cre/*lox* transgenic technology applicable to the zebrafish is high, yet efficient protocols and flexible transgenic lines remain to be established. Below, we will discuss key points concerning transgenesis, such as basic transgene design, Tol2 transposon handling, and post-injection genetics for creation of high-quality single-insertion transgenic lines. Subsequently, we will consider principles and protocols for Cre/*lox* experimentation in zebrafish including CreER^{T2} induction by 4-OHT and the application of transgenic lines for lineage tracing and general transgene manipulation.

II. Transgene Design, Cloning, and *Tol2* Transgenesis

A. Basic Principles of Transgene Design

Key to successful Cre/*lox* experimentation is efficient gene expression from the involved transgenes. Transcription of mRNA from a transgene is a multi-step process including many co- and post-transcriptional processes such as 5′-capping, intron excision, and 3′-polyadenylation. Because gene knock-in technology by homologous recombination is currently lacking in zebrafish, transgene expression requires incorporation of an enhancer/promoter element. Here, we consider the basic principles of eukaryotic RNA pol-II-based gene expression and their systematic implementation to create effective transgenes.

The minimal features of a protein-coding, mRNA-expressing transgene are (Fig. 2):

a) an enhancer or promoter element that contains a minimal promoter with transcription start site.
b) a 5′-untranslated region (5′-UTR) with a variant of the optimal Kozak consensus sequence *5′-cgccRccATGG-3′*, in which the key residue is a purine (abbreviated *R*) three bases upstream (position *-3*) of the protein's start codon (underlined) (Kozak, 1987, 1991). The Kozak sequence is recognized by the ribosome to initiate efficient translation at the start codon of the encoded protein. In practical terms, the *5′* UTR sequence is often defined by the used promoter, which dictates the transcription start site and 5′ sequence of the resulting mRNA. *5′*-UTR*s* have

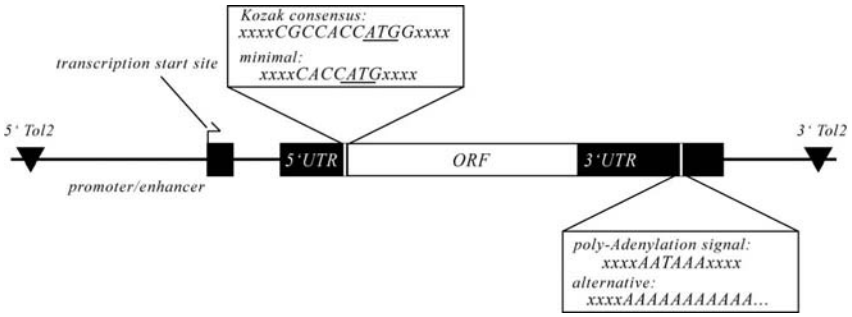


Fig. 2 Overview of individual components required for efficient expression of a transgene ORF from a given promoter/enhancer. See text for details, and compare to Fig. 1.

a profound effect on mRNA stability and translation efficiency. Unfortunately, we know little regarding zebrafish-specific requirements (if any) for effective 5′-UTR*s*.

Kozak sequences show many gene-specific variations, which reflects both the recognition tolerance of the ribosome and gene-specific regulatory features that modulate translation efficiency. Experience has shown that experimental transgenes using the *5′-cgccaccATG-3′* consensus or even *5′-caccATG-3′* (when using *pENTR/D* middle entry clones for the Gateway System, see below) are efficiently translated in zebrafish. 5′-UTR*s* with in-frame and out-of-frame *ATG*s upstream of the main *ATG* may severely impair efficient translation of the desired transgene cargo protein (Kozak, 1996) and should be avoided when designing transgene control elements.

c) a protein-coding open reading frame (ORF), starting with an *ATG* in a Kozak consensus and ending in a stop codon (*TAG*, *TAA*, *TGA*).
d) a *3′*-UTR with a polyadenylation signal sequence (*5′-AAUAAA-3′*). As for *5′*-UTR*s*, little information is available on transgene-tested *3′*-UTR*s* and their impact on RNA stability or protein expression. The *SV40* and *Bovine Growth Hormone* (*BGH*) gene derived minimal 3′-UTR*s* are used in the vast majority of transgenes reported to date. Both are fairly short (∼200 bp) and can be included as a module after an ORF to ensure efficient polyadenylation. A less elegant way to ‘‘ensure’’ polyadenylation is to include a stretch of *A* residues after the ORF, but this approach sacrifices the benefits of natively processed 3′-UTR*s*.

A transgene encoding a *5′*-UTR-*Kozak-[ORF]-3′*-UTR-*polyA* cargo plus a suitable promoter therefore contains the minimal modules to drive protein production. Transgenes intended to express microRNA loci or other non-coding RNAs require different structural features, which are generally poorly defined. *lox* sites are small (34 bp) and can be flexibly placed at any position of a transgene, most often flanking the cargo ORF (Fig. 1).

B. Multisite Gateway® System for Transgene Assembly

Assembly of the modules required in a transgene is easily accomplished via contemporary recombinant DNA methods, in particular by using the Multisite Gateway®*in vitro* recombination cloning system (Invitrogen Corp., Carlsbad, CA). In this modular system, the proprietary LR clonase II Plus enzyme recombines 1) a *5′* entry vector, 2) a middle entry vector, and 3) a *3′* entry vector into 4) a backbone plasmid using matching recombination sites engineered into the individual vectors. Typically, the *5′* entry vector contains a promoter/enhancer as transgene driver, the middle entry vector provides a desired ORF, and the *3′* entry vector contains a minimal *3′*-UTR with polyadenylation sequence. Any promoter can be linked with any ORF once they have been integrated into Multisite Gateway®-compatible backbones.

The recent creation of the Tol2 kit and related zebrafish transgenesis vector collections to create Tol2 transposon transgenes with Multisite Gateway® (Kwan *et al*., 2007; Villefranc *et al*., 2007) provides a stepping stone to contemporary zebrafish transgene design. To create new Multisite Gateway®-compatible *5′* and middle entry vectors from gel-purified PCR products, we prefer to use the *pENTR5′* and *pENTR/D* TOPO cloning vectors (Invitrogen Corp., Carlsbad, CA) (Mosimann *et al*., 2011). *pENTR5′* is a TOPO TA cloning vector and requires the adenine (A)-overhangs on the 3′ ends of PCR products as generated by *Taq* polymerase. To minimize PCR amplification errors, we routinely use the *Taq*-containing Roche Expand Hi-Fidelity polymerase kit (Roche Applied Science, Indianapolis, IN). *pENTR/D* is a directional TOPO vector for cloning blunt-ended PCR products produced by hi-fidelity polymerases (PCR amplicons generated by *Taq*-supplemented polymerase can be cloned into the *pENTR/D* TOPO vector since some product remains blunt-ended.) The *pENTR/D* TOPO reaction preferentially yields directionally cloned inserts if the *5′* PCR primer contains the *5′* overhang sequence 5′-*CACC*-3′. As the majority of middle entry vectors are intended to deliver ORFs, the 5′ *CACC* overhang provides a minimal Kozak consensus when followed by the ORF's *ATG* start codon (as *5′-CACCATGxxx-3′*). Prior to use, all recombinant vectors should be validated by restriction digestion and sequencing of the entire insert.

The Tol2 kit and related plasmid collections provide backbones containing a Tol2 repeat-flanked Multisite Gateway® cassette into which the entry vectors recombine (Kwan *et al*., 2007; Villefranc *et al*., 2007). We keep our entry vectors and backbone working stocks at 40 ng/μL and perform recombination reactions in 5 μL volumes (50% of the volume recommended by the Gateway protocol) for 16 h in a thermocycler set at 25 °C. Although Multisite Gateway® is more expensive than restriction enzyme cloning (clonase costs around US-$ 11 for half-reactions), the system provides unparalleled speed and flexibility, including easy vector sharing.

C. Tol2 Transposon-mediated Transgenesis for Single Insertions

True single-insertion transgenesis in zebrafish, comparable to P-Element transposition in *Drosophila*, became possible when the Medaka-derived *Tol2* transposon

was engineered to carry transgene cargo (Kawakami, 2007; Kawakami *et al*., 2004). The original autonomous *Tol2* transposon contains a transposase gene and requires flanking sequences (200 bp at the *5′* end and 150 bp at the *3′* end) for transposase catalyzed genomic integration. The flanking sequences (called 5′ and 3′ *Tol2* arms) include 12 bp terminal inverted repeats and sub-terminal regions that are recognized by the Tol2 transposase. In non-autonomous *Tol2* transposon transgenes, the 5′ and 3′ *Tol2* arms are joined to the transgene cargo in a plasmid vector backbone. For transgenesis, one-cell stage zebrafish embryos are injected with a mix of Tol2 transposase-encoding mRNA and *Tol2* transposon transgene vector. *Tol2* transgene cargo does not seem to be size-restricted; indeed, *Tol2* transposons have successfully integrated Bacterial Artificial Chromosome (BAC) transgenes in the 70 kb range into the zebrafish genome (Suster *et al*., 2009). Data regarding genomic integration locus preferences are however limited (Kondrychyn *et al*., 2009).

Tol2 transposons are further simple to handle and efficiently transmit through the germline. We routinely observe *Tol2* transposon germline transmission frequencies of 20–100% of injected F0 animals, with 3–100% of F1 embryos per clutch positive for the transgene, similar to the published range (Kawakami, 2007). Together, *Tol2* transgenesis now provides unprecedented flexibility in zebrafish experimental design. *Tol2* transposons are therefore ideally suited for complex cargo transgenes, such as intricate *lox* switch vectors.

D. Preparation of *Tol2* mRNA and Transgene Plasmid DNA for Embryo Microinjection

Tol2 mRNA and the *Tol2* transposon transgene vector can be generated and purified using standard lab techniques and commercial kits. We emphasize that optimal transgenesis is contingent on fastidious DNA preparation and quality. To generate *Tol2* mRNA, we linearize the *pCS3FA-transposase* vector (Tol2kit plasmid number 396) (Kwan *et al*., 2007) by digestion with *NotI*, purify the plasmid template by agarose gel electrophoresis, and *in vitro* transcribe 5′ capped mRNA using Sp6 polymerase (mMESSAGE mMACHINE kit, Applied Biosystems/Ambion, Austin, TX). The resulting *Tol2* mRNA contains the SV40 late polyA signal sequence to mediate polyadenylation in the embryo upon injection. We store *Tol2* mRNA at −80 °C at 100 ng/μL in single-use aliquots of 2.5–3 μL. Our standard 10 μL injection mix contains 25 ng/μL transgene vector plus 25 ng/μL Tol2 mRNA, which translates into 2.5 μL from the 100 ng/μL *Tol2* mRNA stock.

Routine lab practice should include quality control of *Tol2* mRNA and preparation of new batches before working stocks are exhausted. *Tol2* mRNA quality can be reliably assessed by monitoring integration efficiency of an easily monitored transgene, such as *pDestTol2A2_ubi:EGFP* (*ubi:EGFP*) (Mosimann *et al*., 2011), over time. When co-injected with functional 25 ng/μL *Tol2* mRNA, *ubi:EGFP* yields extensive mosaic expression of EGFP which is still detectable 3–4 weeks post injection, with a high probability for germline transmission of *ubi:EGFP*. In comparison, injection of 25 ng/μL *ubi:EGFP* plasmid DNA alone leads to a markedly weaker EGFP mosaicism and the fluorescence fades over the course of a few weeks, as the

vector rarely integrates into the genome and the plasmid is degraded. Observation of adult EGFP expression is facilitated by UV headsets (''goggles'') that allow non-intrusive fluorescence observation (Mosimann *et al*., 2011), especially if working in the transparent *casper* background. To avoid inefficient transgenesis rates, *Tol2* mRNA batches that do not result in extensive EGFP mosaicism over several weeks should be discarded.

The purity of the *Tol2* transposon transgene DNA is critical to successful transgenesis. Contaminants, such as ethanol, chaotropic salts, bacterial DNA/RNA, or RNAse, severely impair embryo viability, and likely, *Tol2* mRNA stability. Plasmid mini preparations (Mini preps), done with the QIAprep Spin Miniprep kit (QIAgen Inc., Valencia, CA) or comparable reagents, are simple and yield DNA of sufficient purity for efficient transgenesis. The following additions to the standard miniprep protocol optimize yield and purity for injection-grade plasmid DNA (use of a vacuum manifold is highly recommended to speed up processing).

Additions to the standard QIAprep miniprep protocol:

- Use fresh and properly stored re-suspension buffer containing RNAse (i.e., QIAprep buffer P1 at 4 °C) for optimal digestion of bacterial RNA.
- After adding lysis buffer (P2) and gentle mixing by tube inversion, incubate minipreps for 5 min at room temperature to enhance bacterial lysis.
- After adding neutralization buffer (N3) and gentle mixing by tube inversion, incubate minipreps for 5–10 min at −20 °C (or on ice) to increase lipid and protein precipitation.
- To pellet lipid and protein debris completely, thus producing a cleaner plasmid-containing supernatant, centrifuge the precipitated sample 12–15 min at 13,000–14,000 rpm, preferably at 4 °C.
- With QIAprep kits, include the recommended PB buffer wash step, which inactivates residual RNAse and removes protein residue.
- After washing the columns (buffer PE) to remove ethanol, centrifuge for 2 min, then discard waste tubes. Leave the columns in fresh collection tubes at room temperature for a few minutes to evaporate the last traces of ethanol.
- Elute the plasmid DNA by addition of 50 μL ultra-pure water to the column membrane. Incubate 1 min, then centrifuge (1 min at 13,000–14,000 rpm). *Do not use elution buffer EB provided with the kit.*

This procedure yields highly concentrated minipreps (100–400 ng/μL) of average-sized plasmids, depending on culture growth time. To elute larger (>15 kb) plasmids off the column matrix, pre-heating of the elution water to 65 °C can increase yield. Injection-grade minipreps should give a 260/230 nm ratio of 1.8–1.9 (lower values indicate salt or ethanol contamination), and a 260/280 nm ratio of 2.0–2.3 (values below 2.0 indicate protein contamination). Pure plasmids can then be processed as described for microinjection into zebrafish embryos for Tol2-mediated transgenesis (Kwan *et al*., 2007; Villefranc *et al*., 2007). Transgenesis success depends on *Tol2* mRNA quality, plasmid vector purity and size, as well as the biological effects of the injected transgene.

III. *Tol2* Transgene Genetics

A. "Every Fish is Unique" – Basic Genetic Principles and Transgenesis Markers

When the *Tol2* transgenesis mix is injected into one-cell stage embryos, some, but not all, of the cells in cleavage will successfully catalyze the Tol2 integration. The resulting embryo is thus mosaic for the transgene, with each cell potentially carrying 0–10 integrations spread randomly across the 50 zebrafish chromosomes (Kawakami, 2007). Consequently, the germ cells of a mosaic F0 animal will differ in transgene integration and will transmit transgenic chromosomes at an unpredictable ratio to the next generation. The non-Mendelian nature of transgene inheritance typically yields 3–100% positive embryos, and each embryo will harbor an unpredictable number of independent insertions. For successful outcomes in transgenesis experiments, each fish of the first generation is considered a unique transgene carrier; predictable Mendelian inheritance of functional transgenes often requires several rounds of out-crossing (Fig. 3 and see also below).

Screening and line maintenance are facilitated with fluorescent reporters and transgenes engineered with a fluorescent marker in *cis*. Current zebrafish transgene markers express a fluorescent protein in a particular cell type or confined developmental structure and are cloned into the Tol2 transgene vector backbone. Currently, the choice for zebrafish transgene markers is rather limited and will greatly benefit from future additions. The cardiomyocyte-specific *cmlc2:EGFP* cassette is common in Tol2 cloning vectors (Kwan *et al*., 2007). Other options are the eye lens-labeling *α-crystallin:YFP* or *Cerulean* cassettes (Hesselson *et al*., 2009; Villefranc *et al*., 2007), as well as the skin-specific *krt3:RFP* (Liu *et al*., 2008). Various *lox* cassette transgenes (see also later) contain a fluorescent *lox* cassette, such as used in *FlEx (Flip Excision)* or *ubi:Switch* (Boniface *et al*., 2009; Mosimann *et al*., 2011).

One caveat to using transgenesis markers is potential activation of the transgene cargo by the marker's regulatory sequences, and *vice versa*. For example, we and others have found transgene cargo under control of different promoters to express aberrantly in cardiomyocytes when coupled in *cis* with *cmlc2:EGFP* (Mosimann *et al*., 2011). Potent enhancer elements in a transgene cassette can act on the promoter of the transgenesis marker and inappropriately activate the marker gene.

B. Transgenesis Screening Scheme

Microinjected zebrafish (F0 generation) are screened for transgene transmission by individually out-crossing to wild-type zebrafish (Fig. 3). Each *Tol2* transposon-injected F0 animal will give rise to a distinctive collection of F1 animals. F0 parents with transgene transmission to the F1 are subsequently called F0 founders. We routinely inject *AB* or *Tu* wild-type strains as the F0 and screen for F0 founders by mating to the *TL* line, which contains *tuepfel* (*tup*, also called leo^{t1}, recessive mutation causing spotted pigment instead of stripes) and *longfin* (lof^{dt2}, dominant

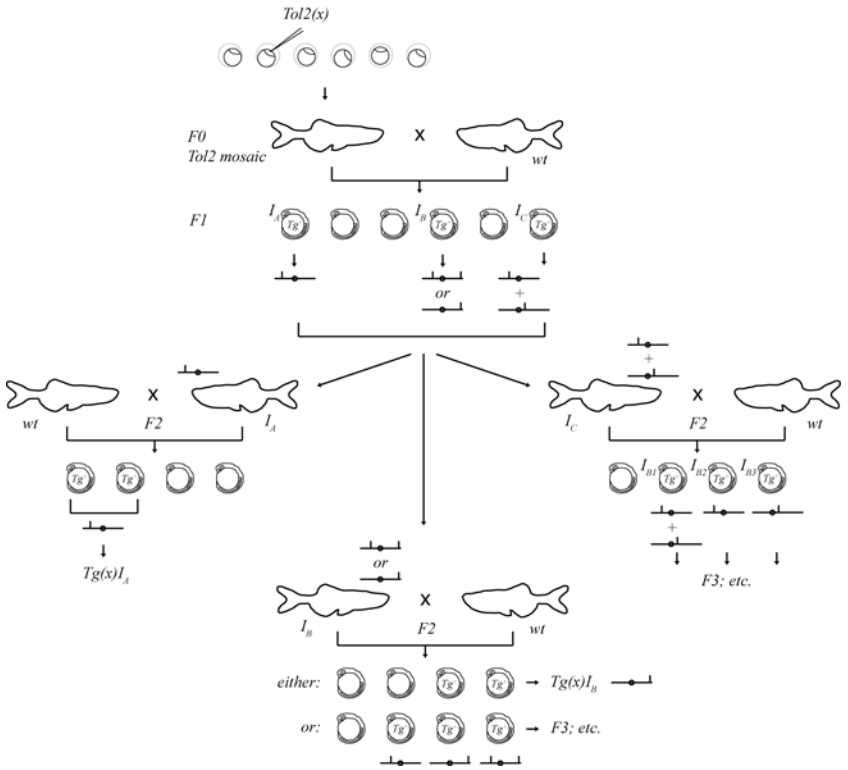


Fig. 3 Genetic scheme outlining single-insertion Tol2 transgenesis from a generic transgene vector (*Tol2(x)*). See text for details. Note the three different possible combinations of transgene insertions in the F1, labeled I_A, I_B, and I_C, and the resulting downstream crossing scheme required to obtain single-insertion transgenic strains. Chromosome schematics contain either one (I_A, I_C) or two (I_B) chromosomal insertions. In the case of two insertions on the same chromosome (I_B), the insertion number cannot be properly deduced by Mendelian genetics in the F1 but can be revealed in the F2 generation (or subsequent generations, depending on the centi-Morgan distance of the two insertions).

homozygous viable mutation causing overgrowing fins) mutations (ZFIN, http://zfin.org). Mating of F0 to *TL* enables us to separate the parents after the cross and introduces the *tup* and *L* markers into transgenic lines. We prioritize F0 animals with visible transgene marker expression as it indicates successful injection (but does not predict successful transgene transmission).

Individual F0 animals that provide offspring are consecutively numbered (1, 2, 3, etc., or Roman numerals) and kept in isolation. Their F1 embryos are subsequently screened for the transgene marker by fluorescence microscopy or by PCR. If a reasonable clutch size (50–100 embryos) is transgene-negative, the corresponding F0 animal is culled. If the clutch is positive, the positive F1 embryos are raised, and the F0 parent is kept for further rounds of outcrosses.

Each F1 offspring is still potentially unique due to unpredictable transgene segregation in the F0 germline. For example, fluorescent reporter transgenes may reveal mixed transgene numbers in F1 clutches by variable fluorophore intensity between siblings. Transgenes can also independently insert onto the same chromosome and cause complex inheritance due to chromosome cross-over (Fig. 3). We label the F1 offspring of a given F0 animal by adding a consecutive letter to the F0 designation (i.e., 2_A, 2_B, 2_C). Between two and five transgene-positive F1 animals from an F0 founder are again individually out-crossed to wild type to assess transgene segregation in the F2 generation. If the F2 embryos are 50% transgene-positive, then it is likely that the F1 parent contains a single transgenic insertion (or may have two or more closely linked insertions on the same chromosome; see also below). Higher F2 percentages reveal multiple transgene insertions. In the latter case, if required, several individual F2 animals from a single F1 parent are further out-crossed to wild type until 50% segregation is observed (Fig. 3).

Tol2 transposons can also integrate onto the same chromosome and segregate by cross-over in subsequent generations, resulting in higher than 50% transgene transmission in subsequent generations. In contrast, we rarely observe lower than 50% transgene segregation, which is indicative of transgene silencing, after the F1; such lines are not pursued further. We regularly detect by PCR silenced transgene copies in transgene-negative embryos from F2 clutches (thus, undetectable by fluorescence microscopy), which is probably due to heterochromatin-mediated silencing or defective transposition in the F0.

As we lack methods to permanently label individual zebrafish, this labor-, time-, and space-consuming crossing scheme results in rapidly expanding numbers of individual fish in isolation tanks. Therefore, we prioritize F1 animals with 50% segregation in the F2 generation because they have the greatest probability of harboring single-copy insertions.

C. Transgene Quality Assessment

Single-insertion zebrafish transgenics feature predictable inheritance patterns and expression, yet long-term transgene stability is critical to sustainable experimentation. Transgene silencing can occur at random and eliminate the utility of any established line. It is good practice to establish at least three (or more) independent transgenic lines for a given transgene. Several independent lines may further help to uncover individual position effects, e.g., non-specific transgene cargo expression in inappropriate cell types. Complementary analyses, such as *in situ* hybridization to detect Cre transgenes, also help to ensure faithful transgene expression.

When using fluorescent reporters, the commonly observed decrease in insertion copy number after the F1 generation will cause progressively dimmer, but more consistent, reporter fluorescence in subsequent generations until single-copy transgenes are established. Thus, F1 animals can lead to unsound overestimations of reporter potency. For the sake of subsequent reproducibility, drawing experimental conclusions from F1 animals should be discouraged.

The high transgenesis rates with Tol2 allow exhaustive screening for high-quality *lox* transgene insertions. Any transgene integration locus contributes chromatin effects that, due to unfavorable histone modifications or protein binding to DNA in the vicinity, can influence Cre access, reactivity at *lox* sites, and resulting recombination efficiency (see also below) (Hans *et al*., 2009, 2010; Mosimann *et al*., 2011). For each of our established *lox* transgenic lines, we had to test several independent founder lines to obtain an efficiently switching, high-quality *lox* transgene insertion.

D. Materials

- *pENTR5′* TOPO TA Cloning Kit; Invitrogen, Cat. No. K591-20
- *pENTR/D* TOPO Cloning Kit; Invitrogen, Cat. No. K2400-20
- LR Clonase® II Plus enzyme; Invitrogen, Cat. No. 12538-120
- Tol2 Kit plasmid collection
- mMESSAGE mMACHINE® SP6 Kit w/Manual, Cat. No. AM1340M
- UV filter goggles for EGFP detection, Modular Fluorescence Head set type: FHS, frame FHS/F-01, light source FHS/LS-1B, emission filter FHS/EF-2G2; BLS Ltd, Budapest, Hungary.

IV. CreER^{T2}-Controlled *lox* Recombination Using 4-OHT

Zebrafish embryos are ideal for lineage tracing studies as they develop rapidly, are nearly transparent, and can be staged precisely. High-quality Tol2 transgenes now greatly facilitate Cre/*lox* applications for lineage tracing in zebrafish. With TAM or 4-OHT for temporal control of CreER^{T2} activity and subsequent *lox* recombination, the components are now in place to generate lineage tracing zebrafish lines.

TAM and 4-OHT are lipid-soluble steroids that can pass through cell membranes and the chorion. The relative efficacies of TAM and 4-OHT in induction of CreER^{T2} are difficult to assess due to their instability and markedly different water solubilities (0.3 mg/L for TAM, 4-OHT is described as insoluble in water). Nonetheless, CreER^{T2} is sensitive to micromolar (μM) concentrations of both drugs in E3 embryo medium (Hans *et al*., 2009, 2010; Mosimann *et al*., 2011). The use of TAM, a pro-drug that requires metabolic transformation to produce the higher affinity 4-OHT, introduces a potential lag in CreER^{T2} response. Although several studies have successfully applied TAM (Hans *et al*., 2009, 2010), we predominantly use 4-OHT for our experiments. As ER^{T2} preferentially binds the *trans* (*Z*) 4-OHT isomer (Katzenellenbogen *et al*., 1984), we advise the use of *trans* TAM or 4-OHT for induction experiments.

A. 4-OHT Chemistry and Handling

Trans 4-OHT is commercially available from various sources (such as H7904 from Sigma-Aldrich, St. Louis, MO). We routinely create 10 mM working stock solutions

in ethanol and store 15 μL single-use aliquots in the dark at −20 °C. Published studies also report 4-OHT stocks of 25 mM in ethanol and 50 mM TAM in DMSO (Hans *et al*., 2009). Prolonged storage decreases the potency of dissolved 4-OHT over time, even with strict storage at −20 °C in the dark. The loss of potency is due to the tendency of 4-OHT to undergo *cis–trans* (*E-Z*) interconversion spontaneously in common laboratory solvents over time or when exposed to light. Desiccated 4-OHT powder kept in the dark at 2–8 °C is stable for years.

CreER^{T2}-mediated *lox* recombination occurs at concentrations as low as 0.5 μM TAM or 4-OHT (Hans *et al*., 2009; Mosimann *et al*., 2011). Zebrafish embryos tolerate 10 μM 4-OHT treatment without obvious phenotypic consequences, and we routinely apply 10 μM to achieve maximal CreER^{T2} responses to test new transgenics or induction time points. As in comparable mouse experiments, the optimal TAM/4-OHT dose for a specific CreER^{T2} driver paired with a *lox* switch requires individual elucidation (Hans *et al*., 2009).

B. The ''Golden Rule'' for Zebrafish Lineage Tracing: Paternal Cre, Maternal *lox* Switch

Successful Cre/*lox* experiments rely on faithful Cre activity in the investigated cell lineage. Because ''tissue-specific'' or ''induction-specific'' Cre transgenes can be active during oogenesis, we advise to use males as the source of the Cre driver. Genuine examples of recombinase transgenes that contribute maternally are *ubi:creER^{T2}* (Mosimann *et al*., 2011) and *hsp70:creER^{T2}* lines (Hans *et al*., 2009, 2010). Tissue-specific Cre/CreER^{T2} drivers may also unexpectedly contribute maternally due to chromatin position effects of individual transgene insertions. Undesired dominant maternal Cre contribution must therefore be avoided.

Therefore, we use males as the source for the Cre driver transgene to assure *de novo* transcription of Cre upon onset of zygotic transcription (Fig. 4). We recommend the ''golden rule'' to always cross Cre transgene males to *lox* switch females for tissue-specific developmental lineage tracing. Cre drivers and *lox* transgenes should also not be propagated together in double-transgenic animals because a leaky Cre transgene would lead to irreversible recombination of the *lox* transgene.

Maternally contributed Cre may further lead to an overestimation of *lox* excision efficiency of new *lox* switch transgenes, such as when testing with the commonly available *hsp70:cre*. Nonetheless, maternal CreER^{T2} from *hsp70:creER^{T2}* or *ubi:creER^{T2}* offers a potential tool to create discrete clones of cells with recombined *lox* switches by 4-OHT induction before the onset of zygotic transcription.

C. 4-OHT-mediated CreER^{T2} Induction

A simple format for embryo CreER^{T2} experiments is to array embryos prior to induction in fresh E3 into multi-well plates, such as 6- or 12-well format (with up to 35 embryos per well in 6-well plates). Collected embryos should be thoroughly rinsed using a tea strainer, and then transferred to fresh E3 prior to arraying. The multi-well format facilitates testing of several transgenic lines as well as subsequent

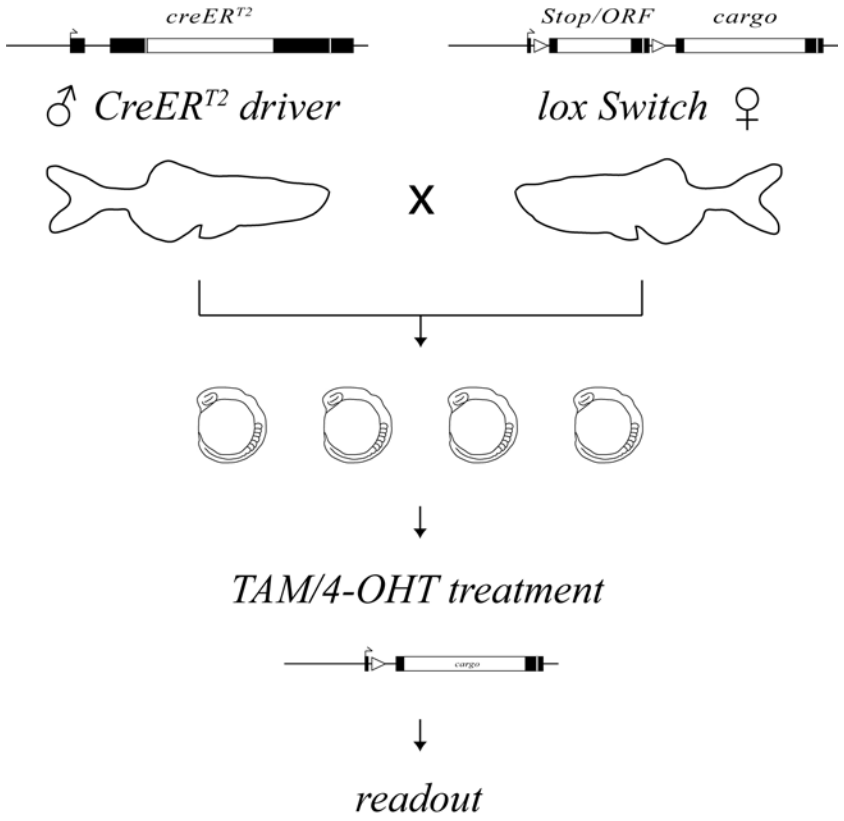


Fig. 4 Basic CreER^{T2}/*lox* experiment crossing scheme. The *lox* switch transgene is provided by the mother to avoid maternal contribution of CreER^{T2}. Experiments with regular Cre-expressing transgenes are done identically, yet without the TAM/4-OHT induction.

handling and microscopy. It also reduces the amount of 4-OHT consumed and allows the induction of embryos in different wells at different time points. De-chorionating the embryos did not significantly increase 4-OHT potency in our hands, yet may be advisable with less sensitive CreER^{T2} drivers.

The instability of 4-OHT in water dictates fresh preparation of 4-OHT-containing E3 induction medium immediately before treatment. In a 50 μL Falcon tube, the desired final volume of E3 medium is prepared and the required 4-OHT added with a pipette from a single-use aliquot, taken from −20 °C storage immediately prior to use. The drug will temporarily go out of solution when the ice-cold ethanol mixes with water and will be visible as a cloudy precipitate trail at the pipette tip. The precipitate can flake and drift to the medium surface. Brief vortexing effectively dissolves the precipitate. The prepared induction medium is best kept in the dark (such as in a drawer) while the embryos are prepared for immediate induction.

The E3 medium on the arrayed embryos is removed using a vacuum pipette with a fine tip while tilting the plate toward the experimenter to facilitate careful yet rapid aspiration starting from the top of the well. The embryos should not be allowed to dry, so apply the 4-OHT induction medium immediately after aspiration. We found 3–5 mL of medium to be sufficient for undisturbed embryo growth in 6-well plates while avoiding accidental spill-over of induction medium into non-induced control wells during handling. We routinely perform experiments combining 4-OHT with bioactive chemicals or Propylthiouracil (PTU) (to inhibit melanization for enhanced microscopy) without detrimental effects on 4-OHT or CreER^{T2} activity.

Consistent with its chemical properties, experimental observations indicate a short half-life of waterborne 4-OHT (Mosimann *et al*., 2011). Extended induction time windows may thus require repeated addition of fresh 4-OHT-containing E3 to maintain CreER^{T2} activity. However, 15 min 4-OHT treatment followed by thorough washing has been shown to trigger a strong response in the majority of cells when tested with *ubi:creER^{T2}*/*ubi:Switch* transgenics (Mosimann *et al*., 2011). Despite its light sensitivity, 4-OHT handling does not require working in dimmed rooms or shielding of embryo-containing plates while performing experiments on the laboratory bench. Incubators should, however, be dark, and handling on bench tops should be minimized during induction.

D. Induction Time Point Considerations

The timing of 4-OHT induction needs to be coordinated with 1) anticipated transgenic CreER^{T2} expression, 2) the 4-OHT activity window in embryo medium, and 3) the optimal 4-OHT concentration to activate sufficient CreER^{T2} to recombine a given *lox* switch at a given time point. Adding 4-OHT prior to CreER^{T2} expression might not leave sufficient active drug to activate newly synthesized recombinase. The dynamics of a given *lox* switch's expression is a further factor to gauge in determining the optimal induction time point. Although successful *lox* recombination is rapidly detectable by PCR, considerable lag time passes between 4-OHT administration and appearance of the actual *lox* switch transgene readout: cytoplasmic CreER^{T2} needs to bind 4-OHT, translocate to the nucleus, and engage two *lox* sites in recombination; the newly formed reporter requires mRNA transcription and subsequent protein translation, followed by proper folding of the fluorescent reporter protein and sufficient accumulation past the detection threshold. Clearly, the rapid development of the zebrafish embryo pushes the limit of how fast transcription- and protein translation-dependent readouts can be.

E. Materials

- E3 embryo medium.
- Cell culture-grade multi-well plates – 6-well, 12-well, or 24-well recommended to facilitate fluorescent microscopy.

- (Z)-4-OHT: Sigma-Aldrich, H7904, ≥98% (HPLC), powder, dissolved in ethanol to stock solutions of 10 mM, kept as single-use aliquots (15 μL) in reaction tubes in the dark at −20 °C.
- (Z)-TAM, Sigma-Aldrich, T5648, ≥99%, dissolved in DMSO to stock solutions of 50 mM.
- Bench top aspirator, equipped with fine pipette tip for controlled medium aspiration and connected to vacuum line.
- Closed and dark zebrafish incubator (21, 28.5, or 32 °C, depending on desired growth speed).

V. Discussion

Tol2-mediated single-integration transgenes have greatly invigorated the zebrafish community and triggered a new boom in creating transgenic zebrafish strains. The transgenesis and Cre/*lox* protocols outlined here will be refined in the near future as the field ventures into more complex molecular genetics using zebrafish. In particular, reliable new transgene markers to complement the few currently existing are clearly needed and will facilitate complex genetic experiments that combine multiple transgenes. Application of TAM/4-OHT transgene induction to adult zebrafish poses different challenges than manifested in embryo experiments (Jopling *et al*., 2010; Kikuchi *et al*., 2010) and warrants experimental attention. The rising number of transgenic zebrafish lines and increasingly sophisticated experimental strategies require community involvement to organize reagent sharing, maintenance of transgenic lines, and publication of detailed protocol descriptions. Nonetheless, we envision that the zebrafish model system will, in the near future, feature an array of transgenic experimental protocols that rivals those of *Drosophila* and the mouse.

Conflict of interest statement

L.I.Z. is a founder and stock holder of Fate, and a scientific advisor for Stemgent.

Acknowledgments

We would like to thank Charles K. Kaufman, Katy L. Lawson, and Alexa Burger for critical input on the described methodology and the manuscript. C.M. received support through an EMBO long-term fellowship and a HFSP long-term fellowship.

References

Bailey, J. M., Creamer, B. A., and Hollingsworth, M. A. (2009). What a fish can learn from a mouse: principles and strategies for modeling human cancer in mice. *Zebrafish* **6**, 329–337.

Bertrand, J. Y., and Traver, D. (2009). Hematopoietic cell development in the zebrafish embryo. *Curr. Opin. Hematol.* **16**, 243–248.

Boniface, E. J., Lu, J., Victoroff, T., Zhu, M., and Chen, W. (2009). FlEx-based transgenic reporter lines for visualization of Cre and Flp activity in live zebrafish. *Genesis* **47**, 484–491.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10887–10890.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757.

Feng, H., Langenau, D. M., Madge, J. A., Quinkertz, A., Gutierrez, A., Neuberg, D. S., Kanki, J. P., and Look, A. T. (2007). Heat-shock induction of T-cell lymphoma/leukaemia in conditional Cre/lox-regulated transgenic zebrafish. *Br. J. Haematol.* **138**, 169–175.

Hans, S., Freudenreich, D., Geffarth, M., Kaslin, J., Machate, A., and Brand, M. (2010). Generation of a non-leaky heat shock-inducible Cre line for conditional Cre/lox strategies in zebrafish. *Dev. Dyn.* **240**, 108–115.

Hans, S., Kaslin, J., Freudenreich, D., and Brand, M. (2009). Temporally-controlled site-specific recombination in zebrafish. *PLoS One* **4**, e4640.

Hesselson, D., Anderson, R. M., Beinat, M., and Stainier, D. Y. (2009). Distinct populations of quiescent and proliferative pancreatic beta-cells identified by HOTcre mediated labeling. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14896–14901.

Iyengar, A., Muller, F., and Maclean, N. (1996). Regulation and expression of transgenes in fish — a review. *Transgenic Res.* **5**, 147–166.

Jopling, C., Sleep, E., Raya, M., Marti, M., Raya, A., and Belmonte, J. C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606–609.

Katzenellenbogen, B. S., Norman, M. J., Eckert, R. L., Peltz, S. W., and Mangel, W. F. (1984). Bioactivities, estrogen receptor interactions, and plasminogen activator-inducing activities of tamoxifen and hydroxy-tamoxifen isomers in MCF-7 human breast cancer cells. *Cancer Res.* **44**, 112–119.

Kawakami, K. (2005). Transposon tools and methods in zebrafish. *Dev. Dyn.* **234**, 244–254.

Kawakami, K. (2007). Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* **8**(Suppl 1), S7.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144.

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., Evans, T., Macrae, C. A., Stainier, D. Y., and Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. *Nature* **464**, 601–605.

Kondrychyn, I., Garcia-Lecea, M., Emelyanov, A., Parinov, S., and Korzh, V. (2009). Genome-wide analysis of Tol2 transposon reintegration in zebrafish. *BMC Genomics* **10**, 418.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887–903.

Kozak, M. (1996). Interpreting cDNA sequences: some insights from studies on translation. *Mamm Genome* **7**, 563–574.

Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P., and Chien, C. B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099.

Langenau, D. M., Feng, H., Berghmans, S., Kanki, J. P., Kutok, J. L., and Look, A. T. (2005). Cre/lox-regulated transgenic zebrafish model with conditional myc-induced T cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6068–6073.

Le, X., Langenau, D. M., Keefe, M. D., Kutok, J. L., Neuberg, D. S., and Zon, L. I. (2007). Heat shock-inducible Cre/lox approaches to induce diverse types of tumors and hyperplasia in transgenic zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9410–9415.

Liu, X., Li, Z., Emelyanov, A., Parinov, S., and Gong, Z. (2008). Generation of oocyte-specifically expressed cre transgenic zebrafish for female germline excision of loxP-flanked transgene. *Dev. Dyn.* **237**, 2955–2962.

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62.

Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J., and Zon, L. I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177.

Sauer, B. (1987). Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol.* **7**, 2087–2096.

Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. *Methods* **14**, 381–392.

Sauer, B., and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5166–5170.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**, 70–71.

Stuart, G. W., McMurray, J. V., and Westerfield, M. (1988). Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**, 403–412.

Suster, M. L., Sumiyama, K., and Kawakami, K. (2009). Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* **10**, 477.

Thummel, R., Burket, C. T., Brewer, J. L., Sarras Jr., M. P., Li, L., Perry, M., McDermott, J. P., Sauer, B., Hyde, D. R., and Godwin, A. R. (2005). Cre-mediated site-specific recombination in zebrafish embryos. *Dev. Dyn.* **233**, 1366–1377.

Villefranc, J. A., Amigo, J., and Lawson, N. D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev. Dyn.* **236**, 3077–3087.

Yoshikawa, S., Kawakami, K., and Zhao, X. C. (2008). G2R Cre reporter transgenic zebrafish. *Dev. Dyn.* **237**, 2460–2465.

Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3789–3794.

CHAPTER 11

Use of Phage PhiC31 Integrase as a Tool for Zebrafish Genome Manipulation

James A. Lister

Department of Human and Molecular Genetics and Massey Cancer Center, Virginia Commonwealth University School of Medicine, Box 980033, Richmond, Virginia, USA

Abstract

On the strengths of forward genetics and embryology, the zebrafish *Danio rerio* has become an ideal system for the study of early vertebrate development. However, additional tools will be needed to perform more sophisticated analyses and to successfully carry this model into new areas of study such as adult physiology, cancer, and aging. As improved tools make transgenesis more and more efficient, the stage has been set for precise modification of the zebrafish genome such as are done in other model organisms. Genome engineering strategies employing site-specific recombinase (SSR) systems such as Cre/*lox* and Flp/*FRT* have become

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00011-2

invaluable to the study of gene function in the mouse and Drosophila and are now being exploited in zebrafish as well. My laboratory has begun to use another such SSR, the integrase encoded by the *Streptomyces* bacteriophage PhiC31, for manipulation of the zebrafish genome. The PhiC31 integrase promotes recombination between an attachment site in the phage (*attP*) and another on the bacterial chromosome (attB). Here I describe strategies using the PhiC31 integrase to mediate recombination of transgenes containing attP and attB sites in *cis* to excise elements with spatial and temporal specificity. The feasibility of the intramolecular recombination approach having been established, I discuss prospects for employing PhiC31 integrase for intermolecular recombination, i.e., transgene integration at defined sites in the genome.

I. Introduction

A. The Zebrafish as a Continually Expanding Model System

In the last 20 years, the zebrafish *Danio rerio* has become well-established as a model organism, particularly for the study of the genetics of early vertebrate development. In the future it is certain that it will become an even more widely used biomedical research model as its reach continues to extend into such realms as cancer (Amatruda *et al*., 2002), infectious disease (Davis *et al*., 2002), physiology (Briggs, 2002), behavior (Fetcho and Liu, 1998), and aging (Kishi *et al*., 2003). Forward genetic tools already exist and have been usefully exploited for the discovery of new genes and pathways. For example, large scale ENU mutagenesis screens first reported 14 years ago produced hundreds of mutants (Driever *et al*., 1996; Haffter *et al*., 1996), and many of these loci have since been identified at the molecular level. The sequencing of the zebrafish genome is approaching completion (http://www.sanger.ac.uk/Projects/D_rerio/) and reverse genetics methods are available (Nasevicius and Ekker, 2000), but additional and more precise genomic tools will be needed to address problems in these new areas.

B. Transgenic Technologies in Zebrafish

The ability to generate transgenic lines, by which exogenous DNA can be stably transmitted from generation to generation, is fundamental to the usefulness of the zebrafish as it is for many other model systems. Beyond simple transgenesis is the means to make specific manipulations in the zebrafish genome. In recent years, more efficient methods for obtaining germline integration of foreign genes in zebrafish have been described (Grabher and Wittbrodt, 2008), and increasingly transgenic lines are appearing in the literature (see zfin.org for an up-to-date and searchable listing). For example, using the Tol2 transposon originally isolated from medaka, Kawakami *et al*. (2004) reported germline transmission rates as high as 50%. Other high efficiency methods include a second transposon, Sleeping Beauty (Davidson *et al*.,

2003), I-*Sce I* meganuclease co-injection (Grabher *et al*., 2004), and injection of pseudotyped retrovirus into blastula stage embryos (Chen *et al*., 2002). Both lines expressing fluorescent markers to label particular cell populations, and to misexpress genes with spatial and temporal precision are needed to realize the full experimental potential of this organism.

C. Genome Modification Using Site-specific Recombinases

Site-specific recombinase approaches extend the usefulness of transgenes by offering a means for their manipulation after they are established in the genome. The Cre/*lox* and Flp/*FRT* are most commonly used in mice for the construction of conditional knockout alleles (reviewed in Branda and Dymecki, 2004), and for similar gene control and chromosomal rearrangement strategies in Drosophila (reviewed in Bischof and Basler, 2008). Both *loxP* and *FRT* sites comprise short inverted repeats around a core sequence, and recombination between two *loxP* sites or two *FRT* sites results in the exchange of sequences flanking each site but preservation of the *loxP/FRT* sites themselves, which may be substrates for additional reactions. Therefore, recombination between two *loxP*/*FRT* sites flanking a transgene in the presence of Cre/Flp causes the intervening sequence to be excised as a circular molecule. Reintegration is formally possible but thermodynamically highly unlikely. For this reason, Cre and Flp have proven to be very useful for deleting transgenes from chromosomes. While Cre and Flp can catalyze intermolecular recombination, as could occur between a *loxP/FRT* site on a plasmid and one on a host chromosome, this creates two *loxP/FRT* sites which can immediately recombine, reversing the integration event (although modifications have been devised to allow Cre and Flp to be used in this way; see Discussion).

The PhiC31 bacteriophage uses a distinct integrase to catalyze directed recombination at sequence-specific sites in the Streptomyces genome (Smith *et al*., 2010). The integrase of the PhiC31 bacteriophage mediates recombination between two different sequences: the attP site in the phage genome and the attB site in the *Streptomyces* chromosome (Fig. 1A). The integrase alone is sufficient for this reaction in the absence of other co-factors from the phage or from *Streptomyces* and can even catalyze recombination between attB and attP sites *in vitro* (Thorpe and Smith, 1998). Importantly, unlike Cre/*lox* and Flp/*FRT*, the attL and attR sites thus created cannot be acted upon by integrase alone, which requires additional, yet to be identified co-factors to reverse the reaction. Therefore, PhiC31 integrase would appear to have great potential as a tool for both intra- and inter-molecular recombination strategies.

Although its natural environment is prokaryotic it was shown that PhiC31 integrase could catalyze intramolecular recombination in human cells in culture (Groth *et al*., 2000). Using an intramolecular recombination (excision) assay, minimal attB and attP sites were defined to be 34 basepairs and 39 basepairs, respectively. Integration was demonstrated in human and mouse cells of plasmids bearing

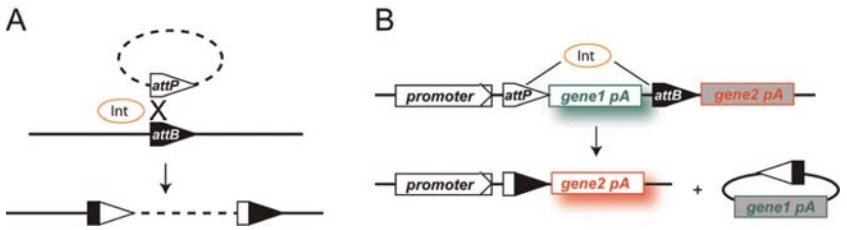


Fig. 1 Mechanism of PhiC31 integrase action. A) The integrase of the PhiC31 Streptomyces bacteriophage (Int) catalyzes recombination between an attachment site in the phage genome (attP) and a site in the bacterial genome (attB). In eukaryotic applications, it has generally been found that recombination is more efficient when the attP site is provided by the host genome rather than by the donor plasmid. B) Generalized strategy for use of Phic31 integrase (Int) for transgene excision. PhiC31 integrase (Int) can catalyze recombination between att sites in the same molecule; if these sites are in the same orientation the intervening sequenced will be excised. In this way, the integrase can be used to inactivate a gene (gene1) and/or activate a second gene (gene2) in a tissue-restricted and/or temporally-regulated manner. (See color plate.)

attB sites but not attP sites (Thyagarajan *et al*., 2001). Directed integration of an attB plasmid into endogenous loci, termed pseudo-attP sites, was observed to occur 10- to 20-fold more efficiently than random integration. In Drosophila, PhiC31 integrase has been shown to efficiently recombine attB-bearing plasmids with transgenes containing attP sites (Groth *et al*., 2004), and based on this elegant transgenesis strategies have been devised by a number of groups (Bateman *et al*., 2006; Bischof *et al*., 2007; Boy *et al*., 2010; Fish *et al*., 2007; Huang *et al*., 2009).

D. Cre and Flp Recombinases Function in Zebrafish

Several reports have now been made of the successful application of the Cre/*lox* system in zebrafish (Boniface *et al*., 2009; Collins *et al*., 2010; Dong and Stuart, 2004; Hans *et al*., 2009; Hesselson *et al*., 2009; Jopling *et al*., 2010; Kikuchi *et al*., 2010; Langenau *et al*., 2005; Pan *et al*., 2005; Seok *et al*., 2010; Thummel *et al*., 2005), and Flp recombinase as well appears to function as expected in the zebrafish embryo (Boniface *et al*., 2009). As is beginning to become clear in the mouse and Drosophila, considerable power lies in the ability to combine more than one tool simultaneously (Branda and Dymecki, 2004; Huang *et al*., 2009). Because they recognize different target sequences, PhiC31 and the other recombinases can potentially be used in parallel to allow independent manipulation of more than one transgene in a single organism. Finally, although there have been reports of chromosomal aberrations resulting from stable expression of PhiC31 integrase in human cells (Liu *et al*., 2009), the potential genotoxicity of Cre is also well-known (Loonstra *et al*., 2001; Schmidt *et al*., 2000), and PhiC31 may turn out to be a less toxic alternative. For these reasons, the development of PhiC31 integrase technology should complement other site-specific recombinases as they are adapted for use in the zebrafish. I focus here on intermolecular recombination (i.e., transgene excision)

using the PhiC31 integrase; successful intermolecular recombination (transgene integration) has not yet been described in the literature, but practice and experience with the former should facilitate the latter advance.

II. Rationale

A. PhiC31 Integrase Functions in Zebrafish

Although native to Streptomyces bacteria, the integrase encoded by the PhiC31 phage can also function in eukaryotic cells, including Drosophila (Groth *et al*., 2004), mouse (Belteki *et al*., 2003), frog embryos (Allen and Weeks, 2005) human cells in culture (Groth *et al*., 2000), as well as a variety of plant species (Khan *et al*., 2005; Lutz *et al*., 2004; Rubtsova *et al*., 2008; Thomson *et al*., 2010). It was therefore not surprising to find that this integrase could also function in zebrafish (Lister, 2010). While the native integrase was active in zebrafish embryos, we found using a plasmid-based reporter assay that a version of the integrase optimized for mouse codon usage (PhiC31o; (Raymond and Soriano, 2007)) gave over twice the frequency of recombination. Others have confirmed the basic utility of this approach (Lu *et al*., 2010), but for simplicity I focus below on results from my own laboratory.

To observe recombination in living embryos, we constructed a reporter (XIpGbR) comprising 1) the Xenopus EF1α promoter and rabbit β-globin intron, 2) a green fluorescent protein (GFP) open reading frame (ORF) and SV40 polyadenylation signal flanked by an attP site and an attB site, and 3) a Discosoma sp. variant red fluorescent protein DsRed-Express ORF and polyadenylation signal. To reduce the chance that the att sites might interfere with expression of the GFP and DsRed ORFs, they were shortened to the minimal lengths found to retain activity in a bacterial assay. It was confirmed that this plasmid could be recombined in bacteria and in embryos if and only if PhiC31 integrase was present. We also confirmed in these transient transgenesis experiments that the GFP and DsRed cassettes were expressed appropriately. (Interestingly, a construct in which the relative position of the attP and attB sites was reversed still expressed GFP, and could be recombined precisely and efficiently, however following excision no expression of DsRed could be observed.)

B. A Stable Reporter Line

A stable transgenic line was generated by co-injecting embryos with XIpGbR, in a Tol2 transposon backbone, along with Tol2 transposase mRNA. From multiple integrations a line (*vc2*) was selected that has shown strong GFP expression, but no detectable red fluorescence, for several generations. Injection of transgenic embryos with mRNA encoding the PhiC31 integrase, however, produces widespread red fluorescence (Fig. 2). Sequencing of PCR products from genomic DNA confirms precise excision of the GFP cassette. We have also constructed a red-to-green reporter and are currently attempting to establish lines with this (Fig. 3). This reporter was assembled by Gateway cloning, described below.

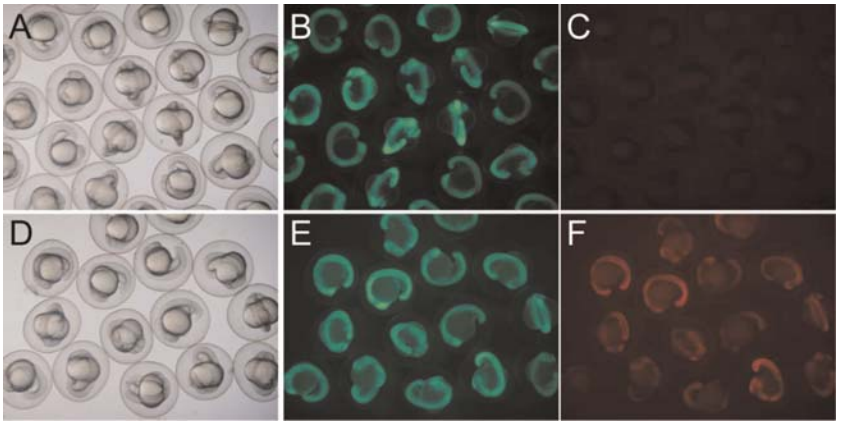


Fig. 2 Recombination in the green-to-red reporter strain, *vc2*. A)–C) uninjected embryos (shown at 18-somite stage) express GFP strongly (B) but do not express DsRed (C). D)–F) Injection of messenger RNA encoding codon-optimized PhiC31 integrase (PhiC31o) induces widespread transgene recombination and expression of DsRed. (Due in part to significant maternal GFP expression, GFP signal is not seen to diminish.) (See color plate.)

The complementary arm to the development of transgenes flanked by attachment sites is the isolation of tissue-specific drivers, for expressing the recombinase in restricted patterns. We have begun to use the green-to-red reporter line to test candidate promoters for appropriateness as drivers in transient assays (Fig. 4) prior to using them to generate stable lines. The reporter line can then be used to screen prospective founders.

C. Conditional Activation of PhiC31 Integrase

While spatial control of recombination may be obtained by tissue-specific expression of integrase, temporal control is another highly desirable feature. Temporal

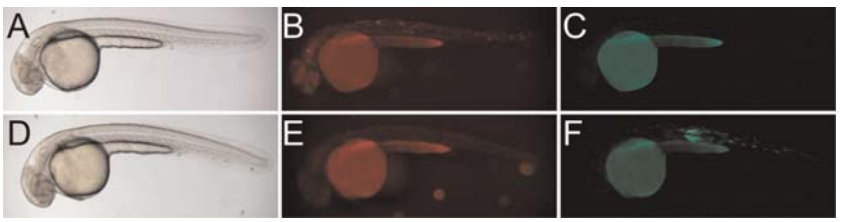


Fig. 3 Demonstration of a red-to-green PhiC31 integrase reporter generated by Gateway cloning. Wild-type embryos were injected with the plasmid pDestTol2pA2-XIpRbG alone (A–C) or with PhiC31o mRNA (D–F). In the absence of integrase activity mosaic DsRed expression (B) but no GFP expression (C) is observed. With addition of PhiC31o, only mosaic GFP expression (F) is seen. (See color plate.)

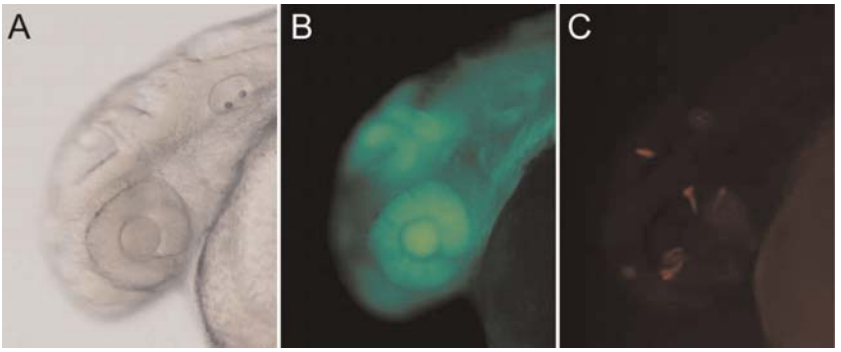


Fig. 4 Tissue-specific recombination. A *vc2* (green-to-red reporter) embryo injected with the plasmid pDestTol2pA2-mitfa0.9-PhiC31o-pA is shown at approximately 36 h post-fertilization A) brightfield, B) near ubiquitous GFP expression, C) recombination and expression of DsRed occurs in a small subset of cells. (See color plate.)

control of recombination in the Cre-lox system has been achieved by fusing the recombinase to a ligand binding domain variant of the estrogen receptor, rendering the chimeric protein inactive unless in the presence of 4-hydroxytamoxifen (4-OHT). It was reported that PhiC31 integrase activity can also be regulated by fusion to the ER ligand-binding domain (Sharma *et al*., 2008). We tested the inducibility of an analogous PhiC31o-estrogen receptor fusion protein in zebrafish. Embryos from the green-to-red reporter transgenic line were injected with mRNA encoding PhiC31o alone or as an N-terminal fusion to the estrogen receptor variant ERT2. Embryos treated with DMSO or 4-OHT beginning 60–90 min after fertilization were examined at approximately 30 h post-fertilization for the presence of DsRed-expressing cells. Recombination was only observed with PhiC31o-ERT2 injection in the presence of 4-OHT (Fig. 5); however the fraction of DsRed-positive cells was much lower with PhiC31o-ERT2 than with PhiC31o alone. The precise reasons for the reduced activity of the fusion protein have not yet been determined; however, these results indicate that improvements to this approach may be required to bring its efficiency up to that of constitutively-active integrase. The present version may still be useful for situations where mosaic recombination is desired or sufficient.

D. Gateway-compatible Vectors

The use of multisite Gateway cloning technology has facilitated the development of complex vectors for zebrafish transgenesis. To take advantage of this modular approach, we have generated and tested a number of entry vectors for use with the Tol2 kit described by Kwan *et al*. (2007). A list of these vectors is given in Table I. (It is important to note that although there is some shared terminology, the attachment sites and recombinases used in the Gateway system are derived from the lambda phage and are distinct from, and do not cross-react with, those of

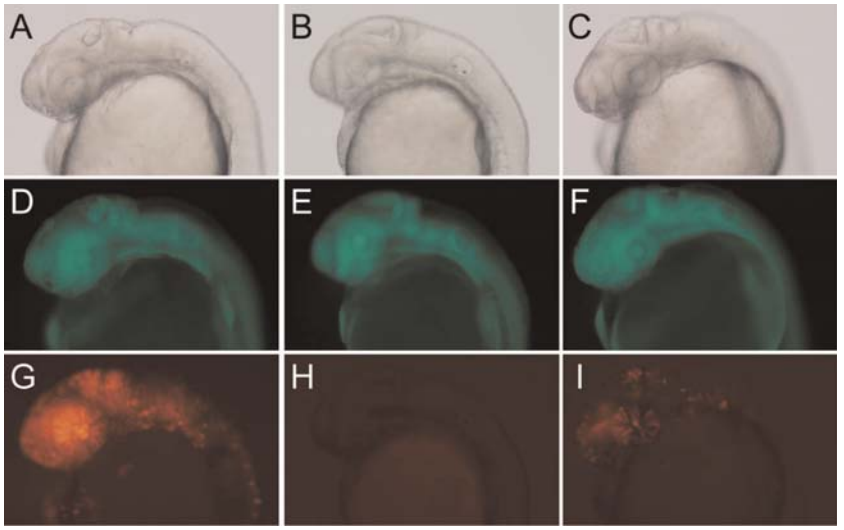


Fig. 5 Tamoxifen-dependent recombination. Individual *vc2* green-to-red reporter embryos injected with mRNA for PhiC31o (A, D, G) or PhiC31o-ERT2 (B, C, E, F, H, I) are shown. With addition of DMSO alone (B, E, H) no DsRed expression is observed, while addition of 300 nM 4-hydroxytamoxifen (C, F, I) induces recombination and red fluorescence (I), although to a much lesser degree than non-chimeric PhiC31o integrase (G). (See color plate.)

PhiC31.) In addition to middle entry vectors, which can be combined with a promoter of choice to generate tissue-specific integrase "drivers", we have constructed integrase targets by inserting the EF1α promoter/β-globin intron/attP-attB cassette as a unit into a 5′ entry vector. Versions include flanked GFP-polyA, DsRed-polyA, and polyA alone. These may thus be combined with a middle entry vector of choice (and 3′ entry and destination vectors) in a multi-site Gateway reaction to generate expression clones in which the middle ORF is only expressed after the construct is

Table I
Gateway entry vectors for use with the PhiC31 integrase system

Name	Type	Description
p5E-XIpGb	5′ entry	EF1α promoter w/excisable GFP cassette
p5E-XIpRb	5′ entry	EF1α promoter w/excisable DsRed cassette
p5E-XIppAb	5′ entry	EF1α promoter w/stop cassette
pME-PhiC31o	Middle	codon-optimized integrase
pME-PhiC31o-ERT2	Middle	4-hydroxytamoxifen-inducible integrase
pME-nlsPhiC31o	Middle	nuclear-localized integrase

acted upon by the PhiC31 integrase. An example of a red-to-green reporter made in this fashion is shown in Fig. 3. (We also tried flanking GFP-polyA with attP and attB sites within a middle entry clone, in combination with a 3′ entry clone containing DsRed with its own start site, but found that the DsRed was only weakly expressed following excision.)

III. Materials and Methods

pME-PhiC31o was generated by removing the codon-optimized integrase sequence from pPhiC31o (Addgene plasmid 13794; (Raymond and Soriano, 2007)) with EcoRI and XbaI and inserting it into the plasmid pENTR3C-CS, a middle entry vector containing the multiple cloning site of pCS2. p5E-XIpRb was constructed by inserting a BglII-XhoI fragment containing the Xenopus EF1α promoter and rabbit β-globin intron, attP site, DsRed-Express with poly A signal, and attB site into the BamHI-XhoI sites of p5E-MCS (Kwan *et al*., 2007). p5E-mitfa0.9 contains 0.9 kb of sequence upstream of the translational start site of the *mitfa* gene, and was made by replacing transferring the *mitfa* promoter from the plasmid pNP-P+ (Lister *et al*., 2001) to the 5′ entry vector p5E-MCS (Kwan *et al*., 2007) as a SalI-HindIII fragment. pDestTol2pA2-XIpRb-EGFP-pA was generated in a multi-site Gateway LR reaction (Invitrogen, Carlsbad, CA) with the entry vectors p5E-XIpRb, pME-EGFP, p3E-polyA, and the destination vector pDestTol2pA2 (Kwan *et al*., 2007). pDestTol2pA2-mitfa0.9-PhiC31o-pA was generated in a multi-site Gateway LR reaction with the entry vectors p5E-mitfa0.9, pME-PhiC31o, p3E-polyA, and the destination vector pDestTol2pA2. Additional cloning details for all plasmids are available upon request.

The plasmids pCS2P+PhiC31o and pCS2P+PhiC31o-ERT2 have been previously described, as has generation of the green-to-red reporter transgenic line, *Tg (XlEef1a1:attP-GFP-attB-DsRed2)*^{vc2} (Lister, 2010). Messenger RNA for each integrase was synthesized using the SP6 mMessage mMachine kit (Ambion/Applied Biosystems, Carlsbad, CA) following plasmid linearization by restriction digest with Not I or BssHII. Microinjections were performed using a pressure injection apparatus from Applied Scientific Instrumentation, on a Nikon stereodissection microscope.

To activate the PhiC31o-ERT2 fusion, 4-OHT (Sigma cat. no. H7904, made up at a stock concentration of 300 uM in DMSO) was added to embryos at 60–90 min post-fertilization to a final concentration of 300 nM, with an equal volume of DMSO alone added to the control dish.

IV. Discussion

Use of the Cre/*lox* system is becoming more widespread in zebrafish (Collins *et al*., 2010; Dong and Stuart, 2004; Feng *et al*., 2007; Jopling *et al*., 2010; Kikuchi *et al*., 2010; Langenau *et al*., 2005; Pan *et al*., 2005; Thummel *et al*., 2005; Wang *et al*., 2008),

and includes conditional approaches (Boniface *et al*., 2009; Hans *et al*., 2009). Intramolecular recombination in zebrafish mediated by the PhiC31 integrase represents another tool now available for the zebrafish researcher (Lister, 2010; Lu *et al*., 2010). The XIpGbR green-to-red transgenic line we previously established (Lister, 2010), *vc2*, is functionally analogous to a G2R Cre reporter that has been reported (Yoshikawa *et al*., 2008), and should be useful for the development and screening of integrase-expressing driver lines, as well as for lineage analysis. To facilitate the construction of drivers and reporters/effectors, we have adapted a number of components for use with Gateway cloning (Table I).

An obvious goal for those working with PhiC31 integrase in zebrafish is to demonstrate its utility for targeted transgene integration, and a number of labs including my own are working on this. The issue of exactly where DNA integrates in the genome is important because of the influence of position effects on transgene expression. This will only become more relevant as more investigators seek to study the function of promoters and proteins *in vivo* at single basepair or single amino acid resolution, and wish to compare variants in the same chromosomal context. While the efficiency of production of transgenic zebrafish continues to improve, at present there is no way to control the locus of transgene integration. In mice, this is typically achieved through homologous recombination in ES cells followed by chimeric embryo generation to establish the alteration in the germline. In zebrafish, ES-like cells have been isolated (Fan *et al*., 2004) and homologous recombination has been reported (Fan *et al*., 2006), but it has not yet been demonstrated that cells that have undergone the necessary selection regimen retain germline potential when put back into embryos. The lone report of targeted gene insertion in zebrafish has come from a group using Cre recombinase along with mutant *loxP* sites that once recombined cannot efficiently perform the reverse reaction with each other (Liu *et al*., 2007), but only one such event could be verified out of 80,000 injected embryos.

Based on work in other vertebrates, namely mouse and frogs (Allen and Weeks, 2005; Belteki *et al*., 2003), intermolecular recombination in zebrafish mediated by PhiC31 integrase should be feasible. To make heritable genomic manipulations, it is necessary that recombination occurs in germ cells. By raising integrase-injected *vc2* transgenic embryos to adulthood we determined that PhiC31 integrase is active in the zebrafish germline, as it is in mouse (Belteki *et al*., 2003). Breeding individual transgenic fish to non-transgenic mates produced a mix of offspring with uniform GFP or DsRed expression almost half the time, regardless of the sex of the transgenic adult (Lister, 2010). This showed that recombination can take place in the germline of either both males and females. Moreover, widespread expression of PhiC31 mRNA in early embryos had no obvious deleterious effect on development.

Our initial experiments suggested that codon optimization had a significant effect on the activity of PhiC31 integrase in zebrafish cells (Lister, 2010), and additional refinements should improve the likelihood of success of both intra- and intermolecular recombination approaches. When tested side by side with Cre in recombination assays on extrachromosomal and integrated targets in mammalian cells, PhiC31 integrase was found to be approximately 50% as active as Cre on plasmids and

10% on integrated transgenes (Andreas *et al*., 2002). However, addition of a C-terminal nuclear localization signal increased these to 80 and 50%. Other possibilities for optimizing the efficiency of the system and improving the chances of integration include employing hyperactive PhiC31 mutants (Keravala *et al*., 2009), as well as the identification and, if possible, specific elimination of interactions with inhibitory proteins (Chen *et al*., 2006; Wang *et al*., 2010) through directed mutagenesis. The successes observed in zebrafish with the PhiC31 integrase within a relatively brief time suggest that the goal of additional uses, notably directed transgenesis through intermolecular recombination (Bateman *et al*., 2006; Groth *et al*., 2004), will eventually be realized.

V. Summary

The PhiC31 integrase is functional in zebrafish cells and at present offers an alternative, or complement, to existing tools such as Cre and Flp for regulated and precise recombination of transgenes. In the future it may be possible to extend the use of this tool by exploiting its inherent unidirectionality to direct the integration of transgenes at pre-determined locations in the genome.

Acknowledgments

I would like to thank AnhThu Nguyen, Katie Lunney, and Leyla Peachy for fish care. This work was supported by grant number R21 HD055311 from the National Institute of Child Health and Human Development.

References

Allen, B. G., and Weeks, D. L. (2005). Transgenic Xenopus Laevis embryos can be generated using phiC31 integrase. *Nat. Methods* **2**, 975–979.

Amatruda, J. F., Shepard, J. L., Stern, H. M., and Zon, L. I. (2002). Zebrafish as a cancer model system. *Cancer Cell* **3**, 229–231.

Andreas, S., Schwenk, F., Kuter-Luks, B., Faust, N., and Kuhn, R. (2002). Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res.* **30**, 2299–2306.

Bateman, J. R., Lee, A. M., and Wu, C. T. (2006). Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. *Genetics* **173**, 769–777.

Belteki, G., Gertsenstein, M., Ow, D. W., and Nagy, A. (2003). Site-specific cassette exchange and germline transmission with mouse ES cells expressing phiC31 integrase. *Nat. Biotechnol.* **21**, 321–324.

Bischof, J., and Basler, K. (2008). Recombinases and their use in gene activation, gene inactivation, and transgenesis. *Methods Mol. Biol.* **420**, 175–195.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. U. S. A.* **140**, 3312–3317.

Boniface, E. J., Lu, J., Victoroff, T., Zhu, M., and Chen, W. (2009). FlEx-based transgenic reporter lines for visualization of Cre and Flp activity in live zebrafish. *Genesis* **47**, 484–491.

Boy, A. L., Zhai, Z., Habring-Muller, A., Kussler-Schneider, Y., Kaspar, P., Lohmann, I. (2010). Vectors for efficient and high-throughput construction of fluorescent Drosophila reporters using the PhiC31 site-specific integration system. *Genesis* **48**, 452–456.

Branda, C. S., and Dymecki, S. M. (2004). Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell.* **6**, 7–28.

Briggs, J. (2002). The zebrafish: a new model organism for integrative physiology. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **282**, R3–R9.

Chen, J. Z., Ji, C. N., Xu, G. L., Pang, R. Y., Yao, J. H., Zhu, H. Z., Xue, J. L., Jia, W. (2006). DAXX interacts with phage PhiC31 integrase and inhibits recombination. *Nucleic Acids Res.* **34**, 6298–6304.

Chen, W., Burgess, S., Golling, G., Amsterdam, A., and Hopkins, N. (2002). High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish. *J. Virol.* **76**, 2192–2198.

Collins, R. T., Linker, C., and Lewis, J. (2010). MAZe: a tool for mosaic analysis of gene function in zebrafish. *Nat. Methods* **7**, 219–223.

Davidson, A. E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., Sivasubbu, S., Cliff, M. P., Hackett, P. B., Ekker, S. C. (2003). Efficient gene delivery and gene expression in zebrafish using the Sleeping Beauty transposon. *Dev. Biol.* **263**, 191–202.

Davis, J., Clay, H., Lewis, J., Ghori, N., Herbomel, P., Ramakrishnan, L. (2002). Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. *Immunity* **17**, 693–702.

Dong, J., and Stuart, G. W. (2004). Transgene manipulation in zebrafish by using recombinases. *Methods Cell Biol.* **77**, 363–379.

Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46.

Fan, L., Crodian, J., and Collodi, P. (2004). Production of zebrafish germline chimeras by using cultured embryonic stem (ES) cells. *Methods Cell Biol.* **77**, 113–119.

Fan, L., Moon, J., Crodian, J., and Collodi, P. (2006). Homologous recombination in zebrafish ES cells. *Transgenic Res.* **15**, 21–30.

Feng, H., Langenau, D. M., Madge, J. A., Quinkertz, A., Gutierrez, A., Neuberg, D. S., Kanki, J. P., Look, A. T. (2007). Heat-shock induction of T-Cell lymphoma/leukaemia in conditional Cre/lox-regulated transgenic zebrafish. *Br. J. Haematol.* **138**, 169–175.

Fetcho, J. R., and Liu, K. S. (1998). Zebrafish as a model system for studying neuronal circuits and behavior. *Ann. N. Y. Acad. Sci.* **860**, 333–345.

Fish, M. P., Groth, A. C., Calos, M. P., and Nusse, R. (2007). Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. *Nat. Protoc.* **2**, 2325–2331.

Grabher, C., Joly, J. S., and Wittbrodt, J. (2004). Highly efficient zebrafish transgenesis mediated by the meganuclease I-SceI. *Methods Cell Biol.* **77**, 381–401.

Grabher, C., and Wittbrodt, J. (2008). Recent advances in meganuclease-and transposon-mediated transgenesis of Medaka and zebrafish. *Methods Mol. Biol.* **461**, 521–539.

Groth, A. C., Fish, M., Nusse, R., and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775–1782.

Groth, A. C., Olivares, E. C., Thyagarajan, B., and Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5995–6000.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, Danio Rerio. *Development* **123**, 1–36.

Hans, S., Kaslin, J., Freudenreich, D., and Brand, M. (2009). Temporally-controlled site-specific recombination in zebrafish. *PLoS ONE* **4**, e4640.

Hesselson, D., Anderson, R. M., Beinat, M., and Stainier, D. Y. (2009). Distinct populations of quiescent and proliferative pancreatic beta-cells identified by HOTcre mediated labeling. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14896–14901.

Huang, J., Zhou, W., Dong, W., Watson, A. M., and Hong, Y. (2009). From the cover: directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8284–8289.

Jopling, C., Sleep, E., Raya, M., Marti, M., Raya, A., Belmonte, J. C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606–609.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell.* **7**, 133–144.

Keravala, A., Lee, S., Thyagarajan, B., Olivares, E. C., Gabrovsky, V. E., Woodard, L. E., Calos, M. P. (2009). Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. *Mol. Ther.* **17**, 112–120.

Khan, M. S., Khalid, A. M., and Malik, K. A. (2005). Phage phiC31 integrase: a new tool in plastid genome engineering. *Trends Plant Sci.* **10**, 1–3.

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., Evans, T., Macrae, C. A., Stainier, D. Y., Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. *Nature* **464**, 601–605.

Kishi, S., Uchiyama, J., Baughman, A. M., Goto, T., Lin, M. C., Tsai, S. B. (2003). The zebrafish as a vertebrate model of functional aging and very gradual senescence. *Exp. Gerontol.* **38**, 777–786.

Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P., Chien, C. B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099.

Langenau, D. M., Feng, H., Berghmans, S., Kanki, J. P., Kutok, J. L., Look, A. T. (2005). Cre/lox-regulated transgenic zebrafish model with conditional Myc-induced T cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6068–6073.

Lister, J. A. (2010). Transgene excision in zebrafish using the phiC31 integrase. *Genesis* **48**, 137–143.

Lister, J. A., Close, J., and Raible, D. W. (2001). Duplicate Mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev. Biol.* **237**, 333–344.

Liu, J., Skjorringe, T., Gjetting, T., and Jensen, T. G. (2009). PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC Biotechnol.* **9**, 31.

Liu, W. Y., Wang, Y., Qin, Y., Wang, Y. P., and Zhu, Z. Y. (2007). Site-directed gene integration in transgenic zebrafish mediated by Cre recombinase using a combination of mutant Lox sites. *Mar. Biotechnol. (NY)* **9**, 420–428.

Loonstra, A., Vooijs, M., Beverloo, H. B., Allak, B. A., van Drunen, E., Kanaar, R., Berns, A., Jonkers, J. (2001). Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9209–9214.

Lu, J., Maddison, L. A., and Chen, W. (2010). PhiC31 integrase induces efficient site-specific excision in zebrafish. *Transgenic Res.* **20**, 183–189.

Lutz, K. A., Corneille, S., Azhagiri, A. K., Svab, Z., and Maliga, P. (2004). A novel approach to plastid transformation utilizes the phiC31 phage integrase. *Plant J.* **37**, 906–913.

Nasevicius, A., and Ekker, S. (2000). Effective targeted gene 'Knockdown' in zebrafish. *Nat. Genet.* 216–220.

Pan, X., Wan, H., Chia, W., Tong, Y., and Gong, Z. (2005). Demonstration of site-directed recombination in transgenic zebrafish using the Cre/loxP system. *Transgenic Res.* **14**, 217–223.

Raymond, C. S., and Soriano, P. (2007). High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162.

Rubtsova, M., Kempe, K., Gils, A., Ismagul, A., Weyen, J., Gils, M. (2008). Expression of active streptomyces phage phiC31 integrase in transgenic wheat plants. *Plant Cell Rep.* **27**, 1821–1831.

Schmidt, E. E., Taylor, D. S., Prigge, J. R., Barnett, S., and Capecchi, M. R. (2000). Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13702–13707.

Seok, S. H., Na, Y. R., Han, J. H., Kim, T. H., Jung, H., Lee, B. H., Emelyanov, A., Parinov, S., Park, J. H. (2010). Cre/loxP-regulated transgenic zebrafish model for neural progenitor-specific oncogenic Kras expression. *Cancer. Sci.* **101**, 149–154.

Sharma, N., Moldt, B., Dalsgaard, T., Jensen, T. G., and Mikkelsen, J. G. (2008). Regulated gene insertion by steroid-induced PhiC31 integrase. *Nucleic Acids Res.* **36**, e67.

Smith, M. C., Brown, W. R., McEwan, A. R., and Rowley, P. A. (2010). Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem. Soc. Trans.* **38**, 388–394.

Thomson, J. G., Chan, R., Thilmony, R., Yau, Y. Y., and Ow, D. W. (2010). PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. *BMC Biotechnol.* **10**, 17.

Thorpe, H. M., and Smith, M. C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5505–5510.

Thummel, R., Burket, C. T., Brewer, J. L., Sarras Jr., M. P., Li, L., Perry, M., McDermott, J. P., Sauer, B., Hyde, D. R., Godwin, A. R. (2005). Cre-mediated site-specific recombination in zebrafish embryos. *Dev. Dyn.* **233**, 1366–1377.

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S., and Calos, M. P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell. Biol.* **21**, 3926–3934.

Wang, B. Y., Xu, G. L., Zhou, C. H., Tian, L., Xue, J. L., Chen, J. Z., Jia, W. (2010). PhiC31 integrase interacts with TTRAP and inhibits NFkappaB activation. *Mol. Biol. Rep.* **37**, 2809–2816.

Wang, L., Zhang, Y., Zhou, T., Fu, Y. F., Du, T. T., Jin, Y., Chen, Y., Ren, C. G., Peng, X. L., Deng, M., Liu, T. X. (2008). Functional characterization of Lmo2-Cre transgenic zebrafish. *Dev. Dyn.* **237**, 2139–2146.

Yoshikawa, S., Kawakami, K., and Zhao, X. C. (2008). G2R Cre reporter transgenic zebrafish. *Dev. Dyn.* **237**, 2460–2465.

CHAPTER 12

Method for Somatic Cell Nuclear Transfer in Zebrafish

Kannika Siripattarapravat[*] and Jose B. Cibelli[†,‡]

[*]Department of Pathology, Faculty of Veterinary Medicine, Kasetsart University, Bangkok, Thailand

[†]Departments of Animal Science and Physiology, Michigan State University, East Lansing, Michigan, USA

[‡]LARCell, ProgramaAndaluz de MedicinaRegenerativa, Seville, Spain

Abstract

Somatic cell nuclear transfer (SCNT) has been a well-known technique for decades and widely applied to generate identical animals, including ones with genetic alterations. The system has been demonstrated successfully in zebrafish. The elaborated requirements of SCNT, however, limit reproducibility of the established model to a few groups in zebrafish research community. In this chapter, we meticulously outline each step of the published protocol as well as preparations of equipments and reagents used in zebrafish SCNT. All describable detailed-tips are elaborated in texts and figures.

Zebrafish has been increasingly utilized as a vertebrate model for human diseases and developmental biology as it possesses advantageous characteristics, many of which are far better than other established animal models. As its uses widen among laboratories, the demand for sophisticated genetic manipulations only grows, i.e., conditional knock in, knockout, and knockdown among others. Many research groups have already established a wide variety of approaches for generation of both

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00012-4

transgenic and mutant zebrafish lines in an effort to meet such a demand (Amsterdam *et al*., 2004; Balciunas *et al*., 2004; Doyon *et al*., 2008; Grunwald and Streisinger, 1992; Meng *et al*., 2008; Nagayoshi *et al*., 2008; Nasevicius and Ekker, 2000; Nüsslein-Volhard and Dahm, 2002; Solnica-Krezel *et al*., 1994). However, there is still ample room for improvement, mostly on the establishment of a first-generation stable mutant founder animal.

Somatic cell nuclear transfer (SCNT) offers the possibility of introducing germline genetic modifications, more specifically in animal species in which embryonic stem cell – chimera system is not readily available (Wilmut *et al*., 1997). This can be done by making genetic modifications, including but not limited to homologous recombination, in cultured somatic cells. The cells can be selected and verified for their identity, the insertion of the transgene and subsequently transplanted into eggs that their genetic materials have been removed. Ultimately fertile mutant cloned offspring can be generated. The technique for nuclear transfer in zebrafish has been previously described (Lee *et al*., 2002); however, up until now, there is no mutant fish line generated by this technique. We have done a comprehensive overhaul of the described SCNT method in zebrafish (Siripattarapravat *et al*., 2009b) streamlining the technique, and demonstrating that it can be implemented in the most common strains, AB and Tubingen.

I. Methods

We described a simple method that uses mature-arrested eggs at metaphase II of meiosis (MII) as recipients in which their genomes are completely inactivated using a laser-firing device. The donor cells are delivered through the micropyle prior to activation of the reconstructed embryos. In doing so, the recipient eggs are at MII stage from the time of nuclear transfer until we purposely trigger egg activation. Cloned embryos are manipulated with their chorion intact; thus less injury is inflicted on the embryos. For verification of cloned zebrafish, we describe a genotyping method using selected set of 11 single nucleotide polymorphism (SNP) markers (Siripattarapravat *et al*., 2009b), and also perform a phenotypic screening and a karyotyping by replication-banding as well (Amores and Postlethwait, 1999). The following is a step-by-step protocol for zebrafish SCNT, including a preparation of Chinook salmon ovarian fluid, easily one of most important reagents for the success of our technique (Fig. 1).

A. Chinook Salmon Ovarian Fluid (CSOF)

The ovarian fluid of Chinook salmon (*Oncorhynchus tshawytscha*) is used because it known as the best source of zebrafish-egg holding media. In our experience, Chinook Salmon Ovarian Fluid (CSOF) could maintain zebrafish eggs at metaphase II of meiosis for up to 6 h (Siripattarapravat *et al*., 2009a). With the help of egg taker's crews (Michigan Department of Natural Resources), Chinook salmon

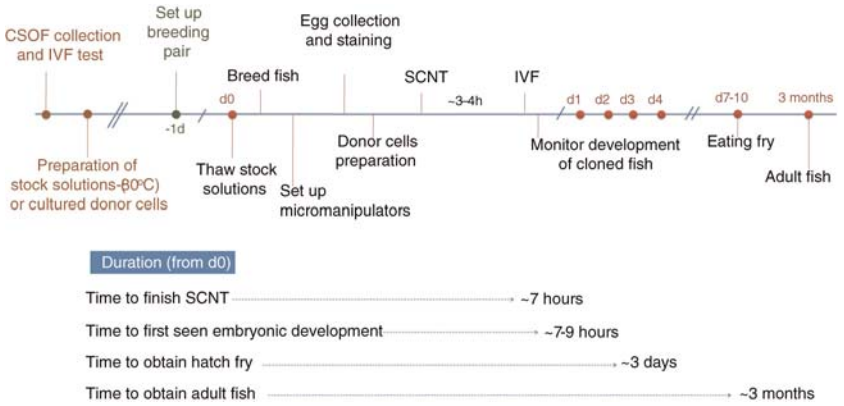


Fig. 1 Timeline for somatic cell nuclear transfer in zebrafish. (See color plate.)

were retrieved from little Manistee weir around fall–winter transition period (mid-September to October). The CSOF is collected and tested as following.

1. Preparation for a Field Work

We used zebrafish breeding apparatus (Rough filter units and reservoirs, with 1 mm^2 stainless grid) to separate the eggs from the ovarian fluid. Other strainer that could allow separation of salmon eggs and ovarian fluid might be used in conjunction with any kind of reservoir. Paper towels are used to clean the genital area prior to egg collection to avoid contamination of water. Other miscellaneous items should be prepared including the following: conical tubes (50 mL), permanent markers, ice boxes and ice, boots and aprons, and plastic bottles with caps (500 mL) for pooling fluid of several females.

2. Collection of Ovarian Fluid

Salmon is anesthetized with tricaine (MS222) and the abdominal area is carefully dried out prior to egg taking to avoid water contamination. Positive air pressure is then in the abdominal cavity of the salmon to release eggs and ovarian fluid into a plastic can. During this time, care should be taken to avoid contamination of blood and feces since they can lower the quality of ovarian fluid. The ovarian fluid is separated from salmon eggs using a strainer. The quality of CSOF is assessed grossly: good quality fluid should be clear to off-whitish in color, and may contain few cell debris without broken eggs or blood contamination. In case of fecal contamination, the fluid turns greenish in color. A pathological examination, if applicable, should be done on site to avoid collection of CSOF from sick fish. CSOF should be placed on ice soon after collection. The fluid is stored and tested either individually – each batch from a single female — or as a pool of several females.

3. Storage

CSOF is aliquot into 50 mL conical tubes, and spun down at 8000 rpm for 10 min in a refrigerated centrifuge (4 °C). If floating sediment is observed, it is removed using a pasture pipette. The fluid is gently poured into a collecting bottle without reaching the sediment part. We recommend setting aside aliquots of 12 mL from each batch for quality testing prior to use. The batch number and collection date should be labeled properly and the CSOF stored in vacuum-sealed bags in a −80 °C freezer.

4. Quality Control

Aliquots of CSOF are thawed and their quality is tested using *in vitro* fertilization (IVF). The standard protocol for IVF has been described in the zebrafish book (Westerfield, 1993). Prior to use, CSOF is filtered through 0.22 μm membrane. Eggs are collected from at least three females and placed directly into CSOF. Approximately 30 eggs are fertilized hourly with freshly prepared milt, and for up to 6 h post egg collection. The developmental rate is recorded at blastula, day 1 post fertilization (pf), and hatch fry, with indicative numbers of the abnormal embryos obtained. Good quality CSOF should maintain a fertilization rate in the first 3 h post-collection as high as 80%. So that the recipient eggs could be maintained at MII throughout the SCNT operation.

5. Preparation of Stock Solution

Batches of CSOF that have passed the quality control described in 4 are thawed in water at room temperature, occasionally shaking and changing the water to minimize thawing time. CSOF is filtered through 0.22 μM (largest) top filter unit. Note that the fluid is very thick and slowly passed through the filter. The tested and filtered CSOF are fractioned into aliquots of 12 mL in plastic tubes with caps, labeled, and stored in −80 °C freezer until used for SCNT. A small portion of CSOF from a same batch is used to prepare: 5% PVP in CSOF (aliquots of 1 mL each) and kept in an −80 °C freezer. We prefer to use a same batch of CSOF in each manipulation to minimize confounding factors.

B. Somatic Cell Nuclear Transfer (SCNT)

The following is a protocol that circumvents the challenge of using MII eggs as recipient cells for SCNT. Our technique, including laser-assisted inactivation of the egg genome and the delivery of the donor cell through the egg's micropyle, allows for a tight control of the egg's cell cycle stage.

1. Breeding Pairs

The day before SCNT manipulation, breeding pairs are prepared in a breeding tank. In general, we prepared only one pair of fish per breeding tank and set up at

least ten pairs. As soon as the light comes, the breeding pairs are allowed to start their courtship behavior. In general, the male would strip the female multiple times over the course of one breeding session. The breeding pairs are separated once the female spawns the first few eggs. The naturally fertilized embryos are collected for quality control and monitored for cleavage.

2. Preparation: Reagents and Micromanipulators

Frozen aliquots of CSOF, 5% PVP in CSOF, and MS222, are thawed out in a water bath at room temperature. All reagents are prepared in Falcon dishes as shown in Fig. 2.

Micromanipulators are set up as shown in Fig. 3; an egg holder with a CellTram® Air (Eppendorf) is connected to one side, and on the other side, an injection needle and a supporting needle with CellTram® Oil (Eppendorf) are also set. It is important that the injection needle and the supporting needle are aligned in parallel. As long as this type of needle setting is allowed, other models of micromanipulators may be used as well.

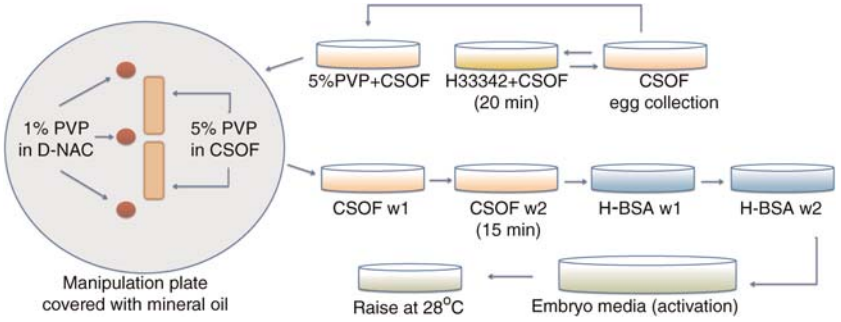


Fig. 2 Reagents for zebrafish SCNT. (See color plate.)

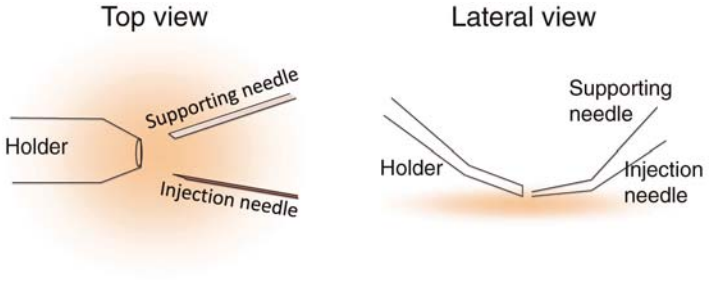


Fig. 3 Micromanipulators. (See color plate.)

3. Recipient Eggs

We evaluate the quality of natural fertilized (NBD) embryos prior to egg collection using cleavage rate and spontaneous activation. A clear sign of proper fertilization is established in the presence of embryos that developed timely and with a well-demarcated cytoplasm. Good egg activation is judged by the display of a chorion that is fully and uniformly separated from the egg. We only stripped eggs from females that gave good quality embryos. Egg isolation is done as previously described with slight modifications (Westerfield, 1993). Briefly, the female is anesthetized with MS222, rinsed in fish water, and placed on Kimwipe® slit supported by the operators' thumb and index fingers (Fig. 4). The abdominal area is fully dried with Kimwipe®. We gently applied pressure on lateral abdomen with two fingers, and then using a glass rod the eggs are gradually stripped out by applying gentle pressure from the anterior part of abdomen toward the genital opening. Eggs are released directly into CSOF. The biopsy of the caudal fin of the female egg donor is also collected at the same time for subsequent genotypic analysis.

The morphology of the MII eggs is assessed for their quality under a stereoscope. Good eggs are granular and yellowish in color. It is not recommended to use eggs with mixed quality, i.e., containing lots of cell debris, or some showing signs of spontaneous activation. Eggs are stained with Hoechst33342 (50 μg/mL in CSOF) in dark-moist chamber for exactly 20 min. Hoechst33342 should be carefully mixed with CSOF as precipitation may affect DNA staining. Stained eggs are moved back to CSOF until used as recipient eggs for SCNT.

4. Donor Cells

Donor cells are prepared just prior to their use. The protocol varies depending upon the source of the donor cells. For freshly isolated cells taken from embryos, we separate tail buds out from a 15–24 hpf old larvae, and mince them briefly in LHC media. Embryos in LHC are mixed using a ratio of 1:1 with 0.05% trypsin

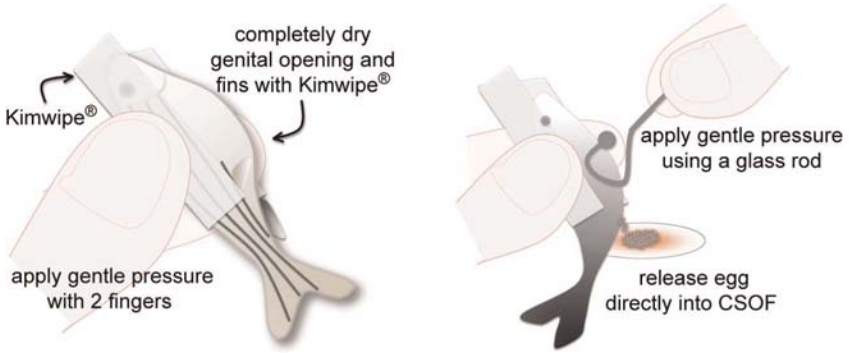


Fig. 4 Egg collection.

EDTA. Trypsinization is done for 10 min at room temperature. An equal volume of 5% FBS is added in LHC to inhibit trypsinization. Cells are centrifuged at 1200 rpm for 5 min, and the supernatant is discarded. Cells are washed with 5% FBS in LHC, and collected after another round of centrifugation. The cell pellet is briefly disassociated by gently pipetting and then mixed in 1% PVP in D-NAC without serum. For adherent cell cultures, cells are washed twice with LHC, and trypsinized using 0.025% Trypsin EDTA in LHC. The 5% FBS is added in LHC to inhibit enzymatic reaction, and then the same washing protocol described above is followed.

5. Micromanipulation

For SCNT manipulation, stained eggs are washed in 5% PVP in CSOF, and placed in a manipulation drop. A laser XY Clone 40× objective lens is set at 100% power with a 500 μS pulse. The egg's metaphase plate is located using a micropyle as a landmark, under 4× objective lens. The egg then rotated and aligned to have its animal pole facing the bottom of the dish. The egg is then gently sucked with the holding pipette. We used both the supporting needle and the injection needle to help in fixing the position of the egg, and to press it closer to the bottom of a dish. The fluorescence-stained metaphase plate of the egg is located first under a 20× objective lens and UV light. Then using the XY Clone 40× objective lens, the metaphase plate is aligned with the laser target and a double firing is aimed at the metaphase plate. This procedure is repeated for the remaining eggs (generally five eggs are manipulated at once). Note that both the viewing of the metaphase plate and the ablation of it are more intense – more effective – when closest to the bottom of the manipulation dish.

Donor cells are placed in a separate drop in the same manipulation dish and allowed to settle at the bottom. It is important to avoid placing of many cells in a drop, as later cell selection could be difficult. Donor cells are individually selected and placed in the injection needle (under 20× objective lens). The needle used for injection is a traditional ICSI needle normally used for human IVF, but with a small internal diameter so that the cytoplasmic membrane brakes at suction. Note that the size of the injection needle varies depending upon the type of cells used. The donor cell, with disrupt cytoplasmic membrane, is moved to a manipulation drop. The enucleated egg is realigned to have its micropyle facing the injection needle, and gentle pressure is applied through a holder to fix the egg in place (under 4× objective lens). It is important to align the micropylar pit in parallel with the injection needle to facilitate the cell transfer. The micropyle has the shape of a cone, with a small pit at the end, so we must use an injection needle with bevel tip to guide it through the pit. In our experience, an injection needle as large as 10 μm in diameter can pass through the micropyle. If the tip of injection needle is not located at the micropylar pit, it is impossible to pass through easily and could injure the embryo. We recommended using an injection needle to glide along a micropylar cone toward the deepest part, where the micropylar pit is located. A single nucleus is then transferred through the micropyle. We generally pick one

cell at a time. A total of five eggs are manipulated in a given round, and processed for the next step.

6. Post-micromanipulation

The reconstructed embryos are washed once in CSOF (CSOF wash #1) to remove PVP, and then incubated in CSOF (CSOF wash #2) for 15 min. Prior to activation, the embryos are washed twice in H-BSA. We used embryo medium to activate the reconstructed embryos at room temperature, and then culture the embryos at 28 °C. We record development at blastula (3 h post nuclear transfer – hpNT), sphere (4 hpNT), germ ring (6 hpNT), 90% epiboly (9 hpNT), 1 day, 2 days, 3 days (hatch), 4 days, and continued every day until the larvae starts eating and reaching adulthood.

7. Quality Control IVF

After the last SCNT group is finished, IVF is performed in the eggs that were not used. The milt is freshly prepared from the male of the same breeding pair, and kept in HBSS (10 μL). The concentration and motility of sperms are evaluated under a microscope prior to using it. IVF is performed as previously described with slight modification (Westerfield, 1993). Briefly, eggs are washed twice in H-BSA, and then moved to a new culture dish (60 mm^2 Falcon) with small amount of H-BSA. Pre-evaluated milt is added in conjunction with embryo media (approximately 500 μL to 1 mL), mixed, and left untouched for at least 3 min. Fertilized eggs are then moved to a culture dish with 5 mL of embryo media and cultured at 28 °C.

II. Materials

1. CSOF—pre-tested using IVF is stored in aliquots at −80 °C.
2. CSOF with 5% Polyvinylpyrrolidone (Sigma-Aldrich, St. Louis, MO) can be dissolved at 4 °C overnight, filtered, and stored at −80 °C.
3. Hank's balance solution (HBSS)with Ca^{2+} and Mg^{2+} (Gibco, Carlsbad, CA).
4. 50 μg/mL Hoechst 33342—bisBenzimidetrihydrochloride (Sigma-Aldrich, St. Louis, MO) in CSOF for egg's DNA staining, 1 μL of Hoechst stock solution (50 mg/mL in HBSS, kept at −20 °C) is added to 1 mL of CSOF prior to use.
5. 0.5% Bovine serum albumin (Sigma-Aldrich, St. Louis, MO) in HBSS—HBSA, stored at −80 °C.
6. LHC basal media (Gibco, Carlsbad, CA).
7. 0.05% Trypsin EDTA (Gibco, Carlsbad, CA).
8. 5% Fetal bovine serum—FBS (Hyclone, Logan, UT) in LHC basal media.
9. 1% PVP in D-NAC—DMEM(Gibco, Carlsbad, CA), supplemented with 2mM N-acetyl-L-cysteine (Sigma-Aldrich, St. Louis, MO), 1 μM ascorbate-2-phosphate (Sigma-Aldrich, St. Louis, MO), and 10 ng/mL bovine insulin (Sigma-Aldrich, St. Louis, MO) for a drop of donor cells, stored at 4 °C for up to 1 month.

10. Plastic petridishes (BD Falcon™); 35 mm^2, 60 mm^2, and 100 mm^2.
11. Fire polished glass pipettes with approximately 1 mm diameter, and screw pipettor.
12. Glass pipette for micromanipulators; holder (200–300 μm, inner diameter–ID with fire polished), supporting needle (20 μm, ID, just cut straight), and intracytoplasmic sperm injection—ICSI needle (select size according to a type of donor cells, Humagen, Charlottesville, VA).
13. Embryo medium (Westerfield, 1993).
14. Thermo-plate and incubator set at 28.5 °C.
15. Fluorescence inverted microscope with UV filter and micromanipulation set.
16. Laser-assisted enucleation, 40× objective lens – XYClone module (Halminton Throne Biosciences, Beverly, MA).

References

Amores, A., and Postlethwait, J. H. (1999). Banded chromosomes and the zebrafish karyotype. *Methods Cell Biol.* **60**, 323–338.

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., and Hopkins, N. (2004). Identification of 315 genes essential for early zebrafish development. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12792–12797.

Balciunas, D., Davidson, A. E., Sivasubbu, S., Hermanson, S. B., Welle, Z., and Ekker, S. C. (2004). Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics* **5**, 62.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702–708.

Grunwald, D. J., and Streisinger, G. (1992). Induction of recessive lethal and specific locus mutations in the zebrafish with ethyl nitrosourea. *Genet Res.* **59**, 103–116.

Lee, K. Y., Huang, H., Ju, B., Yang, Z., and Lin, S. (2002). Cloned zebrafish by nuclear transfer from long-term-cultured cells. *Nat. Biotechnol.* **20**, 795–799.

Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D., and Wolfe, S. A. (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701.

Nagayoshi, S., Hayashi, E., Abe, G., Osato, N., Asakawa, K., Urasaki, A., Horikawa, K., Ikeo, K., Takeda, H., and Kawakami, K. (2008). Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: tcf7 and synembryn-like. *Development* **135**, 159–169.

Nasevicius, A., and Ekker, S. C. (2000). Effective targeted gene "knockdown" in zebrafish. *Nat. Genet.* **26**, 216–220.

Nüsslein-Volhard, C., and Dahm, R. (2002). *Zebrafish: A Practical Approach.* Oxford University Press, Oxford.

Siripattarapravat, K., Busta, A., Steibel, J. P., and Cibelli, J. (2009a). Characterization and *in vitro* control of MPF activity in zebrafish eggs. *Zebrafish* **6**, 97–105.

Siripattarapravat, K., Pinmee, B., Venta, P. J., Chang, C. C., and Cibelli, J. B. (2009b). Somatic cell nuclear transfer in zebrafish. *Nat. Methods* **6**, 733–735.

Solnica-Krezel, L., Schier, A. F., and Driever, W. (1994). Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**, 1401–1420.

Westerfield, M. (1993). *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio).* University of Oregon Press, Eugene, OR.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., and Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813.

PART 3

The Zebrafish Genome and Mapping Technologies

CHAPTER 13

Single Nucleotide Polymorphism (SNP) Panels for Rapid Positional Cloning in Zebrafish

Matthew D. Clark[*], Victor Guryev[†], Ewart de Bruijn[†], Isaac J. Nijman[†], Masazumi Tada[‡], Catherine Wilson[§], Panos Deloukas[*], John H. Postlethwait[§], Edwin Cuppen[†] and Derek L. Stemple[*]

[*]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[†]Cancer Genomics Center, Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

[‡]Department of Cell and Developmental Biology, University College London, Gower Street, London, UK

[§]Institute of Neuroscience, 1254 University of Oregon, Eugene, OR, USA

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00013-6

Abstract

Despite considerable genetic and genomic resources the positional cloning of forward mutations remains a slow and manually intensive task, typically using gel based genotyping and sequential rounds of mapping. We have used the latest genetic resources and genotyping technologies to develop two commercially available SNP panels of thousands of markers that can be used to speed up positional cloning.

I. Introduction

Even with a dense map of 3842 simple sequence length polymorphisms (SSLPs) on the MGH meiotic panel (Shimoda *et al*., 1999), 4073 predominately coding markers on the doubled haploid mapping panel (Woods *et al*., 2005) and a near complete genome sequence (http://www.sanger.ac.uk/Projects/D_rerio/) the positional cloning of mutations can still be time consuming. SSLPs while highly polymorphic between strains do not lend themselves to a highly parallel approach, and only a small subset are useable as agarose scorable markers rather than on radioactive polyacrylamide gels (Geisler *et al*., 2007). The T51 Radiation Hybrid (RH) panel provides the densest and highest resolution map (Geisler *et al*., 1999). However, doubled haploid and RH panel markers are, however, based largely on coding sequences and as such are not as polymorphic as SSLP markers. Furthermore, in assembling the genome we have noted that the T51 panel, while accurate over short distances, disagrees with meiotic panels and the physical maps over longer distances. It is possible that the AB9 cell line used to make both RH panels (LN54 and T51) (Hukriede *et al*., 1999; Kwok *et al*., 1998) had undergone rearrangements. Indeed a large part of the improvements in genome assembly from Zv8 onwards were due to the implementation of a marker weighting scheme that took this into account, giving more weight to meiotic map placements than to RH map placements (see http://tinyurl.com/5trm9uz). Thus while there are significant marker resources, there is a continuing need for more markers, especially those that are polymorphic, meiotically mapped, and are amenable to high throughput genotyping, both for better genome assembly and to accelerate the cloning of mutants.

Well over 650,000 candidate SNPs have been mined from sequence traces from the Washington University Zebrafish EST project (Clark *et al*., 2001) and the Wellcome Trust Sanger Institute's zebrafish genome sequence, with confirmed polymorphism rates of 65–86% (Bradley *et al*., 2007; Guryev *et al*., 2006). In addition to providing many more markers, high throughput SNP genotyping technologies with over one million markers typed in parallel (Frazer *et al*., 2007) (see also http://tinyurl.com/66knryp) are now widely available. Most researchers have access to these genotyping platforms through core facilities, or commercial services. A previous Zebrafish SNP microarray contained only 599 SNPs (optimized for the C32 and SJD strains) at 234 unique map positions with a mean sex averaged distance between markers of 9.8 cM (Stickney *et al*., 2002) but required many PCR amplifications before pooling and SNP detection, and was not widely adopted.

We have constructed a SNP panel where all 3212 SNPs are assayed simultaneously, using molecular inversion probes (MIPs) technology (Hardenbol *et al*., 2005). A single experiment can be used to screen all 3212 SNPs at once for coarse mapping. In addition we have designed and validated a subset of 1072 SNPs, which can be used individually as PCR based assays using an allele specific amplification method known as KASPar (Cuppen, 2007). Both panels are commercially available from Affymetrix (MIPs) and KBiosciences (KASPar), respectively. Data files for the SNP assays described here are available from the Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv).

Using these panels we have characterized the genotypes and genetic relationships of nine common zebrafish strains, have more than doubled the number of confirmed zebrafish SNPs, and added 971 markers to the Heat Shock genetic map (Kelly *et al*., 2000; Woods *et al*., 2005). Finally we have also used these methods to map a mutation with the commonly used bulk segregant approach. We believe that with this commercially available SNP panel, a single researcher could roughly (5-10cM resolution) map up to 96 mutations a week and identify SNPs polymorphic in their cross for finer mapping on individuals, for example by KASPar.

II. Extraction of Zebrafish Genomic DNA

MIPs assays use more concentrated DNA than standard genotyping PCRs, thus any contaminants would also be more concentrated. For MIPs, we use a phenol chloroform isoamylalcohol (PCI) extraction step to make cleaner DNA. For KASPar, since less DNA is needed per reaction, we have used standard proteinase K digestions, heat killing the proteinase and diluting before use, though we obtain better results after precipitation and re-suspension. Both protocols can easily be carried out in 96-well microtiter plates.

A. Adult Fin Clips

This protocol scales well in our hands and has better yields than kits we have tried. Using a 96-tip pipetting robot (e.g., Agilent Bravo or Robbins Hydra) allows easy removal of the supernatant from above the phenol phase. This can also be achieved by multi-channel pipette. Starting at an outer row, and using bevelled tips, judge by eye (through the plate wall) the level of the phenol, and move the tips ~2 mm above this to pipette the supernatant – now note the position of the plate edge relative to the tips, and use the same tip height for rows where you cannot view the tips and meniscus through the plate wall.

Briefly, caudal fin clips are collected into 2 ml Corning 96 Well Clear V-Bottom 2 ml Polypropylene Blocks (Corning Costar Catalogue # 3961) containing 400 μl of ProtK Buffer: (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.2% SDS, 5 mM EDTA pH 8.0 and 100 μg/ml proteinase K). After heat-sealing the plates (Agilent PlateLoc Thermal Plate Sealer), fins are digested overnight at 55 °C floating in a large water bath. The proteinase K is heat inactivated by 30 min at 80 °C in the water bath. The

plate is cooled to room temperature, and 400 μl of PCI is added to each well, the plate is re-sealed and mixed by inversion. After 10 min at 4000 rpm, 300 μl of the supernatant is added to a new Corning plate already containing 300 μl of isopropanol. Plates are re-sealed, mixed by inversion and precipitated for 30 min at 4000 rpm. The pellets are washed and spun twice for 10 min at 4000 rpm with 500 μl of 70% ethanol, the wash solution is removed by inversion, and draining on tissue paper (in these Corning plates the pellets stick well) finally the pellet is air dried including a final 15 min at 70 °C in a hot air oven to remove the last traces of ethanol, and dissolve overnight at 4 °C in 20 μl of T0.1E (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0), heat-sealed, and heated for 10 min in a 70 °C oven. DNA concentrations must be calculated using Picogreen fluorescence (Invitrogen) against standards using a Qubit (Invitrogen) or in a fluorescent plate reader (e.g., Pherastar from BMG labtech). A spectrophotometer can give inaccurate readings due to organic solvents, RNA, and protein contaminants. We routinely obtain 1–5 μg depending on fin size. The concentration is adjusted accordingly to the Picogreen readings to a final of 75 ng/μl, either by hand, or results can be fed into a pipetting robot.

B. Zebrafish Larvae

Assemble single embryos in a 96-well-rigid skirted PCR plate, for example an ABgene Thermo-fast 96 skirted (Catalogue # AB-0800), in which each well is prefilled with 100 μl of 100% methanol and stored at -20 °C. Remove the methanol when ready to prepare DNAs. Dry embryos completely on heating block (until they jump in their wells). Add 25 μl 2.0 mg/ml proteinase K in TE (10 mM Tris-HCl pH 8.0 and 1 mM EDTA pH 8.0). Digest at least 4 h at 55 °C. Heat inactivate for 30 min at 75 °C. DNA yields vary especially depending on developmental stage; for example, one 3-day fry gives about 100–150 ng DNA. For KASPar assays, 5–50 ng will be sufficient (approximately 4 μl of a 4x to 20x dilution in ddH2O). If assays are new or problematic, then DNAs can be cleaned by precipitation in plates by the addition of 2.5 μl 3 M sodium acetate pH 5.2, and 20 μl isopropanol, mixed by pipetting, precipitated for 30 min at 4 k rpm, washed with 100 μl of 70% ethanol, wash solution is drained off by inversion and traces removed by heating in a PCR block for 2 min at 70 °C, the pellet is dissolved in 10 μl of T0.1E (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA pH 8.0 to a final concentration of ~10 ng/μl) and heated to 70 °C for 2 min.

For MIPs, 1 μg is needed for each bulk segregant pool. Take half the DNA from each fry after digestion, assemble mutant and sibling pools of 12–48 fry, extract the pools with an equal volume of PCI (Phase Lock Gel Tubes from Eppendorf or MaXtract tubes from Qiagen make this easier) and then precipitate with 0.6 volumes of isopropanol and 0.1 volumes of sodium acetate pH 5.2, wash pellet with 1 ml 70% ethanol, air dried and re-dissolve in 20 μl T0.1E (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and heat to 70 °C for 2 min. Take 1 μl to quantify using Picogreen assay, and adjust the concentration to 75 ng/μl.

III. Affymetrics Molecular Inversion Probe Panel

The MIPs technology is a development of padlock probes (Nilsson *et al*., 1994) that allows up to 10,000 SNPs to be assayed simultaneously (Hardenbol *et al*., 2005) and is an attractive platform for the rapid design and testing of new SNP panels. In MIPs, a long probe with both ends targeting a SNP's flanking sequence and containing a unique 25 bp "barcode" is manufactured, these can then be pooled together (up to 10,000 probes) and hybridized to the same genomic sample overnight in a PCR machine. After hybridization, the 5' and 3' ends of each long oligonucleotide are designed to match the genomic sequence flanking a SNP such that the SNP base is a 1 bp gap between the ends of the oligonucleotide. At this point, the hybridized samples are split into four (all subsequent steps from hereon are carried out on these split reactions until ready to be hybridized to the Affymetrix Array), and four single base polymerase extension reactions each with just one of the four bases A, C, G, and T is carried out. In a correctly extended probe, the two ends of the probe (still hybridized on the genomic DNA) now abut, and a ligation reaction generates a circular single stranded DNA molecule. At this point all non-circular probes are destroyed by exonuclease digestion. Within the circularized probes there are two universal primer sites which point towards each other (this is the molecular inversion as before circularization they pointed away from each other and towards the ends of the DNA molecules), small amplicons (also containing the unique 25 bp barcodes) can now be PCR amplified only from circularized probes. Four amplifications are carried out on the split reactions, which were extended with the four separate nucleotides, and each amplicon is labeled with a primer specific to that base. The reactions can then be fragmented, pooled, and hybridized to the Affymetrix Universal Tag Arrays, to which the barcodes hybridize. The barcodes identify the SNP assays, and single base extensions and subsequent four-color labeling identifies the base(s) added at that position.

The MIPs technology is highly specific because of the controlled hybridization, extension by a high fidelity polymerase, and by the specificity of the ligase. More details on the MIPs technology, and the use of MIPs in genotyping project are available here: (Absalan and Ronaghi, 2007). In addition, the design requirements for the MIPs probes is very similar to that for the allele specific amplification Kaspar assay, which also uses a single base extension and PCR and maximizes the likelihood that the multiplexed MIPs assays will work as single assays across many individuals for fine mapping.

A. SNP selection

At the time of selection, there were two previous publications on Zebrafish SNPs, both identifying SNPs in gene transcripts (Guryev *et al*., 2006; Stickney *et al*., 2002). Stickney *et al*. characterized SNPs in genes already mapped by single stranded conformation polymorphism (SSCP) on the basis of polymorphisms between C32 and SJD strains. The Guryev *et al*. data set contains over 50,000 predicted SNPs by comparing EST traces from the Washington University Zebrafish EST project

(Clark *et al*., 2001), generally taken from known strain cDNA libraries (e.g., SJD, C32, and AB), to genomic sequence traces from the Sanger Zebrafish genome project (Tübingen strain). More recently a set of SNPs has been defined by comparing whole genome shotgun reads to finished clones from the zebrafish genome sequencing project (Bradley *et al*., 2007). Since the genome project is based on Tübingen strain this should largely identify intra-Tübingen SNPs and the majority will be intergenic SNPs.

In our SNP selection, we prioritized experimentally confirmed SNPs over predicted SNPs (confirmed SNPs are those that are experimentally polymorphic in at least one comparison of the AB, C32, TL, Tu, and WIK strains). We also required that SNPs could not overlap or be within 50 bp of an already placed SNP, finally we attempted to generate an even spread across the genome.

The selection consists of four SNP subsets:

1. Confirmed SNPs from Guryev *et al*. (190 SNPs).
2. Confirmed SNPs from Stickney *et al*. (66 SNPs).
3. High-quality predicted SNPs from Guryev *et al*, with each allele confirmed by two reads with a Phred quality score (Ewing and Green, 1998; Ewing *et al*., 1998) of at least 20 (1245 SNPs).
4. High-quality predicted SNP from Guryev *et al*., where one of the alleles was confirmed with only one read, that has a Phred quality over 30 (6837 SNPs).

We used all of subsets 1, 2, and 3 (1501 SNPs), and then we sampled through subset 4 to fill up to a total of 3501 and to get an even distribution of SNPs throughout the zebrafish linkage groups, filling as many gaps above 158 kbp as possible (Fig. 1). We added another set of SNPs, 20 ENU-induced STOP codons from the Sanger Zebrafish Mutant Resource project, which we reasoned would not occur in wild type populations. We submitted 3521 SNPs with flanking sequences to Affymetrix for assay design and 3212 were designed and manufactured by Affymetrix.

Based on the genotyping data from the MIPs panel a 1078 SNPs subset was selected for KASPar assays. These SNPs were selected based on performance in the MIPs assay across a selection of diverse strains (call rates differ across strains, especially when more divergent from the Tübingen reference genome). The criteria were high call rate in MIPs (>90%), a high reproducibility rate (>95%), a minor allele frequency (MAF) of greater than 0.05, and were called and polymorphic between at least two of the commonly used strains (AB, Tübingen LongFin, Hubrecht LongFin and Wik). Assays were designed and tested by KBiosciences, genotyping four fish for each of the four strains with duplicate reactions.

B. MIPs Protocol

We carry out MIPs assays according to the manufacturer's instructions except we use 1 μg of DNA instead of the suggested 2 μg. Since the zebrafish genome is about

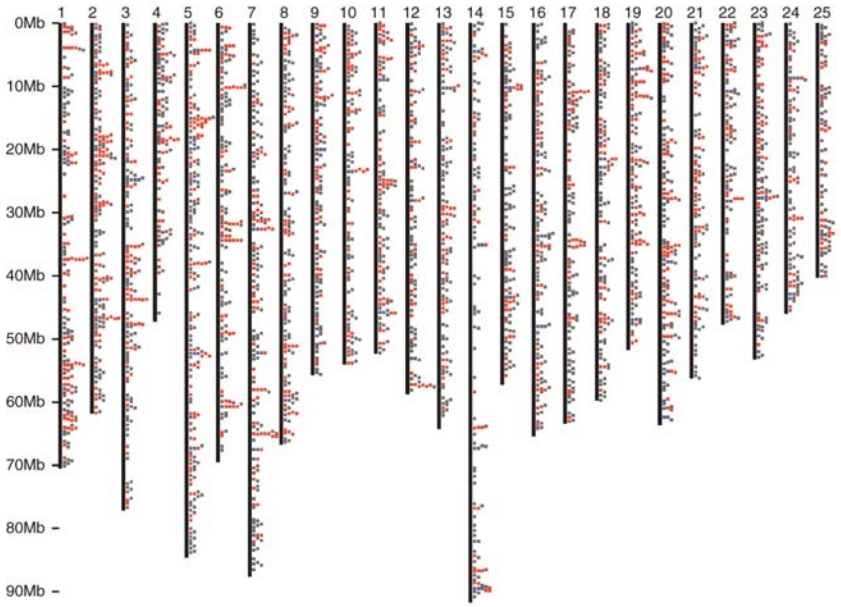


Fig. 1 SNPs in the MIPs panel displayed in relation to zebrafish chromosomes. Experimentally confirmed SNPs from Set 1 (Stickney *et al.*, 2002) and Set 2 (Guryev *et al.*, 2006) are blue, high-quality predicted SNPs from Set 3 are red, and lower-quality predicted SNPs from Set 4 are gray. Original colour pictures are available from Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv). (See color plate.)

half the size of human the concentration of any single copy locus will be the same. A control DNA from a single Hubrecht Long Fin male was included in each plate.

For analysis, we use the standard Affymetrix Targeted Genotyping software with default settings. For quality control:

(1) The coefficient of variation of control elements must be less than 30%.
(2) The median ratio of allelic signals to non-allelic signal must be greater than 20. For each assay, there are only two pre-defined bases. This can cause problems in tri- or quad- allelic systems but these are rare.
(3) Greater than 80% of the 3212 SNPs must be called on a sample for it to be included.
(4) Less than 10% of SNPs should be marginally called (the genotype appears between homozygous and heterozygous, see below for more details on calling genotypes). Pooled samples often fail this, but can be manually passed if they pass the other criteria.

After QC and normalization calls are assigned to homozygous and heterozygous genotypes, for example AA, AB, and BB calls, based on the ratio of signals for each

barcode (SNP assay) in each of the four colors (bases). Only the two known alleles (bases) designated during design are considered. A confident homozygous call has >80% of the signal in one channel, and a confident heterozygous call has >90% of all the signal in the two channels and at a ratio of close to 1:1. Many assays cannot be confidently assigned and are marked as marginal. Such marginal calls are then clustered together with the confident calls for each SNP assay, and marginal calls are assigned to the close genotypes using the allele ratio to measure distance. Because of this clustering poor experiments (or pooled experiments) if manually passed increase the misassignment of calls. Thus, it is worth deleting poor experiments, and keeping a copy of a core set of high quality experiments across wild type strains with which to cluster your experiments.

Note: This system requires the use of an Affymetrix four-color GeneChip Scanner 3000 7G, which is less common than the standard one-color models. Check with your genotyping core or service provider.

C. Strain Polymorphism

Our first tests were on the common strains. We fin clipped, prepared DNA, and genotyped 12 fish of each of the common strains using our MIPs assay: Tübingen, Tübingen LongFin, AB, and Wik as well as some less common inbred strains SJD and C32, and the G0 founders of the MGH cross (Knapik *et al*., 1996). The result of the genotyping of an inbred (SJD) and outbred line (Tübingen) as seen by Affymetrix's supplied software is shown in Fig. 2.

Using the 2221 SNPs that passed the quality control criteria described above, we discretized the calls into A (Homozygous for A allele), B (Homozygous for B allele), and C (Heterozygous). This makes the data anonymous with respect to the allele type, and the nucleotide types. The tree in Fig. 3 was generated using this data and fpars, which is part of the Phylip package http://evolution.genetics.washington.edu/phylip.html (Felsenstein, 1989). Due to the results of these experiments, we split two different strains that we originally thought were both the Tübingen Long Fin strain into two groups, Tübingen Long Fin and Hubrecht Long Fin. The Hubrecht Long Fin strain was used to generate ENU libraries for Zebrafish Mutation Resource and was originally obtained from a Hubrecht pet shop.

D. Genetic Mapping

We had already tested the G0s of the two main meiotic panels MGH and Heat Shock. We decided to genotype the other 42 members of the heat shock panel. On this cross 1072 SNPs appeared to be polymorphic, of which 997 could be assigned a map position with high confidence when merged with the extant markers (Kelly *et al*., 2000; Woods *et al*., 2005). These markers were added to the data used in the genome assembly from Zv8 onwards, and are present in ENSEMBL (Flicek *et al*., 2011) in the marker DAS track.

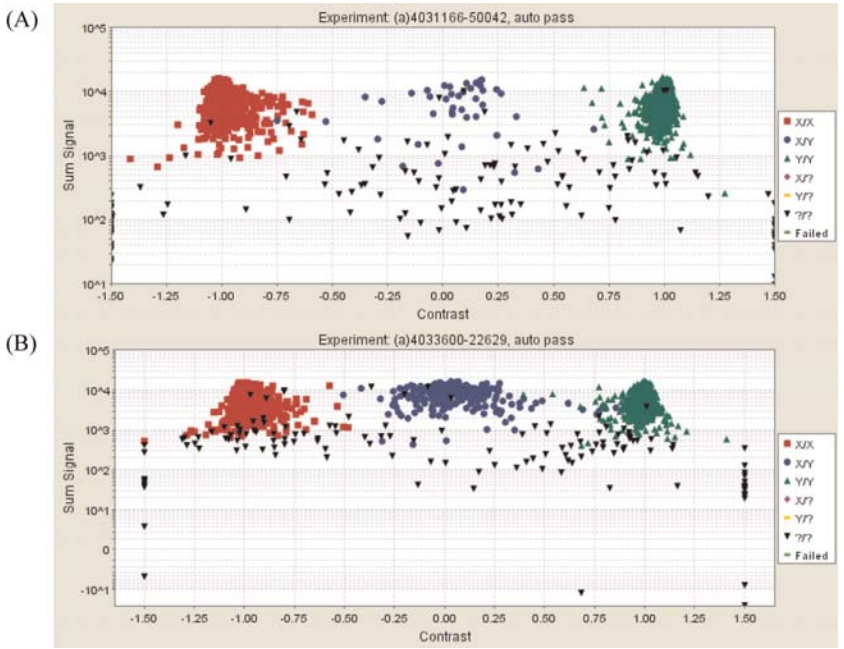


Fig. 2 Results of genotyping 3212 SNPs on individual fish. Each position is the result of one of the 3212 SNP assays. Red is homozygous XX allele, green is homozygous YY, blue is heterozygous (XY), and black are assays that could not be called. (A) The genotype of a single male of the SJD strain, a highly inbred line (1.3% heterozygous); (B) the genotype of a single male from the Tübingen strain, which is not an inbred line (8.3% heterozygous). Original colour pictures are available from Vertebrate Development and Genetics page at the Wellcome Trust Sanger Institute (http://tinyurl.com/3bsoolv). (See color plate.)

E. Pooling Tests

To be able to use bulk segregant mapping it is important to be able to identify alleles within a pooled population. To test this requirement, we picked rare induced ENU alleles from the Zebrafish Mutation Resource (TILLING) project, as these ENU induced alleles were seen only once in over 6000 sequenced fish. In serial twofold dilutions, we diluted these ENU allele carrying fish with the control DNA (a Hubrecht Long Fin male used as a control on each plate), and also with a more stringent test, a pool of 96 mixed fish (24 fish of four strains: Tübingen, Tübingen Long Fin, AB, and Wik). The pool is a particularly demanding test since a second SNP within the primer sequences for the assay can lead to poor annealing and even subsequent allele bias when genotyping the target SNP.

Figure 4 shows the SNP genotyping of the induced Gli3^{hu2965} allele in wild type, heterozygous, and homozygous fish. Figure 5 shows the same SNP, increasingly diluted by the wild type allele: from homozygous, to heterozygous, and heterozygote diluted with an equal amount of control DNA (C) or pooled wild type strains (P). The

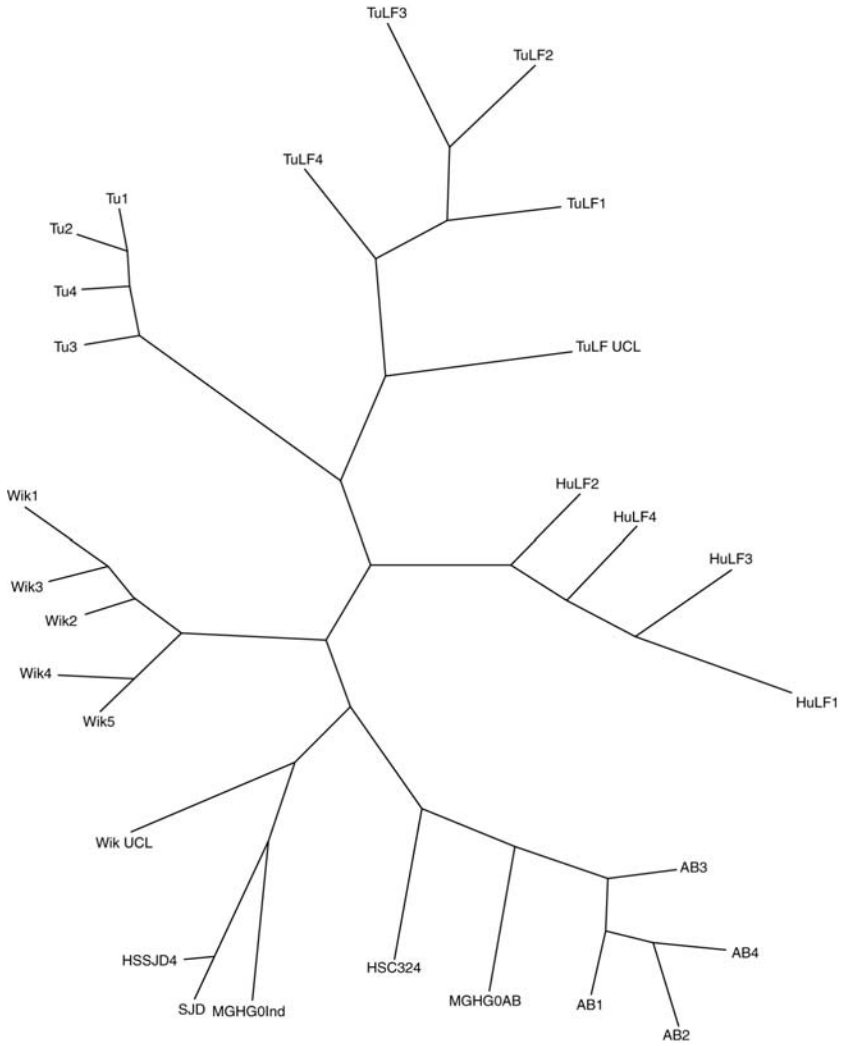


Fig. 3 Phylogenetic tree of strains. Genotypes were discretized to calls of A, B, and C, and used fpars, part of the Phylip package. Tu is Tübingen, TuLF is Tübingen LongFin, TuLF UCL is a Tübingen LongFin individual from the University College of London, HuLF is Hubrecht LongFin, AB is AB strain, MGHG0AB is the AB strain G0 individual used to make the MGH mapping panel, HSC324 is the C32 strain (inbred isolate of AB) G0 individual used to make the Heat Shock mapping panel, MGHG0Ind is the India strain G0 individual used to make the MGH mapping panel, SJD is Steve Johnson Darjeeling (inbred Indian wild type), HSSJD4 is the SJD strain G0 individual used to make the Heat Shock mapping panel, Wik1 to Wik4 are individuals of the Wik strain, and Wik UCL is a Wik individual from the University College of London. All fish are from the Sanger facility and originally sourced from ZIRC unless otherwise stated. The individuals from the mapping panels correspond to the strain as it was at that time, for example 10–15 years ago.

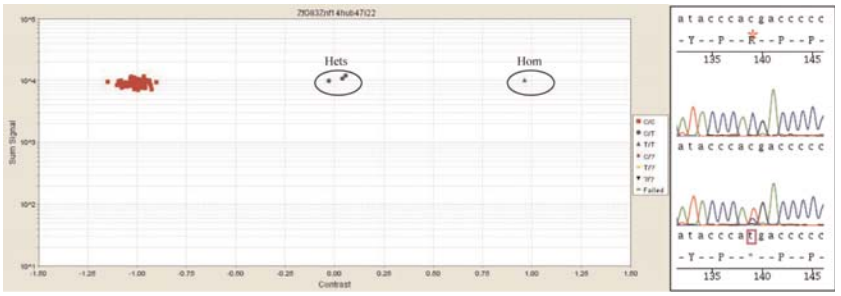


Fig. 4 MIPs assay of homozygous wild-type individuals (C/C: red squares), three heterozygous carriers of a nonsense mutation (C/T: blue circles), and one homozygous mutant individual (T/T: green triangle). The X-axis indicates the spectral contrast between the two dyes of the assay, and the Y-axis is the log of the fluorescence signal. To the right is shown capillary sequencing traces indicating the induced mutation in allele *gli3*^{hu2965} (lower) in comparison with wild type (upper). The MIPs homologous flanking sequence is**: gli3_hu2965 Zv9_24_11311698** atggtcctgaagcccatgtcaccaaaaaacaacgtggggagacatacccca [**C/T**] gacccccacctcagccacgagaacctggagcaaccactgaccaaaaggag. (See color plate.)

full range allele dilutions are: 1/1 (undiluted mutant), 1/2 (heterozygote), 1/4 (heterozygote diluted with equal amount of wild type DNA), 1/8, 1/16, 1/32, 1/64, and 1/128. The 1/8 dilution is scored as containing the ENU allele with both control and pooled dilutions, and the 1/16 dilutions still appear separate from the Wild type samples. Since bulk segregant mapping normally involves comparing the genotypes of mutants (homozygotes) to siblings (a mix of heterozygotes and wild types) a 1/16 sensitivity is more than is needed for to map a monogenic Mendelian recessive

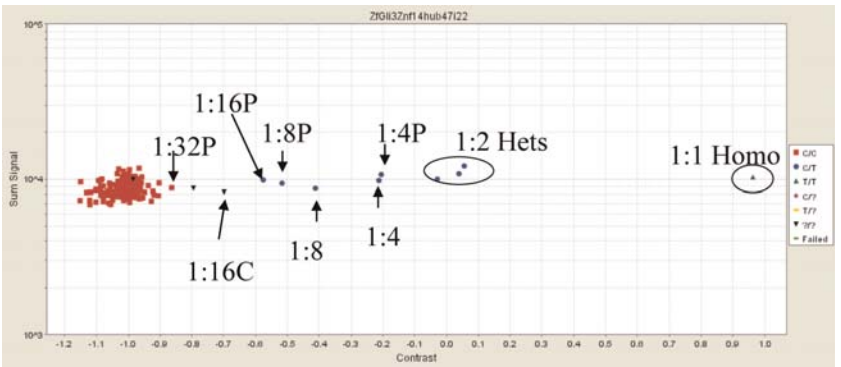


Fig. 5 The *gli3* MIPs was tested by combining DNA from a single wild-type fish (e.g., 1:16C) with heterozygous *gli3*^{hu2965} DNA at different ratios or by combining the pooled DNA of 100 adults (e.g. 1:16P) with *gli3*^{hu2965} DNA different ratios. Also shown is the signal from many wild-type (C/C: red squares) fish, as well as three heterozygous *gli3*^{hu2965} carriers (C/T: blue circles) and one homozygous mutant individual (T/T: green triangle). The X-axis indicates the spectral contrast between the two dyes of the assay, and the Y-axis is the log of the fluorescence signal. (See color plate.)

mutant. It should be possible to screen pools for the presence of a given allele, for example, a specific induced allele.

F. Bulk Segregant Mapping

Next we mapped a mutation found to enhance the *silberblick* phenotype (Heisenberg *et al.*, 2000). The allele was isolated on a TuLF background carrying *silberblick* and crossed onto Wik for mapping. We tested the grandparents, parents of the cross, and ools of mutants and siblings: 6, 12, and 24 embryos deep. In addition, we tested a further improvement using low DNA input. In this case, we amplified 10 ng of DNA from single embryos using the GenomiPhi V2 kits from GE Life Sciences according to the manufacturer's protocol: the amplifications were cleaned with Qiagen Qiaquick columns and Picogreen quantified using 1 μg inputs for MIP assays.

The results were promising: we saw a good signal that was around the *siberblick* locus, and a very strong signal of four SNPs showing linkage on LG21 that on Zv9 (http://www.ensembl.org/Danio_rerio) spans less than a 2.6 Mb region. The signal was strong in pools (Table I) and in individuals (Table II). In a separate experiment we also use GenomiPhi amplification of embryos that had been fixed and hybridized with *in situ* probes, to generate sufficient DNA for a MIPs assay 1 μg (data not shown).

Table I
Bulk segregant mapping

SNP	Zv9 pos	G1	G2	P1	P2	24M	12M	6M	24S	12S	6S
ss49818686	21:2913135	BB	AA	AB	AB	BB	BB	BB	AB	AB	AB
ss49817351	21:4563832	AA	BB	AB	AA	AA	AA	AA	AB	AB	AB
ss49835121	21:3548874	AB	BB	AB	AB	AB	AA	AA	AB	AB	AB
ss49818143	21:2002074	AB	BB	AB	AB	AA	AA	AA	AB	AB	AB

There are four SNPs that show strong linkage with the *silberblick* enhancer; the SNPs are within a 2.6-Mb region of LG21 in Zv9 (between bases 2,002,074 and 4,563,832). G1 and G2 are grandparents – TLF carrier and WIK map outcross, respectively, P1 and P2 are parents, and 24M, 12M, and 6M are mutant pools of 24, 12, and 6 fry, respectively, whereas 24S, 12S, and 6S are the equivalent sibling pools. The pools of 24, 12, and 6 mutants are homozygous (with the exception of ss49835121 in 24M a likely genotyping error), whereas the pools of siblings are all heterozygous.

Table II
Single embryo mapping

SNP	Mut1	Mut2	Mut3	Mut4	Mut5	Mut6	Sib1	Sib2	Sib4	Sib5	Sib6
ss49818686	BB	BB	BB	BB	BB	BB	AB	AB	AA	AB	AB
ss49817351	AA	AA	AA	AA	AA	AA	AB	BB	BB	AB	AB
ss49835121	AA	AA	AA	AA	AA	AA	AB	AB	BB	AB	AB
ss49818143	AA	AA	AA	AA	AA	AA	BB	AB	BB	AB	BB

The same four SNPs as in Table I, on DNA from single larvae amplified using GenomiPhi V2. There are six mutants and five Siblings (one failed to pass QC filters). All mutants are homozygous; no sibling embryos are homozygous for the same allele as a mutant.

Interestingly, on further investigation and cDNA sequencing it was found this *silberbick* enhancer is a mutation in *hmgcrb,* which was isolated previously (D'Amico *et al.*, 2007).

IV. Suggested Mapping Strategy

For a complete mapping strategy, we suggest a two-step approach, rough mapping followed by fine mapping. For rough mapping, we suggest the use of the 3212 SNP MIPs array and bulk segregant pools ideally no more than 24 fry deep, in this way just two chips should establish linkage for about $2–300. Since four sets of 48 samples can be processed in the MIPs protocol per person per week 96 mutants per week could be roughly mapped in this way. This compares well with the time and total cost for the SSLP approach (Geisler *et al*., 2007). For even greater confidence one could genotype two mutant and two sibling pools, or add the parents. However, linkage should be obtained using just two pools (one mutant and one phenotypically wild type) and can be quickly tested by fine mapping. This may not be true in some cases, especially if closely related strains or carrier fish of an admixture of strains are used. If linkage fails to be established, additional pools or parental carriers could be subsequently genotyped to help.

For fine mapping, candidate SNPs can be assayed using the KASPar panel across large numbers of mutants, for example 372 and a few siblings, for example 12. The candidate SNPs should clearly be polymorphic on the MIPs, and preferably in the extant KASPar panel, though it is relatively inexpensive to order new assays.

V. General Considerations

A. Genetic Diversity

In the design of any mapping strategy the correct choice of strains is critical in maximizing the polymorphic marker rate and the number of individuals that need to be genotyped. Essentially a mapping strain distant from the mutant line should be chosen. We have genotyped the commonest stains, several individuals from each strain, often from independent labs, and established the alleles within strains and polymorphisms between them. This is largely summarized by the tree in Fig. 3, which allows the quick identification of distant strains. If an admixture line is used then genotyping the grandparents and parents helps identify the SNP alleles origins and makes subsequent analysis easier.

B. Usefulness of the SNP Panel

After initial rough mapping using the MIPs panel, the emphasis switches from many markers on few pools, to a few closely linked markers on many individuals.

With a large number of extant cloned and well characterized mutants, many laboratories are performing suppressor or enhancer screens. A combined screening and cloning strategy has already been described (Rawls *et al*., 2003) combined with a quick and cost effective method of high-resolution genotyping, such screens become even more attractive.

More recently RAD-tag marker microarrays containing 13,824 elements have been used to map mutations (Miller *et al.*, 2007); however, the sequence, genetic, and genomic positions of these elements is unknown: in addition such RAD markers assayed by microarrays are dominant and non-quantitative and therefore heterozygosity can only be seen after sequencing the products and uncovering additional SNPs. Combining RAD-tags and Illumina sequencing, that is RAD-seq makes this more attractive (Baird *et al.*, 2008), especially for organisms without many existing markers or finished genomes, but still requires considerable bioinformatics analysis.

With the increasing numbers of predicted SNPs, the ability to identify new SNPs using RAD-seq, and other sequencing-based experiments, SNP panels will become increasingly used for mapping zebrafish mutations and QTLs.

References

Absalan, F., and Ronaghi, M. (2007). Molecular inversion probe assay. *Methods Mol. Biol.* **396**, 315–330.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Bradley, K. M., Elmore, J. B., Breyer, J. P., Yaspan, B. L., Jessen, J. R., Knapik, E. W., Smith, J. R. (2007). A major zebrafish polymorphism resource for genetic mapping. *Genome Biol.* **8**, R55.

Clark, M. D., Hennig, S., Herwig, R., Clifton, S. W., Marra, M. A., Lehrach, H., Johnson, S. L., Group t, W. (2001). An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Res.* **11**, 1594–1602.

Cuppen, E. (2007). Genotyping by allele-specific amplification (KASPar). *CSH Protoc. 2007* pdb prot4841.

D'Amico, L., Scott, I. C., Jungblut, B., and Stainier, D. Y. (2007). A mutation in zebrafish hmgcr1b reveals a role for isoprenoids in vertebrate heart-tube formation. *Curr. Biol.* **17**, 252–259.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., *et al*. (2011). Ensembl 2011. *Nucleic Acids Res.* **39**, D800–D806.

Frazer, K. A., Eskin, E., Kang, H. M., Bogue, M. A., Hinds, D. A., Beilharz, E. J., Gupta, R. V., Montgomery, J., Morenzoni, M. M., Nilsen, G. B., *et al*. (2007). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053.

Geisler, R., Rauch, G. J., Baier, H., van Bebber, F., Bross, L., Dekens, M. P., Finger, K., Fricke, C., Gates, M. A., Geiger, H., *et al*. (1999). A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23**, 86–89.

Geisler, R., Rauch, G. J., Geiger-Rudolph, S., Albrecht, A., van Bebber, F., Berger, A., Busch-Nentwich, E., Dahm, R., Dekens, M. P., Dooley, C., *et al*. (2007). Large-scale mapping of mutations affecting zebrafish development. *BMC Genomics* **8**, 11.

Guryev, V., Koudijs, M. J., Berezikov, E., Johnson, S. L., Plasterk, R. H., van Eeden, F. J., Cuppen, E. (2006). Genetic variation in the zebrafish. *Genome Res.* **16**, 491–497.

Hardenbol, P., Yu, F., Belmont, J., Mackenzie, J., Bruckner, C., Brundage, T., Boudreau, A., Chow, S., Eberle, J., Erbilgin, A., *et al*. (2005). Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* **15**, 269–275.

Heisenberg, C. P., Tada, M., Rauch, G. J., Saude, L., Concha, M. L., Geisler, R., Stemple, D. L., Smith, J. C., Wilson, S. W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81.

Hukriede, N. A., Joly, L., Tsang, M., Miles, J., Tellis, P., Epstein, J. A., Barbazuk, W. B., Li, F. N., Paw, B., Postlethwait, J. H., *et al*. (1999). Radiation hybrid mapping of the zebrafish genome. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9745–9750.

Kelly, P. D., Chu, F., Woods, I. G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y. L., Zhou, Y., *et al*. (2000). Genetic linkage mapping of zebrafish genes and ESTs. *Genome Res.* **10**, 558–567.

Knapik, E. W., Goodman, A., Atkinson, O. S., Roberts, C. T., Shiozawa, M., Sim, C. U., Weksler-Zangen, S., Trolliet, M. R., Futrell, C., Innes, B. A., *et al*. (1996). A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. *Development* **123**, 451–460.

Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, R., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N., Schmitt, K. (1998). Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. *Nucleic Acids Res.* **26**, 3562–3566.

Miller, M. R., Atwood, T. S., Eames, B. F., Eberhart, J. K., Yan, Y. L., Postlethwait, J. H., Johnson, E. A. (2007). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol.* **8**, R105.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B. P., Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088.

Rawls, J. F., Frieda, M. R., McAdow, A. R., Gross, J. P., Clayton, C. M., Heyen, C. K., Johnson, S. L. (2003). Coupled mutagenesis screens and genetic mapping in zebrafish. *Genetics* **163**, 997–1009.

Shimoda, N., Knapik, E. W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., Fishman, M. C. (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**, 219–232.

Stickney, H. L., Schmutz, J., Woods, I. G., Holtzer, C. C., Dickson, M. C., Kelly, P. D., Myers, R. M., Talbot, W. S. (2002). Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Res.* **12**, 1929–1934.

Woods, I. G., Wilson, C., Friedlander, B., Chang, P., Reyes, D. K., Nix, R., Kelly, P. D., Chu, F., Postlethwait, J. H., Talbot, W. S. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* **15**, 1307–1314.

CHAPTER 14

Molecular Cytogenetic Methodologies and a BAC Probe Panel Resource for Genomic Analyses in the Zebrafish

Kimberly P. Dobrinski,[*,†] Kim H. Brown,[*,†] Jennifer L. Freeman[‡] and Charles Lee[*,†]

[*]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, USA

[†]Harvard Medical School, Boston, Massachusetts, USA

[‡]School of Health Sciences, Purdue University, West Lafayette, Indiana, USA

Abstract

Molecular cytogenetics is a field that emerged in the 1980s, based on a technique referred to as fluorescence *in situ* hybridization, (FISH). Using FISH methodologies, a specific DNA sequence or collection of DNA fragments may be selectively labeled

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00014-8

with a hapten molecule or fluorescent dye and hybridized to denatured chromosomes, interphase cells, or even chromatin fibers. DNA hybridization kinetics permit these labeled probes to anneal to their complementary sequences on such chromosomal DNA preparations allowing for direct visualization of the sequence of interest in the genome being interrogated. If present, the relative chromosomal position of the sequence can sometimes also be ascertained.

Progress in molecular cytogenetic research has advanced the genetic characterization of zebrafish models of human diseases as well as assisted with accurate annotation of the zebrafish reference genome by anchoring large DNA fragments to specific chromosome regions. Using the procedures described in this chapter, hundreds of ambiguous zebrafish bacterial artificial chromosome (BAC) clones have already been assigned to individual genetic linkage groups . Molecular cytogenetic techniques can also be used to study gene duplication events and study the molecular mechanisms by which they arise . Moreover, the availability of a new molecular cytogenetic technique, array-based comparative genomic hybridization (aCGH), is now able to identify gains and losses of DNA segments in zebrafish DNA samples in a genome-wide manner and in a single assay.

I. Introduction

The zebrafish genome contains 1.35×10^9 bp of DNA, approximately half of the genome size of most mammals, and is organized into 25 pairs of chromosomes ($2n = 50$) (Genome Research Consortium, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/zebrafish/data/index.shtml). Classical cytogenetic studies of zebrafish have shown that individual chromosomes are difficult to unequivocally identify on the basis of chromosome size, morphology (i.e., position of the centromere), and banding pattern (Amores and Postlethwait, 1999). It is therefore not surprising that since 1968 at least 12 different zebrafish karyotypes have been published (Sola and Gornung, 2001). In part, due to the multiple karyotypes that have been published, attempts have been made to assign genetic linkage groups to specific chromosomes based on chromosome size and arm ratios. In one study (Phillips *et al*., 2006), relative chromosome size and linkage group-specific fluorescence *in situ* hybridization (FISH) markers were used to assign each chromosome to a specific linkage group using the short arm of chromosome 3 (3p) as an internal control for chromosome size. In another study, linkage group chromosomes were also identified by chromosome-specific FISH markers, and the size of each zebrafish chromosome was determined using flow cytometry (Freeman *et al*., 2007). While there was good concordance between these two studies for chromosomal assignment of cloned DNA fragments for the largest four chromosomes (i.e., linkage groups 3, 4, 5, and 7) and the smallest three chromosomes (i.e., linkage groups 22, 24, and 25), there was considerable discordance for the remaining 18 chromosomes. Some of the discrepancy in chromosome assignment is likely due to the similar sizes of many of the zebrafish chromosomes. Indeed, based on flow cytometry data, only

35.7 Mb of DNA separates the largest and smallest chromosomes. Among the 18 linkage groups with discordant chromosomal assignments, the size separation is greatly reduced with only 9.1 Mb separating these chromosomes (Freeman *et al*., 2007).

To provide optimized DNA probes for zebrafish genomic analyses, a second-generation zebrafish bacterial artificial chromosome (BAC) probe panel has now been developed (Freeman *et al*., 2007). The methods described below can be used to map specific BAC probes from this panel to chromosome preparations or cells to assess genomic imbalances at specific chromosome regions, cellular ploidy, and genomic instability.

In contrast to examining genomic variation at specifically targeted chromosomal loci, the technique of comparative genomic hybridization (CGH) permits the discovery of genomic imbalances across the genome without *a priori* knowledge of chromosome, chromosomal regions, or genes that may be aberrant. In classical chromosomal CGH, two genomes are differentially labeled, hybridized to metaphase cell preps, and ratios calculated between the labeled DNA samples for measurement of DNA changes in the genome (Kallioniemi *et al*., 1992). However, classical CGH is technically demanding, has low reproducibility, and exhibits a low resolution of around 5–10 Mb for single copy gains and losses (recently reviewed in Peterson and Freeman, 2009). Some of these challenges were overcome with the advent of array-based comparative genomic hybridization (aCGH). The technique of aCGH uses glass slide microarrays containing DNA sequences mapped to the genome and hybridized to differentially labeled DNAs. DNA from a test sample is co-hybridized with DNA from a reference individual to identify relative genomic gains and losses. The intensity ratio between the two different fluorophores at each probe on the microarray is calculated to measure relative copy number changes for specific genomic regions (Fig. 1).

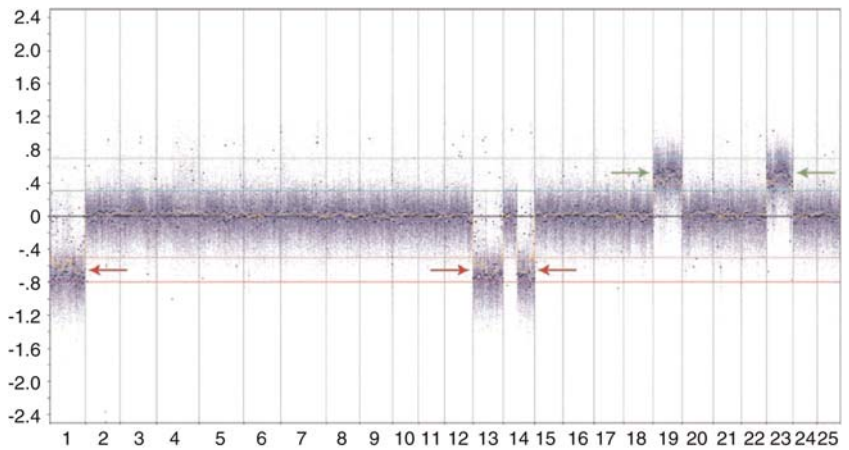


Fig. 1 Example of array comparative genomic hybridization (aCGH) on a zebrafish tumor cell compared to a normal reference sample. Chromosomes 1, 13, and 14 show chromosomal losses (red arrows) compared to the chosen reference DNA source, while chromosomes 19 and 23 show chromosomal gains (green arrows). *X*-axis = chromosome number, *Y*-axis = $\log_2$ ratios. (See color plate.)

Multiple aCGH platforms are now available for zebrafish, which facilitates their use in identifying genomic imbalances throughout the whole genome for disease studies as well normal inter-individual genetic variation. aCGH is technically less challenging than classical CGH and has increased the effective resolution for the detection of gains and losses in zebrafish genomes to chromosomal distances as small as 5 kb (see below).

II. Cytogenetic Methods

Note: These methods, which are updated periodically, can also be found on the following website: http://www.chromosome.bwh.harvard.edu.

A. Zebrafish Chromosome Preparations

1. Metaphase Chromosomes from Embryos

1. Take ~100 embryos at 23 h postfertilization (hpf).
2. Add colchicine (Acros Organics, Fair Lawn, NJ) to a final concentration of 4 mg/mL. Incubate at 28.5 °C for 6 h.
3. Dechorionate embryos using Pronase treatment (Westerfield, 2007). Dilute 1 mL of Pronase stock (50 mg/mL) in 30 mL of fish water (Westerfield, 2007). Add 20 mL Pronase for 10 min. Mix embryos and Pronase solution by use of a transfer pipet.
4. Rinse embryos three times in 20 mL of 0.48 mg/mL Instant Ocean Sea Salt (Instant Ocean, Cincinnati, OH) in fish water. *Hint:* Remember to keep embryos in water at all times.
5. Remove excess water and transfer embryos to a clean microfuge tube on ice. Remove all but 100 μL of fish water and homogenize (approximately 10 strokes) using a pellet pestle (Kontes Glass Company, Vineland, NJ).
6. Add 1 mL of ice-cold 0.9× PBS (Fisher Scientific, Pittsburgh, PA), 10% fetal bovine serum (Atlanta Biologicals, Lawrenceville, GA).
7. Using a transfer pipet, apply the homogenate to 100 μm Cell Trics filter (Partec, Swedesboro, NJ) and collect the filtrate in a clean 15 mL centrifuge tube.
8. Collect the filtrate from step 7, and apply 50 μM Cell Trics filters (Partec, Swedesboro, NJ). Collect the second filtrate in a clean 15 mL centrifuge tube.
9. Add 5 mL of ice-cold 0.9× PBS, 10% fetal bovine serum.
10. Centrifuge at 210 *g* for 10 min at 4 °C.
11. Discard the supernatant and gently disperse the pellet by flicking the tube.
12. Add 10 mL of hypotonic colchicine solution (1 mL 11.0% sodium citrate, 1 mL colchicine (40 mg/mL), and 8 mL fish water).
13. Incubate at room temperature for 25 min.
14. Add 1 mL of ice-cold fixative (3:1 methanol:glacial acetic acid).
15. Centrifuge at 400 *g* for 10 min at 4 °C.

16. Remove supernatant and add 10 mL of ice-cold fixative.
17. Centrifuge cells at 400 *g* for 10 min at 4 °C.
18. Repeat once more with fresh ice-cold fixative.
19. Store cell pellet at −20 °C until use or perform one more wash and drop slides (see Section A.3).

2. Metaphase Chromosomes from Established Adherent Cell Cultures

Note: Several established zebrafish cell lines are now available from the American Type Culture Collection (http://www.atcc.org) including ZF4, SJD.1, ZEM2S, ZFL, and AB.9. To ensure proper chromosome assignment, mapping should be carried out on early passage lines only, because later passages frequently contain cell culture artifacts (e.g., aneuploidy). Metaphase chromosomes can be prepared from these and other independently derived adherent cell lines as follows:

1. Grow an established zebrafish cell line to ∼80% confluency.
2. *Hint:* Depending upon the characteristics of a cell line, multiple flasks may need to be combined to attain cell quantities for optimal metaphase preparation.
3. Add 10 μL of colcemid (0.1 mg/mL) (Irvine Scientific, Santa Ana, CA) and incubate at 28.5 °C and 5% CO_2 for 6 h.
4. Add 1 mL of ethidium bromide (10 mg/mL) per milliliter of culture medium to increase chromosome length. Incubate for 45 min at 28.5 °C.

   Note: This step can be deleted if good metaphase spreads are achieved without addition.

5. Remove media and add 2 mL sterile 1× PBS. Let it sit for 2 min.
6. Remove PBS and add 2 mL of 0.25% Trypsin/0.53 mM EDTA (ATCC, Manassas, VA). Place in a 28 °C incubator for 5 min. *Hint:* If using ZF4 cells, trypsin CANNOT contain EDTA.
7. Add 8 mL of media and transfer the cells to a clean 15 mL centrifuge tube.
8. Centrifuge cells at 200 *g* for 5 min at room temperature.
9. Remove the supernatant and gently resuspend cell pellet in residual media.
10. Combine cells and residual media from each flask into a 15 mL centrifuge tube and bring to a total volume of 1.2 mL.
11. Add 6.8 mL of 0.067 M KCl. Let it sit at room temperature for 22 min.

    Note: Make KCl fresh daily.
    Note: Incubation duration depends on cell line characteristics.

12. Add 1 mL of fixative (3:1 methanol: glacial acetic acid) and mix thoroughly.
13. Spin at 400 *g* for 10 min at 4 °C.
14. Remove supernatant and resuspend pellet in residual solution. Add 7 mL of fresh fixative and let it sit on ice for 30 min. Centrifuge at 400 *g* for 10 min at 4 °C.
15. Remove supernatant and resuspend pellet in residual solution. Add 10 mL fixative and centrifuge at 400 *g* for 10 min at 4 °C.

16. Repeat Step 14, reducing fixative volume to 5 mL.
17. Store cell pellet at −20 °C until use or immediately drop slides (see Section A.3).

B. Preparation of Bacterial Artificial Chromosome (BAC) Probe DNA

1. Warm agar plates to room temperature. Streak BAC clones from glycerol stocks onto LB-agar plates with the appropriate antibiotic.
2. Grow overnight in a 37 °C incubator.
3. Pick a single colony and place in a 50 mL conical tube containing 12 mL of terrific broth (TB) plus 6 µL of the appropriate antibiotic.
4. Incubate in a shaking incubator at 2 *g* overnight at 37 °C.
5. Isolate DNA using a plasmid purification Midi Kit (Qiagen #12643, Valencia, CA)
    a. Chill the P1 (50 mM Tris-Cl, pH 8.0; 10 mM EDTA; 100 µL/mL RNAse A) and P3 (3 M potassium acetate, pH 5.5) buffers at 4 °C.
    b. Centrifuge cells at 1000 *g* for 15 min in a clean and sterile 15 mL centrifuge tube. Pour off the supernatant into a flask containing bleach.
    c. Resuspend pellet in 2 mL of ice-cold P1 buffer. Vortex until pellet is thoroughly resuspended.
    d. Add 2 mL of P2 buffer (200 mM NaOH; 1% sodium dodecyl sulfate (SDS)) down the side of the tube. Roll tube gently and allow cells to lyse for 2 min at room temperature.
    e. Add 2 mL of ice-cold P3 buffer. Invert tube gently.
    f. Keep cells on ice for 20 min.
    g. Centrifuge lysate for 30 min at 2800 *g* to remove cellular debris.
    h. Avoiding the top layer of white film, transfer supernatant to a clean 15 mL centrifuge tube.
    i. Add 5 mL of isopropanol and gently invert tube to precipitate DNA. Centrifuge for 30 min at 2800 *g*.
    j. Decant isopropanol and add 10 mL of cold 70% ethanol. Centrifuge for 10 min at 2800 *g*.
    k. Decant ethanol. Invert tube on paper towel to dry pellet or place in a 55 °C oven to dry rapidly.
    l. Resuspend pellet in 100 µL of nuclease-free water. Place in a 55 °C water bath for 15 min.
    m. Store at 4 °C until ready for Repli-G amplification.
6. BAC DNA amplification using a Repli-G Kit (Qiagen #150045, Valencia, CA)
    a. Add 2.2 µL TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and 0.3 µL BAC DNA (100 ng/µL) to a 0.2-mL PCR tube.
    b. Prepare fresh dilution A (4.5 mL dH_2O, 0.4 mL 5 M KOH, and 0.1 mL 0.5 M EDTA).
    c. Prepare denaturation solution. Dilute solution A (provided by kit) 1:8.
    d. Prepare neutralization solution. Dilute solution B (provided by kit) 1:10.

e. Add 2.5 μL of denaturation solution to each PCR tube and let it sit at room temperature for 3 min.
f. Add 5 μL of neutralization solution to each tube to stop reaction.
g. Prepare Reaction Master Mix on ice as in Table I.
h. Add 40 μL Reaction Master Mix to each tube. Immediately place the tubes in a thermal cycler for two incubations. The first incubation is for 16 h at 30 °C, the second for 5 min at 65 °C, after which samples should be held at 4 °C until retrieved.
i. Tubes can now be stored at −20 °C until ready to proceed.
j. The resulting mix will be very viscous. Add 50 μL of dH_2O to dilute and facilitate precipitation. Transfer to a sterile 1.5 mL microcentrifuge tube.
k. Add 10 μL 3 M sodium acetate to each tube.
l. Add 250 μL 100% ethanol to each tube and invert to mix.
m. Spin tube at full speed (16,000 *g*) in a microfuge for 10 min to pellet DNA.
n. Pour off supernatant and add 1 mL ice-cold 70% ethanol.
o. Centrifuge at full speed in microfuge for 10 min.
p. Decant supernatant and air-dry pellet by leaving tube inverted for 5–10 min.
q. Resuspend pellet in 100 μL sterile ddH_2O. Incubate at 65 °C for 1 h, then overnight at 4 °C.
r. Store at 4 °C until ready to label for FISH.

C. Fluorescence *In situ* Hybridization (FISH)

1. Preparation of C_0t1 DNA

1. Take freshly anesthetized zebrafish and place in mortar. Thoroughly freeze the zebrafish in liquid N_2.
2. Using a pestle, grind the animal to a fine pulp, intermittently adding liquid N_2 to maintain the tissue as a frozen slurry. Transfer the pulp into a sterile 50 mL centrifuge tube and swirl to allow the remaining liquid N_2 to evaporate.

   Note: Tissue needs to remain frozen at all times to attain high molecular weight DNA.

Table I
Reaction master mix volumes for Repli-G BAC DNA amplification

	Volume (μL) per reaction	Volume (μL) per 100 reactions
Sterile water	27	297
4 × Repli-G buffer	12.5	137.5
DNA polymerase	0.5	5.5
Total volume	40	440

3. Add 20 mL of Proteinase K lysis solution (50 mM Tris-Cl, pH 8; 100 mM EDTA; 100 mM NaCl; 1% SDS; 100 μg/mL Proteinase K (Roche, Indianapolis, IN) to the tube. Shake to resuspend pellet. Gently rock the solution in a 55 °C incubator for 24 h.
4. Immediately prior to use, centrifuge Phase Lock Gel (PLG) light 50-mL tubes by centrifuging for 2 min at 1,500 *g* (5 Prime Inc, Gaithersburg, MD).
5. Transfer zebrafish pulp in lysis solution into the PLG tubes. Add 10 mL buffered phenol (phenol-Tris saturated, pH = 8.0, Roche, Indianapolis, IN). Gently invert tubes 20 times to mix the suspension and denature protein.

   Note: Buffered phenol from the above source routinely gave us the best quality DNA.

6. Add 10 mL chloroform/isoamyl alcohol (24:1), and mix thoroughly as in Step 5.
7. Spin at 1500 *g* for 5 min at room temperature to separate the phases.
8. Transfer the upper aqueous phase to a new 50 mL PLG tube. Repeat Steps 5–7.
9. Transfer the upper aqueous phase to a new sterile 50 mL centrifuge tube. Add 0.1× volume of 3 M sodium acetate, mix thoroughly, and then add 1× volume of isopropanol. Gently mix the solution until the genomic DNA starts to aggregate.
10. Let it sit at room temperature for 20 min.
11. To pellet the DNA, centrifuge at 800 *g* for 10 min at 4 °C. Carefully pour off the supernatant.
12. Wash the DNA pellet with 20 mL of ice-cold 70% ethanol. Centrifuge again as in Step 11.
13. Decant supernatant without dislodging the pellet. Using residual liquid, transfer the DNA pellet into a 2 mL tube. Invert tubes at room temperature to evaporate the residual ethanol (1–2 h).
14. Rehydrate the DNA pellet by resuspension in 1 mL of sterile TE (10 mM Tris-HCl and0.1 mM EDTA) and incubation overnight at 55 °C. *Note:* Mixing during incubation will promote efficient rehydration of DNA pellet.
15. Store at 4 °C until ready to proceed further.
16. Obtain DNA concentration by standard spectrophotometric measurements. Following DNA quantification, calculate the volume necessary to make 500 μL of total genomic DNA at a final concentration of 10 mg/mL.
17. Shear 500 μL genomic DNA (10 mg/mL) in a sonicator (Sonics Materials Inc., Newton, CT) to an average fragment size of 400 bp (10 min, amplitude 100, 10 s bursts). Check fragment size on a 1% agarose gel.
18. Denature the DNA by incubation at 100 °C for 15 min.
19. Calculate the time required for the C_0t1 fraction of the DNA sample to reanneal. *Note:* Use the equation: [5.92/DNA concentration in mg/mL] to determine the time in minutes needed for C_0t1 fraction to reanneal.
20. Place the DNA in a 65 °C water bath for 4 min.

21. Add 0.25 mL of 1 M NaCl to the 0.5 mL of DNA to yield a final concentration of 0.3 M NaCl. *Note:* Total volume should equal 0.75 mL.
22. Allow DNA to reanneal at 65 °C for the interval calculated in Step 17.
23. Add 1× volume of ice-cold 2× S1 nuclease buffer (Fisher Scientific, Pittsburgh, PA). Add 1 unit of S1 nuclease (Fisher Scientific, Pittsburgh, PA) per microgram of genomic DNA.
24. Incubate at 37 °C for 30 min.
25. Centrifuge the genomic DNA solution in a 15-mL PLG light tube for 2 min at 1500 *g*. Transfer supernatant to a 15 mL PLG light tube.
26. Add 10 mL phenol:chloroform:isoamyl alcohol (25:24:1), mix thoroughly, and spin at 1500 *g* for 5 min at room temperature to separate phases.
27. Set centrifuge temperature to 4 °C.
28. Transfer upper aqueous phase to a new sterile 15 mL centrifuge tube. Add 0.1× volume of 3 M sodium acetate mix thoroughly, then add 1× volume of isopropanol. Gently mix solution to precipitate DNA. Centrifuge at 3000 *g* for 10 min at 4 °C.
29. Remove supernatant and allow pellet to air dry for 1 h with tube inverted at angle.
30. Rehydrate pellet in 0.5 mL of sterile ddH_2O at 55 °C for 2 h. Store at −20 °C until ready to use.

2. Probe Synthesis

1. Using the Nick Translation Reagent Kit (Abbott Molecular Inc., Abbott Park, IL), mix the following in a microfuge tube:
   - 1 μg of probe DNA. Using sterile water, bring volume to 17.5 μL.
   - 5 μL of 10× nick translation buffer
   - 5 μL of 0.1 mM dTTP
   - 10 μL of dNTP mixture (0.1 mM dATP, dGTP, dCTP each)
   - 2.5 μL of 0.2 mM Spectrum Red/Orange dUTP or 3.5 μL of 0.2 mM Spectrum Green dUTP (Abbott Molecular Probes, Abbott Park, IL).
   - 10 μL of nick translation enzyme
2. Mix and incubate for 11 h at 15 °C and hold at −5 °C. *Note:* It is best to do this step in a thermocycler for optimal temperature control.
3. Electrophorese a 5 μL aliquot on a 1% agarose gel to check the probe size. The ideal probe should contain a spectrum of fragments <500 bp.
4. Purify the labeled DNA using a Sephadex-G50 column (Zymo Research, Orange, CA). *Note*: Follow the Zymo purification protocol (included).
5. Add 25 μg of zebrafish (Section C1 above) or herring sperm C_0t1 DNA (Promega Inc., Madison)
6. Add 0.5 μL of 0.5 M EDTA to preserve sample.
7. Store in −20 °C until needed for hybridization.

3. Making Zebrafish Metaphase Chromosome Preparations

Note: Metaphase slides can be prepared from either the uncultured or cultured cell method described above in *A*.

1. Boil water in a 100-mL beaker.
2. If making chromosome preparations from a stored cell pellet: Wash 2× in ice-cold fixative. (i.e., centrifuge at 450 *g* for 10 min at 4 °C. Remove supernatant and add 10 mL fresh ice-cold fixative.)
3. Make a humidified chamber by wetting a paper towel and putting a box over it.
4. Exhale on slide to warm and moisten slide.
5. Apply three drops of cell suspension along the length of the glass slide.
6. Place slide under the humidified chamber for about 1–2 min (while preparing the next slide).
7. Place slide over boiling water so that the steam spreads the chromosomes.
8. Wipe off the back of the slide with a paper towel.
9. Place slide on a slide warmer.
10. Using a phase contrast microscope, examine each slide to check for mitotic index and extent of chromosome spreading.

4. Pretreatment of Slides and Probe Hybridization

1. Place slides in 3:1 methanol:glacial acetic acid at −20 °C for 15 min.
2. Wash slides in sterile 1× PBS for 2 min. Repeat once more.
3. Place slides in a Coplin jar containing 60 mL Protease II solution (Abbott Molecular Probes,Abbott Park, IL) for 5 min at 37 °C.
4. Wash slides in sterile 1× PBS for 5 min. Repeat.
5. Dehydrate chromosome preparations by placing slides consecutively in 70%, 90%, and 100% ethanol (2 min each), then air dry. Use slides immediately.
6. While slides are drying, take 10 μL of probe mixture and centrifuge in a Speedvac (Thermo Scientific, Rockford, IL). Add 10 μL of Hybrizol (Oncor, Gaithersburg, MD) to the dried probe. Vortex briefly to resuspend pellet.
7. Apply 10 μL of the probe mixture to the dehydrated slides.
8. Cover with a 22 × 22 mm glass coverslip and seal edges with rubber cement.
9. Co-denature in a Hybrite (Abbott Molecular, Des Plaines, Il) at 70 °C for 3 min and then bring to 37 °C.
10. Once slides are at 37 °C in the Hybrite, transfer slides to a humidified chamber, in the dark, for 48 h at 37 °C.

5. Post-hybridization Washes and Counterstain

1. Gently peel away rubber cement and remove coverslip.
2. Perform two 5-min washes in 50% form amide plus 2× SSC at 45 °C in Coplin jars.
3. Perform two 5-min washes in 2× SSC at 45 °C in Coplin jars.

4. Wash slides in 4× SSC, 0.05% Tween-20 (Sigma-Aldrich, St. Louis, MO) for 8 min at 37 °C.
5. Apply 30 μL of antifade mounting medium with DAPI (Vector Labs, Burlingame, CA) and cover with a 24 × 60 mm coverslip.
6. Incubate in the dark at room temperature for 5 min.
7. Observe hybridized chromosome preparations by fluorescence microscopy with appropriate filter sets.

Note: Multiple colors can be used in a hybridization to detect different DNA fragments (e.g., DAPI (blue), Spectrum Green, Spectrum Orange, Spectrum Red, and Spectrum Aqua can be used along with appropriate excitation and emission filters.)

III. Application of Cytogenetic Methods in Zebrafish Studies

The development and application of molecular cytogenetic probes has bridged the longstanding gap between the DNA sequenced-physical maps and the more well-established genetic maps, providing cytogenetic anchors to facilitate the annotation of the most recent zebrafish reference genomes, Zv8 and Zv9, respectively. A second-generation zebrafish-chromosome-specific BAC probe panel was subsequently developed (Table II), which consisted of FISH-verified BAC clones that are closer to the centromere and telomeres of each zebrafish chromosome (Freeman *et al*., 2007) compared to the first-generation probe panel (Lee and Smith, 2004). BAC clones were selected from the CHORI 211 (Children's Hospital Oakland Research Institute BAC-PAC Resources Center, 2004, http://bacpac.chori.org) or Danio Key BAC libraries (Deutsche Ressourcenzentrum für Genomforschung, RZPD, 2004), initially on the basis of their proximity to the ends and centromere of each chromosome, inferred using linkage map data. Each BAC clone was end-sequenced for clone identity verification and subsequently mapped by FISH to metaphase chromosomes from wild-type zebrafish embryos as described by Freeman *et al*. (2007). For each FISH experiment, two different BAC clones for a given linkage group were simultaneously hybridized to metaphase chromosome spreads, thus providing visual confirmation of chromosomal synteny of the hybridized BAC clones (e.g., Fig. 2A and B). BAC clones mapping near the primary constriction (centromere) of each chromosome were designated as the near-centromeric clones, and near-telomeric probes were assigned to either the short arm or long arm termini of the respective chromosome. Distinct improvements in the second-generation BAC probe panel include 31chromosomal probes that are more closely located to their respective centromeres and telomeres and the inclusion of a BAC clone (zC079A18) from the long arm terminus of chromosome 4, which replaces the heterochromatin probe (zC027E19) previously used (Fig. 2C).

LG chromosome 4 corresponds to a zebrafish chromosome previously reported to contain substantial amounts of constitutive heterochromatin along most of the long arm (Sola and Gornung, 2001). Constitutive heterochromatin is typically comprised

Table II
Second-generation probe panel consisting of near-telomeric and near-centromeric BAC clones for each zebrafish chromosome (bold and italicized clones were in the first-generation probe panel; Lee and Smith, 2004)

LG	Short arm telomere	Long arm telomere	Centromere
1	***zC093G23***	***zC141F18***	***zC022O06***
2	***zK014G06***	***zC009D09***	zK127K09
3	***zK007C07***	zK030G05	***zC115J06***
4	***zK030C13***	zC079A18	zC091G03
5	***zC087E10***	***zC150K20***	***zK007B18***
6	zK023D07	zK166J19	zC112J16
7	***zK009M06***	***zC128L16***	***zK014N10***
8	***zC069A12***	zK149F22	***zC103G04***
9	***zC115B08***	***zC012N08***	zC212N06
10	***zC128P08***	***zC022E09***	zC136O04
11	zC159E12	***zC115I06***	zK014H17
12	***zC121C04***	***zC086E02***	***zK022H21***
13	***zK006L12***	zC113I09	zK016I06
14	***zC117N19***	***zC125N22***	***zC117E17***
15	***zC055C01***	***zC059M05***	zK151P21
16	zC119D19	zK246M23	zC213B07
17	***zK013L17***	***zK014B13***	zC042B22
18	cZ241L24	***zK014D24***	***zK005J13***
19	zK089B17	zK201G07	***zC132A16***
20	zK077B17	zC153J24	zK033I22
21	***zC122A16***	***zK014M09***	***zC065O02***
22	***zK002J07***	zC118M01	zC206A19
23	***zC041B11***	zC214H13	***zC051C19***
24	zK018N12	zK124I03	***zK001A04***
25	***zC096F02***	***zC087L10***	zK044K01

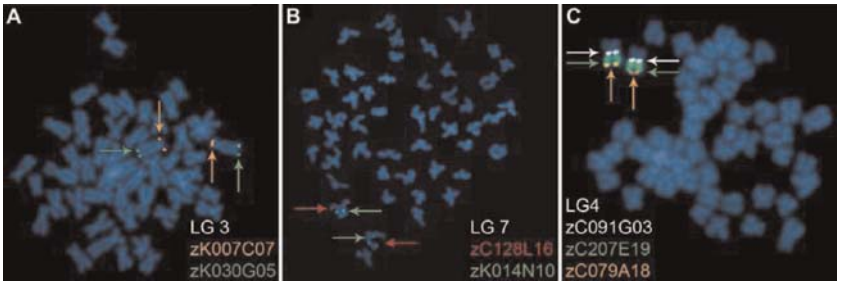


Fig. 2 Representative FISH images from the second-generation BAC probe panel. Panels A and B present two color FISH images: A) near-telomeric short-arm probe (*orange*) and near-telomeric long-arm probe (*green*) of LG chromosome 3; B) near-centromeric probe (*green*) and near-telomeric long-arm probe (*red*) of LG chromosome 7. Panel C compares the use of the new near-telomeric long-arm probe (*orange*) and the previously used heterochromatin probe (*green*), as well as a near-centromeric probe (*white*) for LG chromosome 4. (See color plate.)

of highly repetitive DNA (John, 1988), such as 5S rDNA, which is known to be present on this chromosome (Gornung *et al*., 2000; Phillips and Reed, 2000). Previous BAC clones that localized to the long arm of LG chromosome 4 revealed a hybridization pattern similar to that observed for 5S rDNA, with hybridization observed along most of the long arm. By contrast, BAC clone zC079A18 specifically labels the telomeric region of chromosome 4.

In humans, most centromeres contain repetitive sequences that are chromosome-specific. These DNA sequences are ideal because they provide prominent hybridization signals for straightforward chromosomal identification in metaphase chromosome preparations and accurate chromosomal enumeration in interphase nuclei (Choo, 1997; Lee *et al*., 1997a). Unfortunately, the zebrafish-centromeric DNAs characterized to date are similar to those of mouse (e.g., Wong and Rattner, 1988) and other non-primate vertebrate species (e.g., Lee *et al*., 1997b) in that they do not have chromosome-specific repetitive DNA sequences (Sola and Gornung, 2001). Therefore, chromosomal enumeration probes must be chosen from loci near the primary constriction of each chromosome. Indeed, near-centromeric BAC clones have already been successfully used to determine ploidy in interphase cells of *retsina* mutants (e.g., Paw *et al*., 2003) and *mps1*^{zp1} mutants (e.g., Poss *et al*., 2004).

Zebrafish near-telomeric DNA probes have proved useful for enumeration studies to identify and confirm interchromosomal translocations and terminal deletions (i.e., deletions that extend from a given chromosomal locus to the chromosomal terminus). For example, the zebrafish mutant *T3(hoxb)* was thought to have a translocation involving the LG3 chromosome long arm (Fritz *et al*., 1996). Two color FISH assays with near-telomeric probes for the LG3 short arm and long arm confirmed the chromosomal rearrangement in this mutant (Fig. 3). Furthermore, retrotransposed tandem gene duplications within the zebrafish genome have been localized by use of near-telomeric BAC probes from the panel (Yoder *et al*., 2008).

Both near-telomeric and near-centromeric BAC clones can be used to evaluate genomic instability in cancer and cell cycle mutants. For example, a commonly used method for assessing genomic instability involves a recessive mutation in the solute carrier family 24, member 5 gene (*slc24a5*, formerly *golden*) (e.g., Driever *et al*., 1996; Lamason *et al*., 2005) – a gene which resides near the telomere of LG chromosome 18. Zebrafish mutants suspected of having genomic instability are crossed to *golden* heterozygotes. If the mutants promote genomic instability, some of the heterozygous progeny will develop a chromosomal aberration that deletes the one normal *golden* locus in the heterozygotes, which leads to a lack of eye and body pigmentation. Unfortunately, this method for assessing genomic instability assesses only one chromosomal locus and therefore conceivably underestimates the extent of genomic instability in many mutants. In addition, this method for assessing genomic instability is incompatible with early embryonic lethals. Such limitations can be overcome by multicolor FISH experiments using several BAC probes simultaneously, which provides rapid and accurate assessment of genomic instability in the zebrafish (e.g., Fig. 4). Such studies have been performed to evaluate genomic instability and cancer susceptibility in haploid zebrafish assays (e.g., Shepard *et al*., 2005, 2007).

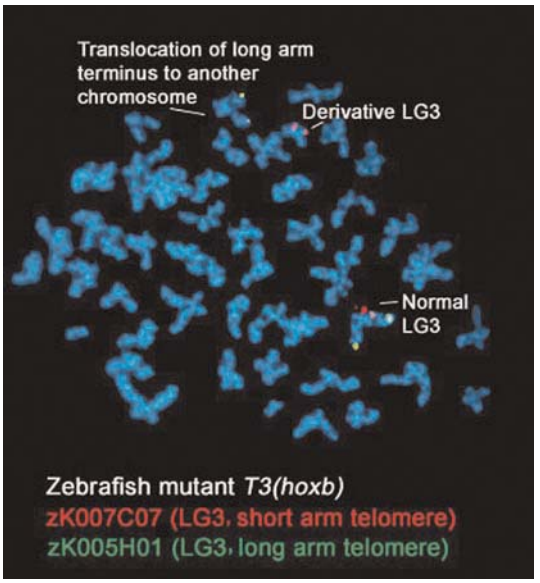


Fig. 3 FISH confirmation of a chromosome translocation in the zebrafish mutant, T3(*hoxb*). Two colored FISH assay shows syntenic hybridization of a near-telomeric short-arm probe and a near-telomeric long-arm probe on one normal LG3 chromosome. The nonsyntenic hybridization of these same probes to two different chromosomes in this zebrafish mutant is consistent with a chromosomal translocation involving one LG3 chromosome and another nonhomologous chromosome. (See color plate.)

IV. Methods for Array CGH

A. DNA Isolation

1. Take freshly anesthetized zebrafish and place in mortar. Thoroughly freeze the zebrafish in liquid nitrogen.
2. Using a pestle, grind the zebrafish into a fine pulp, intermittently adding liquid nitrogen to carefully maintain the tissue as a slurry. Transfer the pulp into a sterile 50 mL centrifuge tube and swirl to allow the remaining liquid nitrogen to evaporate.
3. Add 20 mL of Proteinase K lysis solution (50 mM Tris, 100 mM EDTA, 100 mM NaCl, 1% SDS, 100 μg/mL Proteinase K) to the tube. Shake to resuspend the frozen pellet without causing excess bubbles from the SDS in the lysis solution. Gently rock the solution in a 55 °C incubator for a minimum of 24 h.

   Note: Alternatively, the zebrafish pulp can be frozen in the lysis solution without the proteinase K and kept at −80 °C until ready to isolate DNA. When you are ready to continue, thaw and add 100 μL (20 mg/mL) Proteinase K for a final concentration of 100 μg/mL. Then place in 55 °C incubator for 24 h and continue with protocol.

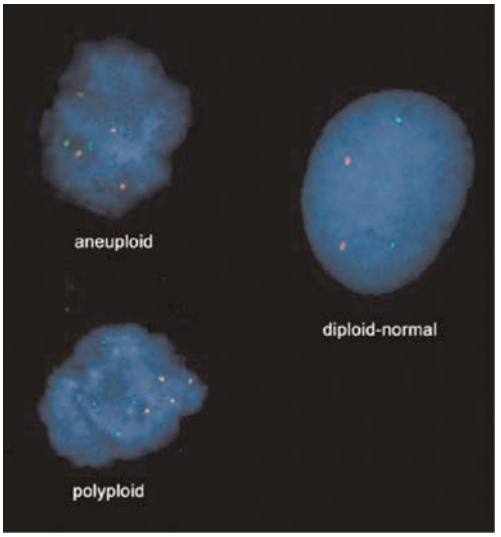


Fig. 4 Use of BAC probes to assess genomic instability. A normal hybridization pattern would show two copies of each BAC probe in a given nucleus ("diploid-normal"). A polyploid cell would show three or more signals for each probe with the number of signals for each BAC probe being the same for a given nucleus. For example, the polyploid cell shown in this figure has four green and four red signals; consistent with a tetraploid cellular content. Aneuploid cells would be expected to have a different number of signals for each BAC probe in a given nucleus. In this figure, the aneuploid cell shown has two copies of the green-labeled BAC probe, zK127K09, for chromosome LG2, and four copies of the red-labeled BAC probe, zC213B07, for chromosome LG16. (See color plate.)

4. Immediately prior to use, prepare a PLG light (50 mL tube) by centrifugation at 1500 *g* for 2 min (5 Prime Inc, Gaithersburg, MD).
5. Transfer lysed solution (20 mL) to the PLG.
6. Add 10 mL buffered Phenol (Phenol-Tris saturated, pH ~8.0; Roche, Indianapolis, IN). Gently rock/invert tube about 20 times to mix the suspension to denature the protein.
7. Add 10 mL chloroform:isoamyl alcohol (24:1) and mix thoroughly as in Step 6.
8. Centrifuge for 5 min at room temperature at 1500 *g* to separate the phases.
9. Transfer the upper aqueous phase into a new PLG 50 mL tube. Repeat Steps 6–8. Cool centrifuge to 4 °C.
10. Transfer the upper aqueous phase to a sterile 50 mL tube. Add 0.1× the volume of 3 M sodium acetate (typically ~2 mL) and mix thoroughly. The volume of the resulting solution will be ~20–25 mL.
11. Add 1× volume (~25 mL) of isopropanol. Gently mix until the DNA starts to aggregate into viscous mass.
12. Incubate at room temperature for 20 min.
13. Pellet the DNA at 800 *g* at 4 °C for 10 min and decant the supernatant.

14. Wash the pellet in 20 mL of ice cold 70% ethanol and centrifuge at 800 *g* at 4 °C for 10 min.
15. Decant the supernatant without dislodging the pellet. Using residual liquid, transfer the DNA pellet into a 2 mL tube. Place tubes upside down at room temperature to evaporate the residual ethanol for 1–2 h.
16. Rehydrate the DNA pellet in 1 mL autoclaved 1× TLE buffer. Incubate overnight (16 h) at 55 °C.
17. Obtain DNA concentration by spectrophotometry.

Note: Alternatively, the DNA pellets can be stored at −20 °C until ready to use. Determine the DNA concentration on the day of use.

B. DNA Labeling

Note: The following steps are for the BioPrime® Total Genomic Labeling Kit (Invitrogen, Carlsbad, CA). The protocol is slightly modified from one provided by the manufacturer (www.invitrogen.com).

1. Use 0.5–1.0 μg DNA for arrays. Adjust volume to 21 μL with H_2O. Use 1.5 mL Eppendorf tube.
2. Denature the DNA at 95 °C for 2 min.
3. Cool the sample on ice for 5 min.
4. Add 20 μL 2.5× random primer solution.
5. Heat at 95 °C for 5 min.
6. Cool on ice for 5 min.
7. Make master mix for each reaction: 5 μL10×dCTP, 3 μL Cy3-dCTP or Cy5-dCTP (PerkinElmer, Waltham, MA), 1 μLexo-Klenowfragment.
8. Add 9 μL of master mix for each reaction for a total of 50 μL.
9. Store at 4 °C until ready for next step.

C. Purification of Labeled Genomic DNA

Note: Set temperature of centrifuge to 4 °C

1. Add 300 μL of 0.1× SSC to a Millipore Amicon Ultra column (Millipore, Billerica, MA)
2. Add 50 μL labeled DNA to the column.
3. Spin at 16,000 *g* for 12 min at 4 °C, then discard the flow through.
4. Add 300 μL of 0.1× SSC to the column.
5. Spin at 16,000 *g* for 12 min at 4 °C, then discard the flow through.
6. Add 50 μL of dH_2O to the column.
7. Invert column onto a new 1.5 mL tube.
8. Spin at 2400 *g* for 2 min at 4 °C.
9. Bring total volume to 80.5 μL using dH_2O.

D. Quantification and Measurement of Dye Incorporation

1. Prepare Nanodropspectrophotometer (Thermo Scientific, Wilmington, DE) by selecting "Microarray" and blanking with water.
2. Use 1.5 μL of sample.
3. Record absorbance at 260, 550, or 650 nm.
4. Calculate (A260/A550) × 23.15 = nt/Cy3
5. Calculate(A260/A650) × 38.58 = nt/Cy5

Note: nt = dye incorporation, which should be between 50 and 80.

E. Pre-hybridization and Hybridization

Note: The following steps are applicable for oligonucleotide-based arrays printed by Agilent technologies (Santa Clara, CA). The protocol is slightly modified from one provided by the manufacturer (www.agilent.com).

1. For each array combine the reagents in Table III.
2. Denature samples at 95 °C for 3 min.
3. Pre-hybridize the chamber at 37 °C for 30 min.
4. Preheat hybridization oven (Agilent, Santa Clara, CA) to 65 °C and set rotation to 20 rpm.
5. Place clean gasket slide (Agilent, Santa Clara, CA) with the Agilent logo facing upwards in the bottom of the chamber.
6. Place 230 μL (for the 400 K feature) or 490 μL (for the 1 million feature) DNA mixture onto the center of the gasket slide and spread along the central axis.
7. Place array slide with Agilent logo facing downwards on top of the gasket slide.
8. Seal slides together using the SureHyb hybridization chamber (Agilent, Santa Clara, CA).
9. Rotate sealed chamber in a circle to dislodge small air bubbles; examine from the back. It is normal to have one large air bubble freely rotating.
10. Place chamber in 65 °C oven for 24 h. Make sure rotating apparatus is balanced.

Table III
Reagents for hybridization of array CGH

	Amount for 1 million feature array	Reagents
39.5 μL	79 μL	Test DNA (labeled with Cy5-dCTP)
39.5 μL	79 μL	Reference DNA (labeled with Cy3-dCTP)
25 μL	50 μL	Herring sperm DNA (Promega, Madison, WI)
26 μL	52 μL	Blocking reagent (# 5188-5220, Agilent, Santa Clara, CA)
130 μL	260 μL	2× Hybridization buffer (# 5188-5220, Agilent, Santa Clara, CA)
260 μL	520 μL	Total volume

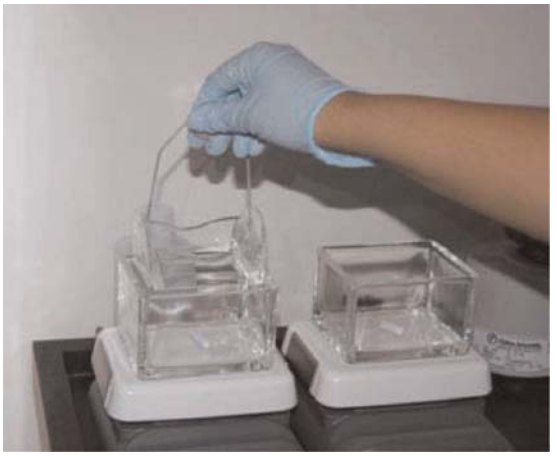


Fig. 5 Slides being placed in wash buffer 1. (See color plate.)

F. Post-hybridization Wash

Note: Use the following wash buffers from Agilent. Oligo aCGH Wash Buffer 1 (#5188-5221, Agilent, Santa Clara, CA), and Oligo aCGH Wash Buffer 2 (#5188-5222, Agilent, Santa Clara, CA).

1. Preheat Wash 2 to 37 °C.
2. Wearing gloves, disassemble slides while immersed in Wash 1.
3. Place slides in Wash 1 for 5 min at room temperature on stir plate (stir setting 4) (Fig. 5).
4. *Note*: Maximum of five slides per wash.
5. Place slides in Wash 2 at 37 °C for 3 min on stir plate (stir setting 4).
6. Remove slide(s) very slowly and allow to dry.
7. Scan slides with a 2 μm scanner as soon as possible to avoid degradation of Cy-dyes. *Note*: Ozone levels above 10 ppb may have an effect on Cy5 when exposed for 10–30 s. Cy5 is also disproportionately affected by photobleaching. It takes higher levels of ozone (greater than 100 ppb) to have an effect on Cy3 (Fare *et al*., 2003).

V. Application of Array CGH in Zebrafish Studies

aCGH is an effective tool for examining an entire genome for regions exhibiting DNA gains and losses [i.e., copy number variants (CNVs)]. The first aCGH platform for zebrafish (Freeman *et al*., 2009) was developed using BACs containing

sequences orthologous to human oncogenes and tumor suppressor genes. This array illustrated the effectiveness of aCGH for investigating genomic variation in zebrafish cancer models. Some cancer studies in humans are limited by the inability to obtain matched tumor/normal samples from the same individual, thus creating the challenge of distinguishing somatic changes in tumors from germline copy variants. This limitation is avoided with zebrafish aCGH studies, because control DNA and tumor DNA can be obtained from the same fish, thus resulting in increased detection of tumor-specific alterations. Using this approach with zebrafish T-ALL samples, serial tumor cell transplantations (i.e., allo-transplantations) with three zebrafish malignancies were recently examined using an oligonucleotide aCGH array containing 400,000 features to select for aggressive disease (Rudner *et al*., 2011). Somatic differences resulting from the transplantations were detected (Fig. 6). Because these cells were transplanted through multiple fish, the accumulation after passage of additional somatic or "*de novo*" copy number mutations could be detected (Fig. 6).

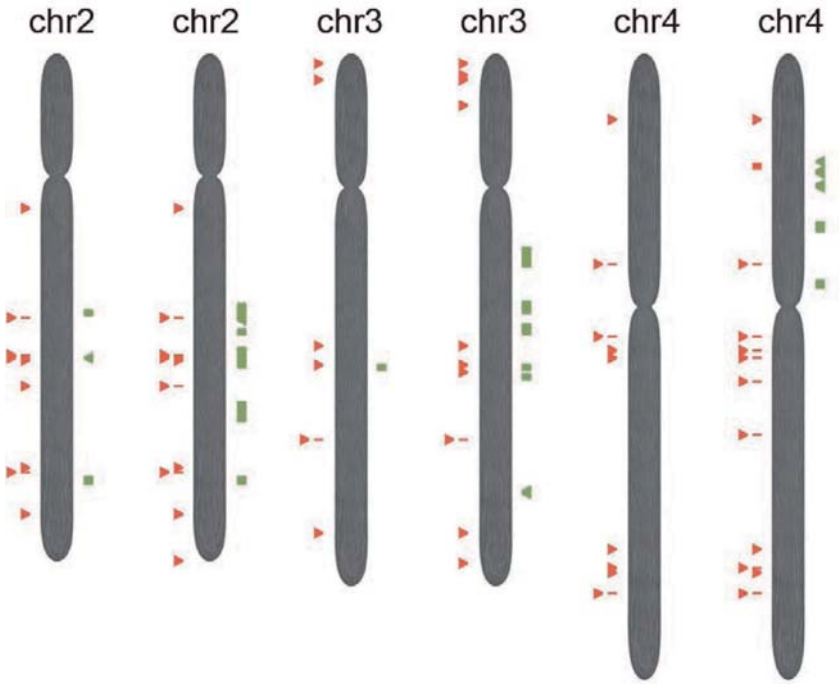


Fig. 6 Accumulation of genomic imbalances in a zebrafish T-ALL model. Data from three chromosomes in one T-ALL clone that has undergone allo-transplantation. Note the additional aberrations that are detected after "passaging" (left chromosome – before passaging, right chromosome – after passaging). Red markings indicate losses, green markings indicate gains. (See color plate.)

VI. Summary

Improvements to FISH technologies and the arrival of aCGH technology have led to rapid technological advancements in molecular cytogenetics. Use of these new technologies in zebrafish cytogenetics, including the BAC probe panel, has resulted in improved assembly of the zebrafish reference genome. Array CGH has been used to identify genomic imbalances throughout the zebrafish genome (e.g., Freeman *et al*., 2009). Moreover, combining these techniques with new technologies, such as next generation sequencing, will soon lead to the characterization of genomic alterations with breakpoints at the nucleotide level. While neither of these technologies alone can illuminate all possible aberrations within a genome, their combination will facilitate a greater understanding of the genetic alterations that are present in specific disease models.

Acknowledgments

The authors would like to thank Mr. Jason Dobrinski for his assistance with Figure 6 and Mrs. Shona Stephan for her assistance with Figures 1–5.

References

Amores, A., and Postlethwait, J. H. (1999). Banded chromosomes and the zebrafish karyotype. *Methods Cell Biol.* **60**, 323–338.

Choo, K. H. A. (1997). *The Centromere.* Oxford University Press, Oxford.

Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., and Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46.

Fare, T., Coffey, E., and Hongyue, D., *et al*. (2003). Effects of atmospheric ozone on microarray data quality. *Anal. Chem.* **75**, 4672–4675.

Freeman, J., Adeniyi, A., Banerjee, R., Dallaire, S., Maguire, S., Chi, J., Ng, B., Zepeda, C., Scott, C., Humphray, S., Rogers, J., and Zhou, Y., *et al*. (2007). Definition of the zebrafish genome using flow cytometry and cytogenetic mapping. *BMC Genomics* **8**, n195.

Freeman, J. L., Ceol, C., Feng, H., Langenau, D. M., Belair, C., Stern, H. M., Song, A., Paw, B. H., Look, A. T., Zhou, Y., Zon, L. I., and Lee, C. (2009). Construction and application of a zebrafish array CGH platform. *Genes Chromosomes Cancer* **48**, 155–170.

Fritz, A., Rozowski, M., Walker, C., and Westerfield, M. (1996). Identification of selected gamma-ray induced deficiencies in zebrafish using multiplex polymerase chain reaction. *Genetics* **144**, 1735–1745.

Gornung, E., De Innocentiis, D., Annesi, F., and Sola, L. (2000). Zebrafish 5S rRNA genes map to the long arms of chromosome 3. *Chromosome Res.* **8**, 362.

John, B. (1988). *In* "Heterochromatin: Molecular and Structural Aspects," (and R. S. Verma, ed.), pp. 1–30. Cambridge University Press, Cambridge.

Kallioniemi, A., Kallioniemi, O., Sudar, D., Rutovitz, D., Gray, J., Waldman, F., and Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821.

Lamason, R. L., Mohideen, M. A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X. Y., Humphreville, V. R., Humbert, J. E., Sinha, S., and Moore, J. L., *et al*. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786.

Lee, C., and Smith, A. (2004). Molecular cytogenetic methodologies and bacterial artificial chromosome (BAC) probe panel resource for genomic analysis in zebrafish. *In* "Methods in Cell Biology," (H. W. Detrich, L. I. Zon, and M. Westerfield, eds.), pp. 241–255. Elsevier Academic Press, San Diego, CA.

Lee, C., Wevrick, R., Fisher, R. B., Ferguson-Smith, M. A., and Lin, C. C. (1997a). Human Centromeric DNAs. *Hum. Genet.* **100**, 291–304.

Lee, C., Court, D. R., Cho, C., Haslet, J. L., and Lin, C. C. (1997b). Higher-order organization of subrepeats and the evolution of cervid satellite I DNA. *J. Mol. Evol* **44**, 327–335.

Paw, B. H., Davidson, A. J., Zhou, Y., Li, R., Pratt, S. J., Lee, C., Trede, N. S., Brownlie, A., Donovan, A., Liao, E. C., Ziai, J. M., and Drejer, A. H., *et al*. (2003). Cell-specific mitotic defect and dyserythropoiesis associated with erythroid band 3 deficiency. *Nat. Genet.* **34**, 59–64.

Peterson, S., and Freeman, J. (2009). Cancer cytogenetics in the zebrafish. *Zebrafish* **6**, 355–360.

Phillips, R. B., Amores, A., Morasch, M. R., Wilson, C., and Postlethwait, J. H. (2006). Assignment of zebrafish genetic linkage groups to chromosomes. *Cytogenet. Genome Res.* **114**, 155–162.

Phillips, R. B., and Reed, K. M. (2000). Localization of repetitive DNAs to zebrafish (Danio rerio) chromosomes by fluorescence in situ hybridization (FISH). *Chromosome Res.* **8**, 27–35.

Poss, K. D., Nechiporuk, A., Stringer, K. F., Lee, C., and Keating, M. T. (2004). Germ cell aneuploidy in zebrafish with mutations in the mitotic checkpoint gene mps1. *Genes Dev.* **18**, 1527–1532.

Rudner, L., Brown, K. H., Dobrinski, K. P., Bradley, D. F., Downie, J. M., Meeker, N. D., Smith, A. C. H., Garcia, M. I., Gutierrez, A., Mullighan, C. G., Look, A. T., Downing, J. R., Schiffman, J. D., Lee, C., Trede, N. S., and Frazer, J. K. (2011). Shared acquired genomic changes in zebrafish and human T-ALL. *Oncogene.* (in press).

Shepard, J. L., Amatruda, J. F., Finkelstein, D., Ziai, J., Finley, K. R., Stern, H. M., Chiang, K., Hersey, C., Barut, B., Freeman, J. L., Lee, C., and Glickman, J. N., *et al*. (2007). A mutation in separase causes genome instability and increased susceptibility to epithelial cancer. *Genes Dev.* **21**, 55–59.

Shepard, J. L., Amatruda, J. F., Stern, H. M., Subramanian, A., Finkelstein, D., Ziai, J., Finley, K. R., Pfaff, K. L., Hersey, C., Zhou, Y., Barut, B., and Freedman, M., *et al*. (2005). Zebrafish bmyb mutation causes genome instability and increased cancer susceptibility. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13194–13199.

Sola, L., and Gornung, E. (2001). Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. *Genetica* **111**, 397–412.

Westerfield, M. (2007). *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 5th edn. University of Oregon Press, Eugene, OR.*

Wong, A. K., and Rattner, J. B. (1988). Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res.* **16**, 11645–11661.

Yoder, J. A., Cannon, J. P., Litman, R. T., Murphy, C., Freeman, J. L., and Litman, G. W. (2008). Evidence for a transposition event in a second NITR gene cluster in zebrafish. *Immunogenetics* **60**, 257–265.

CHAPTER 15

Conserved Synteny and the Zebrafish Genome

Julian M. Catchen[*], Ingo Braasch[†] and John H. Postlethwait[†]

[*]University of Oregon, Center for Ecology and Evolutionary Biology, Eugene Oregon, USA

[†]University of Oregon, Institute of Neuroscience, Eugene Oregon, USA

Abstract

Zebrafish offers significant opportunities for the investigation of vertebrate development, evolution, physiology, and behavior and provides numerous models of human disease. Connecting zebrafish phenogenetic biology to that of humans and other vertebrates, however, requires the proper assignment of gene orthologies. Orthology assignments by phylogenetic analysis or by reciprocal best sequence similarity searches can lead to errors, especially in cases of gene duplication followed by gene

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00015-X

loss or rapid lineage-specific gene evolution. Conserved synteny analysis provides a method that helps overcome such problems. Here we describe conserved synteny analysis for zebrafish genes and discuss the Synteny Database, a website specifically designed to identify conserved syntenies for zebrafish that takes into account the teleost genome duplication (TGD). We utilize the Synteny Database to demonstrate its power to resolve our understanding of the evolution of nerve growth factor receptor related genes, including *Ngfr* and the enigmatic *Nradd*. Finally, we compare conserved syntenies between zebrafish, stickleback, spotted gar, and human to understand the timing of chromosome rearrangements in teleost genome evolution. An improved understanding of gene histories that comes from the application of tools provided by the Synteny Database can facilitate the connectivity of zebrafish and human genomes.

I. Introduction

Zebrafish provides many models that have advanced our understanding of human disease. For example, ferroportin mutations leading to hemachromatosis were first discovered in zebrafish and later in human patients (Donovan *et al*., 2000; Montosi *et al*., 2001); the role of the microRNA miR-140 in cleft palate was elucidated first in zebrafish and then found to be responsible for cleft palate in some human families (Eberhart *et al*., 2008; Li *et al*., 2010); epithelial-macrophage-bacterial interactions first identified in zebrafish promise a host-targeting tuberculosis therapy for humans (Volkman *et al*., 2010); *slc24a5*, which was first identified as a pigmentation gene in pale zebrafish mutants, was found to contribute to the evolution of light skin pigment in northern human populations (Lamason *et al*., 2005); and the validation of *pdzd7* in zebrafish as a modifier of known human Usher syndrome genes led to the addition of *PDZD7* to the genetic testing panel provided to potential Usher patients (Ebermann *et al*., 2010). Direct application of such knowledge from zebrafish to humans requires the accurate assignment of gene orthologies. Orthologs are pairs of genes (one in each of two different species) that evolved from a single gene in the last common ancestor of those two species. In contrast, paralogs are genes that are related to each other by gene duplication events occurring within a lineage. The paralog of a gene in one species can be confused with that gene's ortholog, especially in cases of gene duplication, lineage-specific gene loss, and rapid or asymmetric evolution. To connect genetic information from zebrafish to humans correctly, one must properly identify the human ortholog of the zebrafish gene.

II. Rationale: Gene Duplication and Orthology Assignments

Analysis of gene phylogenies is one of the best ways to help assign orthologs. The examination of gene trees to assign orthologies, however, can lead to difficulties in cases of gene duplication, rapid and unequal evolutionary rates, and lateral gene

transfer. For example, consider an initial gene *g* embedded in a chromosome with other nearby genes (Fig. 1A). After a round of duplication R1, the paralogs *g1* and *g2* begin to diverge from each other in sequence and perhaps in function. After a speciation event S, orthologs begin to diverge from each other (*g1* from *G1* and

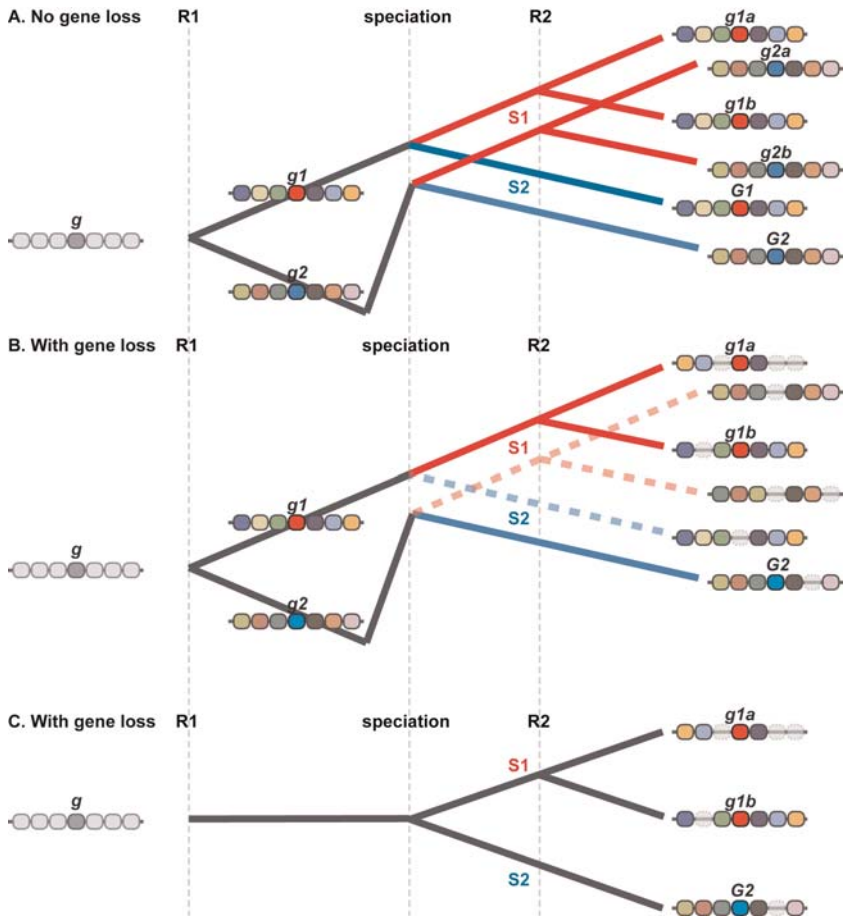


Fig. 1 Gene duplication followed by lineage-specific gene loss can sometimes lead to erroneous assignments of orthology. A. A chromosome segment with seven genes including the one of interest, *g*, becomes duplicated in an R1 round of duplication to give genes *g1* and *g2*, followed by divergence in sequence and perhaps in function of these duplicated genes. After a speciation event leads to two species, S1 and S2, each will have orthologs of both *g1* and *g2* (called *G1* and *G2* in Species-2). A second round of duplication (R2) of this segment in the lineage of Species-1 would lead to duplicates of both *g1* and *g2*. Regular rates of sequence divergence would result in phylogenetic trees that would recreate gene history. B. If the history in part A were modified by the loss of *g2* in S1 and the reciprocal loss of *G1* in S2 (dashed lines; ghost gray genes represent losses), then phylogenetic analyses would fail to reconstruct history. C. The tree produced after the gene history in part B would lead to the erroneous conclusion that *g1a* and *g1b* are co-orthologs of *G2*. (See color plate.)

g2 from *G2*). Finally, a second round of genome duplication R2 in the lineage of Species-1 can give gene duplicates, including *g1a, g1b, g2a*, and *g2b*. Assuming relatively regular rates of gene evolution and no gene losses, phylogenetic analysis will accurately recreate this history.

In the case of gene losses, phylogenies can give incorrect orthology assignments (Fig. 1B). For example, if before the R2 duplication, Species-1 loses *g2* but Species-2 loses *G1*, then the resulting gene tree would suggest that *g1a* and *g1b* are co-orthologs of *G2* (Fig. 1C), a result inconsistent with true gene histories. Incorrect ortholog assignments such as this can obfuscate the connectivity of zebrafish and human functional genomics. As an example of this type of lineage-specific gene loss after gene duplication, consider *hox* clusters in teleost fish: the lineages of zebrafish and percomorphs (including stickleback, pufferfish, medaka) reciprocally lost different entire *hox* clusters that were present in duplicated copies in basally diverging teleosts: *hoxda* was lost from the zebrafish lineage and *hoxcb* was lost from the percomorph lineage (Amores *et al*., 1998; Guo *et al*., 2010; Kurosawa *et al*., 2006; Prohaska and Stadler, 2004). As a result, zebrafish has no ortholog of any *hoxdb* gene in percomorphs and percomorphs have no ortholog of any *hoxcb* genes in zebrafish.

Gene duplication can cause a problem with orthology assignments due not only to gene losses (Fig. 1) but also due to the rapid and uneven rates at which duplicated genes sometimes evolve compared to singletons within the same lineage. Members of a pair of duplicated genes often show substantial asymmetries in evolutionary rates and frequently more radical amino acid substitutions (Byrne and Wolfe, 2007; Leong *et al*., 2010; Steinke *et al*., 2006; Yampolsky and Bouzinier, 2010). These rapid divergence rates may signal the origin of novel, positively selected functions as indicated by experiments in yeast that showed that rate increases occur soon after the whole genome duplication (WGD) (Byrne and Wolfe, 2007). Alternatively, selection may become relaxed on one or the other gene duplicate leading to rapid neutral drift in sequence. Long branches can lead to an artifact in which one member of a duplicate pair diverges basal to the position appropriate for its true history; artifacts such as these can lead to a problem in orthology designation.

A. Types of Gene Duplication Events

Gene duplications that can obscure orthology assignments are of three main types: retrotranspositions, tandem duplications, and WGDs. In retrotransposition, gene transcripts are reconverted to DNA and inserted at some other, often distant, location in the genome. The retrotransposed gene usually lacks introns, is missing its native regulatory elements, and has new gene neighbors different from those of the ancestral gene copy. The usual fate of a retrotransposed gene is to become a processed pseudogene. Some retrotransposed genes, however, may by chance enter the genome near regulatory elements that allow it to be expressed, although the expression pattern is unlikely to be related to its original expression domain. If a

retrotransposition favors fitness, then natural selection may preserve the retrotransposed gene. The zebrafish genome has at least 652 retrocopies of which 437 appear to be transcribed (Fu *et al*., 2010).

In tandem duplication events (or segmental duplications), a segment of chromosome containing one or more genes can duplicate with the two copies occupying adjacent genomic locations in head-to-head, head-to-tail, or tail-to-tail orientation. Tandem duplication events may duplicate all or only part of a gene's regulatory elements; as a consequence, one of the duplicate copies may have regulatory differences from its sister paralog from the very beginning; in addition, regulatory elements between the genes may help to regulate them both. Many gene families show such organization, like the *hox* clusters and *dlx* tandem duplicates.

In whole genome duplication (WGD or tetraploidization) events, all genes in a genome become duplicated with their regulatory elements preserved intact. WGD is especially important for vertebrate lineages because vertebrates experienced multiple rounds of WGD – two rounds (called R1 and R2) early in vertebrate evolution (Dehal and Boore, 2005; Garcia-Fernandez and Holland, 1994; Spring, 1997). After the divergence of cephalochordate and urochordate lineages from the vertebrate lineage, the R1 and R2 events may have happened in quick succession or they may have been separated in time with one event before and one after the divergence of jawless and jawed vertebrates (Force *et al.*, 2002; Kuraku *et al*., 2009; Putnam *et al*., 2008). An additional round of WGD was shared by all teleosts, including zebrafish (the teleost genome duplication, TGD) (Amores *et al*., 1998; Jaillon *et al*., 2004; Postlethwait *et al*., 1998; Taylor *et al*., 2003; Woods *et al*., 2000). As a legacy of the TGD, zebrafish and other teleosts often have duplicates of many genes present in single copy in tetrapods.

B. Evolution After Gene Duplication

In the case of gene duplication by autotetraploidy, the two gene paralogs (called ohnologs in honor of S. Ohno (Ohno, 1970; Wolfe, 2000)) are initially identical, so the loss of one copy (non-functionalization) invokes a selective penalty only in the case of dosage-sensitive genes; loss is the most frequent fate for duplicated genes (Lynch and Conery, 2000; Watterson, 1983). Gene loss after genome duplication leads to chromosomes with gene contents like those shown at the right of Fig. 1B, in which duplicated chromosome segments contain the original gene order with gaps due to missing genes. As an alternative to gene loss, the functions of two WGD paralogs may diverge so that both eventually become positively selected, either because they reciprocally lost essential ancestral subfunctions by neutral evolution (subfunctionalization), which results in both genes becoming essential for normal fitness, or because one or both evolved novel, favorable functions (neofunctionalization) (Force *et al*., 1999; Postlethwait *et al*., 2004). Once neutral evolutionary processes have preserved subfunctionalized genes, adaptive evolution can fine-tune retained subfunctions to improve fitness relative to the ancestral, unduplicated state (Des Marais and Rausher, 2008; Force *et al*., 1999; Hughes, 1994; Innan and

Kondrashov, 2010; Jarinova *et al*., 2008; Jovelin *et al*., 2007; Postlethwait *et al*., 2004; Turunen *et al*., 2009; van Hoof, 2005; Winkler *et al*., 2003). Subfunctionalization does not cause genes to develop new functions, but instead it preserves duplicates long enough to provide an opportunity for the evolution of adaptive change. A large survey in yeast found that duplicated genes resulting from a WGD event diverged with respect to regulatory control more often than they diverged in their biochemical functions (Wapinski *et al*., 2007). Given the breadth of the evidence, it is likely that neofunctionalization and subfunctionalization are both active evolutionary processes (Innan and Kondrashov, 2010).

In summary, to connect zebrafish developmental genetic models to human disease-related genes, we must develop strategies to overcome the difficulties imposed on orthology calls by gene duplications. Here we discuss various ways to assign orthologies, including the use of conserved syntenies. We present an automated system to infer historical relationships between genes that takes into account WGD. Additionally, the system integrates genomic conservation of synteny, providing an additional source of evidence for orthology assignments that complements phylogenetic methods. We use the system here to examine the evolution of nerve growth factor receptor related genes, including *Ngfr* and the enigmatic *Nradd*, and compare conserved syntenies in human, zebrafish, and a basally diverging rayfin fish to elucidate the timing of chromosome rearrangements relative to the TGD.

III. Methods

A. BLAST-based Orthology Calling

One of the most commonly used methods to assign orthologies is to search for gene or protein sequences that are the most similar to a query gene or protein. Alignment algorithms determine sequence similarity and use statistical significance to infer historical relatedness. Algorithms search for a gene (the "subject" or the "hit" ') in one species that aligns best to a sequence (the "query") in another species; the hit then becomes the query sequence and the algorithm repeats the search in the opposite direction. If the reciprocal search identifies as best hit the original query sequence, then the algorithm has identified a reciprocal best hit (RBH) (Wall *et al*., 2003), and infers that the pair of sequences represent orthologs. BLAST (Basic Local Alignment Search Tool) (Altschul *et al*., 1997) is a commonly used alignment similarity algorithm.

An algorithm that automates this methodology, INPARANOID, assigns paralogy and orthology initially by using BLAST to identify candidate homologs between two gene datasets and then sorting hits to find the most appropriate matches (Berglund *et al*., 2008; Remm *et al*., 2001). Given species *A* and *B*, INPARANOID calculates sequence similarity scores between all genes in set *A* versus set *A*, all genes in set *A* versus set *B*, *B* versus *A*, and finally *B* versus *B*. Reciprocal best hits are recorded when relationships are unambiguous and the user can manipulate several variables to

limit genes returned in these pairwise comparisons, including a BLAST-score cutoff and a minimum alignment length. Next, the reciprocal best hits serve as seeds to create an initial set of gene clusters; a series of heuristic rules guide the addition of genes to the clusters and the combination and division of clusters. When the algorithm is complete, sequence clusters represent groups of orthologous genes where members of each species-specific cluster represent paralogs of the orthologous gene. BLAST scores serve as a measure of distance between genes under the assumption that the evolutionary rates between paralogs and orthologs are about equal. INPARANOID's heuristic, BLAST-based approach is fast and can examine large datasets rapidly; its assumption of an equal evolutionary rate among paralogs and orthologs is problematic, however, and the use of arbitrary, manual cut-off limits can cause some inconsistency in sequences that it retains for the clustering portion of the algorithm.

An alternative clustering method (OrthoMCL) dispenses with heuristic clustering rules and processes multiple species together in a unified analysis (Li *et al*., 2003). OrthoMCL also uses BLAST to obtain initial pairwise homology scores for all genes considered and it uses reciprocal best hits to identify initial sets of paralogs and orthologs. OrthoMCL then normalizes scores between genes from different genomes and models gene homologies as a graph, with nodes representing genes and edges connecting nodes as BLAST hits weighted by the BLAST score. This graph is fed into a Markov clustering algorithm (Enright *et al*., 2002), assuming that evolutionarily related genes will be highly connected; matrix transformations exacerbate the natural structure of the graph until separate clusters become disconnected and these disconnected subgraphs define the final groupings of orthologs and paralogs.

B. Conserved Syntenies

An alternative method to help assign orthologies despite genome duplication incorporates evidence from conserved synteny. A *synteny* occurs if two genes are located physically on a single chromosome within a single species (syn: same; tene: thread) (Ruddle, 1972). Even if the two genes are so distant that they are unlinked in a meiotic mapping cross (i.e., they assort independently according to Mendel), they are still syntenic if they are on the same chromosome. A synteny differs from a *conserved synteny*, which occurs if (1) a pair of genes is syntenic in one species; (2) the orthologs of that pair of genes are syntenic in another species; and (3) the ancestors of those genes were syntenic in the last common ancestor of the two species. Note that people sometimes talk of genes as being "syntenic" as shorthand to mean they represent a "conserved synteny".

Conserved syntenies fall into four increasingly stringent categories. In the first category, two or more genes from a single chromosome in one genome are orthologous to two or more genes on a single chromosome in a second genome. In the second category, regions of conserved synteny also exhibit conservation of gene order. The third category adds conservation of transcript orientation for all genes in the region, and the fourth category represents a conserved block – including conserved gene order, transcription orientation, and no intervening or lost genes.

In an autotetraploidization event, duplicated chromosomes (homeologs) initially conserve syntenies completely with each other and with the ancestral unduplicated chromosome, like the two chromosomes after R1 in Fig. 1B. Between the duplication event and the speciation event, however, genes can be lost from one homeolog or the other; such a gene loss causes the gene content of the two homeologs to sum to the gene content in the ancestral chromosome segment. Chromosomal translocations involving one or the other homeolog and some other chromosome can disrupt conserved syntenies. In contrast to translocations, inversions disrupt only gene orders, not conserved syntenies. Regulatory elements embedded in distant locations can inhibit the fixation of chromosome translocations that disrupt conserved syntenies as well as inversions, which disrupt gene orders (Kikuta *et al*., 2007).

C. Using Conserved Syntenies to Help Infer Orthologies

Conserved syntenies help identify orthologies because gene neighborhoods tend to persist during evolution; as a result, entire gene neighborhoods in different species tend to be orthologous; over time, however, conserved syntenies tend to degrade due to the fixation of translocations in diverging populations. In Fig. 1, we could test the hypothesis that genes *g1a* and *g1b* in Species-1 are co-orthologous to gene *G2* in Species-2 by searching for the orthologs of the neighbors of *G2* in Species-1. If the Species-1 orthologs of *G2* neighbors are syntenic with *g1a* and/or *g1b*, then the hypothesis that *g1a* and *g1b* are co-orthologs of *G2* is supported. If, however, the orthologs of *G2* are on a pair of chromosomes that do not have a *g*-related gene, as is the case in Fig. 1B, then the hypothesis that *g1a* and *g1b* are co-orthologs of *G2* is not supported. Note that this method uses information about *G2*'s position, but not its sequence similarities; thus, this method is independent of the observed phylogeny of gene *G2* itself.

Several systems are available for the examination of conserved syntenies of zebrafish with other organisms. Synorth specializes in identifying genomic regulatory blocks – clusters of conserved non-coding elements that regulate distant developmental genes and thereby maintain syntenies. Synorth (http://synorth.genereg.net/) can follow evolutionary fate after the TGD and uses the March 2006 version of the human genome (Dong *et al*., 2009). Narcisse (http://www.narcisse.com/) allows the exploration of segment conservation in animal genomes but does not deal well with duplications and is currently not available on line (Courcelle *et al*., 2008). The AutoGRAPH synteny viewer (http://genoweb.univ-rennes1.fr/tom_dog/AutoGRAPH/) functions for six mammals and does not include zebrafish (Derrien *et al*., 2007). Cinteny (http://cinteny.cchmc.org) allows one to identify regions of conserved synteny between individual zebrafish chromosomes and several tetrapods, but no other fish (Sinha and Meller, 2007). Cinteny provides a visual connection of markers from one chromosome to the other, and provides the locations of the connected genes but not their names. Genomicus (http://www.dyogen.ens.fr/genomicus) is a conserved synteny browser that includes all genomes available at Ensembl in a phylogenetic context and has the useful feature that it can reconstruct

ancestral gene organizations (Muffato *et al*., 2010). One can enter a gene name and select a species, and Genomicus will show alignments, including both zebrafish homeologous chromosomes in the case of duplicated zebrafish genes. Here we discuss The Synteny Database (Catchen *et al*., 2009), a system specially developed to view and interpret conserved syntenies and duplicated regions in zebrafish. The Synteny Database has two components: an mRBH Pipeline and a Synteny Pipeline.

1. The mRBH Analysis Pipeline

The modified Reciprocal Best Hit (mRBH) Pipeline (Catchen *et al*., 2009) was designed to accommodate WGD. mRBH identifies paralogous gene groups in a primary genome (rather than a single "best" hit) and then anchors those gene groups to an ortholog in an outgroup genome using a BLAST-based approach. The pipeline creates paralogous groups relative to the last WGD that occurred in the lineage of the primary genome but not in the lineage of the outgroup genome. For example, if the lineage of the primary genome experienced a genome duplication after it diverged from the lineage of the outgroup genome, then the pipeline will produce gene groups of size two in the primary genome. If, on the other hand, genome duplication occurred before the two species diverged, then the pipeline reverts to a simple ortholog pipeline with a one-to-one correspondence between genes in the primary and outgroup genomes. In practice, recent tandem gene duplication, gene loss, and sequence divergence can influence the number of genes per group.

The mRBH Pipeline first takes the protein sequence of every gene in the primary genome and performs a BLASTp search against all other proteins encoded by the primary genome. In the case of multiple splice variants, the pipeline performs a search for each transcript. Following the within-primary-genome search, the pipeline conducts a BLASTp search using each protein from the primary genome as a query against the outgroup genome; any sequences found then serve as queries to search back into the primary genome (a retro-BLAST). The mRBH Pipeline uses NCBI BLAST with the BLOSUM62 substitution matrix (Henikoff and Henikoff, 1992) and records only BLAST hits with an *E*-value below 1×10^{-5}.

The mRBH Pipeline next uses the collected BLAST results to build paralogy groups. Given paralogs A, B, and C, only two can be reciprocal *best* hits. Transitivity, however, can accommodate multiple duplication events: If genes A and B are traditional reciprocal best hits, then if gene C's best hit is either A or B and A or B's next best hit is C, then genes A, B, and C should all be considered modified reciprocal best hits. Modeling genes as nodes in a graph and using the BLAST hits between genes to connect the nodes into a directed graph can identify these transitive groups. The graph can then be traversed using a single-linkage clustering algorithm (Van de Peer, 2004), implemented by traversing a directed graph, to build paralogous groups.

2. Noise Reduction

A major issue governing the size of inferred paralog groups is the number of BLAST hits made available to the single-linkage clustering algorithm. If a gene has

one or more conserved domains, or even if a gene contains weakly conserved motifs, then BLAST will pick up those regions in its search for statistically significant local alignments. Because each BLAST hit is a potential edge in the directed graph, the system must limit those edges to hits that are likely to provide information to infer real paralogy and orthology, not simply a small, well-conserved protein domain. Several heuristic approaches attempt to eliminate noise from BLAST results (Hahn, 2007; Li *et al*., 2003; Remm *et al*., 2001), but they tend to produce sometimes arbitrary results based on the use of cutoff values that do not scale with respect to evolutionary distance. The mRBH Pipeline uses a novel noise reduction algorithm based on a standard hierarchical clustering algorithm that excludes insignificant BLAST hits for each query gene. To avoid the problem of arbitrary cutoffs and to compare genomes at different evolutionary distances, the pipeline decides how to cluster the BLAST results by permuting search results to create a null distribution; it then applies the gap statistic (Tibshirani *et al*., 2001). BLAST hits that fall into clusters below the selected significance level are discarded as background noise. The pipeline applies this procedure to each set of BLAST hits that it generates.

3. Outgroup Anchoring

Anchoring a group of paralogs to the outgroup helps define the origin of gene duplication events. Prior to outgroup anchoring, the analysis pipeline constructs paralogous groups in the primary genome. The system then identifies the top BLAST hit in the outgroup genome for each member of each paralogy group in the primary genome. If a group member does not have a significant BLAST hit in the outgroup, the pipeline drops that group member from further consideration. If members of a paralogous group have best BLAST hits to different genes in the outgroup, then the pipeline splits the group, with each subset of the original group being anchored to the appropriate (orthologous) outgroup gene. BLAST hits for outgroup genes are then checked to ensure that the outgroup gene hits the original gene in the primary genome (although it does not have to be the top hit). If an outgroup gene does not retro-BLAST back to one of the genes in the original paralogy group, then that gene from the primary genome is eliminated from the group. Finally, the system performs the outgroup anchoring analysis on all genes in the primary genome that had not been assigned to a paralogous group, i.e., singletons, to attempt to identify orthologs for all genes. The end result is a series of paralogous gene groups from the primary genome each anchored to a single gene in the outgroup.

4. The Problem of Variable Rates of Paralog Evolution

Ohnologs often show asymmetry in evolutionary rate – one member of a pair of ohnologs often evolves at a faster rate than the other. Evidence suggests that rate acceleration occurs soon after a WGD event in one of the duplicates (Byrne and Wolfe, 2007). This phenomenon is a major limiting factor both for RBH-based

orthology assignment algorithms and for phylogeny-based methods (Fares *et al*., 2006; Hahn, 2007). Rate asymmetries can lead to incorrect assignments in cases involving paralogs that are substantially diverged because the BLAST search favors the strongest sequence conservation between genes, regardless of true ancestry.

To illustrate the pernicious effects of gene duplication, lineage-specific gene loss, and asymmetrical rates of paralog evolution, consider the MSX gene family (Akimenko *et al*., 1995; Ekker *et al*., 1997; Jabs *et al*., 1993; Jezewski *et al*., 2003; Shimeld *et al*., 1996). Zebrafish has five MSX paralogs, mouse has three, and human has two active genes and a pseudogene. Orthologies in this family are difficult, but conserved synteny analysis suggests orthologies for *zebrafish*/*Mouse/HUMAN* as: *msxe*/*Msx1*/*MSX,* (*msxa & msxd*)*/Msx2/MSX2*, and (*msxb & msxc)/Msx3*/'MSX2P', a pseudogene orthologous to *Msx3* in mouse, despite its official name (Catchen *et al*., 2009; Finnerty *et al*., 2009). Reciprocal BLAST analysis suggests that *msxa*, *msxb*, *msxc*, and *msxd* are all more closely related to *Msx2* than to *Msx1* or *Msx3*. These relationships, however, are likely to be incorrect due to rate asymmetries – phylogenetic trees show long branches for *Msx3* in mouse relative to *Msx1* and *Msx2*; furthermore, the conversion of the human ortholog of *Msx3* into a pseudogene suggests that selection pressures maintaining its sequence integrity have been weak or lacking. BLAST search results for *msxb* against the mouse genome return *Msx2* as its top BLAST hit (expect-value 1.1e-55) with the second best BLAST hit as *Msx3* (expect-value 4.2e-53). The reverse-BLAST search using *Msx2* as a query against the zebrafish genome returns *msxb* as the fourth best hit, giving *msxd*, *msxa*, *msxc*, and *msxb* in that order (expect-values of 9.6e-65, 4.3e-61, 2.3e-59, and 8e-56, respectively); this result is not expected if *msxb* and *Msx2* are orthologs. Performing a reverse-BLAST with *msxb*'s second best forward hit *Msx3* as query returns *msxc* and *msxb* as the top two hits in that order (expect-values of 1.6e-54 and 7e-51, respectively). The scores for these two hits, however, are both lower than all four of the BLAST hits for *Msx2* against zebrafish, consistent with asymmetric gene evolution rates. The reciprocal best hit method does not have the power to make the proper assignment in such cases with radically different rates of evolutionary divergence.

Rate asymmetry becomes problematic when comparing genomes that are highly diverged. The most constructive approach to this problem is to increase taxon sampling for RBH analyses. For example, a plausible approach to solve this problem would be to compare zebrafish MSX genes first to a more closely related but non-duplicated rayfin fish, such as spotted gar or bowfin (Amores *et al*., 2011; Inoue *et al*., 2003, 2005), and then to compare the genomes of these fish to the mouse and human genomes. Another approach, which we discuss in this paper, is to use the conserved synteny of neighboring genes to aid in making the proper assignments.

5. The Problem of Lineage-Specific Gene Loss

The *Msx* case illustrates not only the problem of assigning orthologies in cases of heterogeneous rates of gene evolution, but also what can happen in cases of gene loss. If full genome data were available for human but not for mouse, then BLAST

analysis would have erroneously assigned *msxc* as a co-ortholog of human *MSX2* rather than as the ortholog of a human pseudogene paralog of *MSX2*.

6. The Synteny Pipeline

The analysis of conserved syntenies can help resolve cases like *msx* where phylogenetic trees and RBH fail. In the Synteny Database, output from the mRBH Analysis Pipeline feeds into the Synteny Pipeline (Catchen *et al*., 2009). The Synteny Pipeline can cluster paralogous and orthologous genes into regions of conserved synteny by employing a sliding window strategy. The sliding window strategy identifies chromosome segments within the primary genome and between the primary genome and user-selected outgroup genomes that have been conserved since the last WGD event. The sliding window allows for small-scale changes in gene order, gene orientation, and gene loss within conserved regions. The Synteny Database presents results to the user as a searchable, web-based database of conserved syntenic clusters. The system allows for the analysis of fully or partially assembled genomes (Bridgham *et al*., 2008), and is optimized for the investigation of individual gene families in multiple lineages. The Synteny Database reveals chromosome inversions and translocations and can identify cases of ohnologs gone missing. Below, we describe the implementation of the Synteny Database and present a case study to demonstrate its utility (Canestro *et al*., 2009; Catchen *et al*., 2009).

7. The Synteny Pipeline Algorithm

Given a pair of paralogous index genes in the primary genome, one in chromosome region **A** and the other in chromosome region **B**, we want to locate additional pairs of paralogs in which one member of the pair lies near the index paralog in region **A** and the other member of the pair is near the index paralog in region **B** (Fig. 2A). We define the chromosome neighborhood by placing a pair of *windows* of a user-selected size around the index paralogs, where the size of the window is measured in numbers of contiguous genes (Fig. 2A). The algorithm begins by comparing Chromosome 1 of the primary genome (region **A)**, to Chromosome 2 (region **B)**. The system starts with the first gene of region **A** and checks to see if it has a paralog in region **B**. If it does not, the system slides the window directly to the second gene in region **A**. Alternatively, if the first gene *does* have a paralog in region **B**, the system marks the start of a conserved syntenic cluster (Fig. 2A), and then checks the second gene in region **A** to see if it has a paralog in region **B**. If the second gene has no paralog in region **B**, then the algorithm moves to the next gene within the window on region **A** and continues searching. If the system identifies an additional pair of paralogs, one within the window in region **A** and one within the window in region **B**, it marks this as a conserved synteny (Fig. 2B) and slides the window forward; the new paralogs in the synteny cluster mark the window start. If instead, the search reaches the tail of either window without finding another pair of paralogous genes, then the pipeline closes the cluster and records it (Fig. 2C). The system

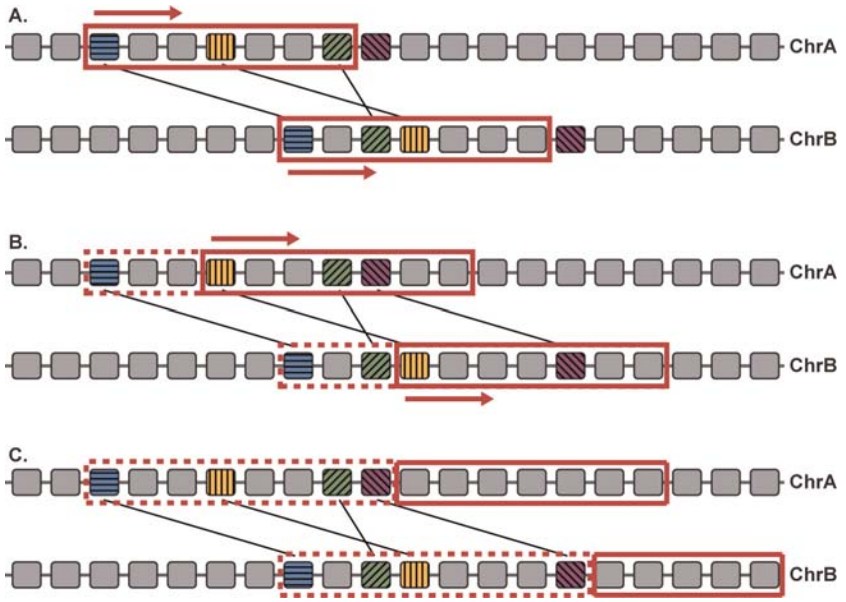


Fig. 2 Sliding window analysis. A. The system searches for modified best reciprocal blasts starting at the first gene at the left end of chromosome A, moves into region A (RegA), and then marks the first ortholog and establishes a window consisting of a set number of genes. B. After making comparisons from RegA to RegB for the set window size, the system slides the window to the next gene and searches, again marking orthologs. C. When the window finally slides to a region that finds no orthologs, the cluster of conserved synteny is considered closed. (See color plate.)

resets the position of the first window to the next gene on chromosome **A** that was not part of the last syntenic cluster and restarts the search. The analysis pipeline continues this process until it has examined all paralogous genes on chromosomes 1 and 2, then compares chromosomes 1 and 3, and so on.

To be certain that the system has not missed any conserved syntenic areas due to changes in gene order arising from inversions (e.g., paralogous genes are ordered left to right on one chromosome segment and right to left on the other segment), the pipeline restarts the search and runs the two windows in opposite directions, again recording found clusters. The software continues this analysis on every pair of chromosomes in the primary genome – comparing the first and third chromosomes, the first and fourth chromosomes, and so on. Finally, this analysis is repeated between the primary genome and the outgroup genome to derive a genome-wide representation of paralogons.

The sliding window algorithm can detect three types of architectural features in the genome. First, it can detect regions that represent simple conserved synteny, for example, in (Fig. 3A), the green horizontally hatched genes in region A are all paralogous to the green horizontally hatched genes in region B. As the algorithm searches forward in the sliding window, it detects each additional pair of paralogous

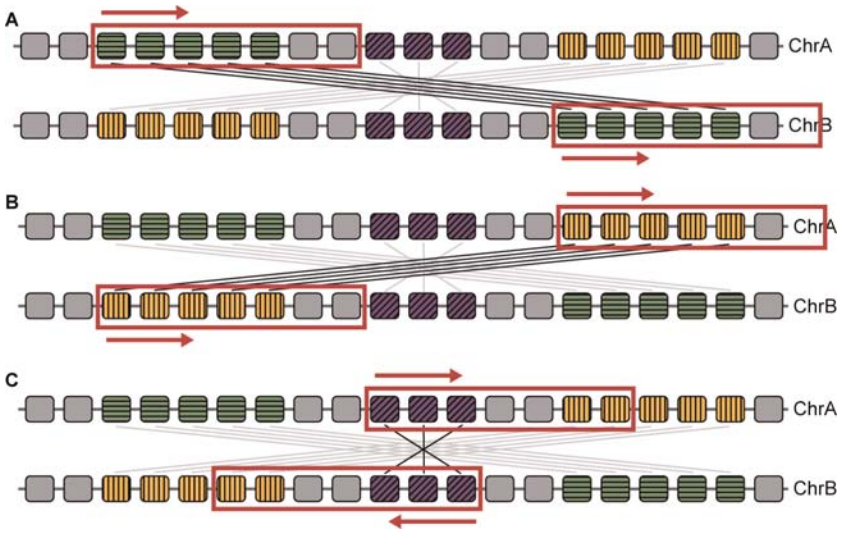


Fig. 3 Detecting rearrangements. A. The system detects the green/horizontally hatched region of conserved synteny as the windows slide in the same direction. B. The system also detects the syntenic group of yellow/vertically hatched genes but initially does not include them in the conserved segment with the green/horizontally hatched genes because they are transposed on RegB relative to RegA. C. Moving the sliding windows in opposite directions permits the combination of the several segments of conserved syntenies. (See color plate.)

green horizontally hatched genes and moves the window forward. When the window reaches the first yellow vertically hatched gene, however, it will not add it to the cluster because the corresponding genes on chromosome B fall before the start of the cluster – their positions have been inverted. The system accommodates this situation by closing the cluster and then resetting the window to the first yellow gene on region A (Fig. 3B). Now, the system places the corresponding window at the first yellow vertically hatched gene on chromosome B, thus detecting this chromosome rearrangement. The green/horizontal and yellow/vertical sets of paralogs provide two distinct clusters when the algorithm has completed examining all paralogs in regions A and B. The third type of genomic feature the algorithm can detect is an inversion (Fig. 3C). The algorithm detects these clusters by running the two windows in opposite directions on the two chromosomes.

IV. Using the Synteny Database: A Case Study of Nerve Growth Factor Receptors

A. The Nerve Growth Factor Receptor (NGFR) Gene Family

The nerve growth factor receptor (Ngfr), alias p75 NT receptor, belongs to the tumor necrosis factor transmembrane receptor superfamily. Ngfr binds to nerve

growth factor (Ngf) and other neurotrophins, proteins that stimulate nerve growth and regulate differentiation of sympathetic neurons and some sensory neurons (Micera *et al*., 2007). By interacting with numerous co-receptors and activating several signaling cascades, Ngfr acts in multiple context-dependent biological functions, including neuronal cell death and survival, neurite outgrowth, and neuronal differentiation. Ngfr signaling is thought to play a role in several human neurodegenerative conditions including Alzheimer's disease (for reviews see Chao (2003), Nykjaer *et al*. (2005), and Underwood and Coulson (2008).

Most vertebrate genomes contain the *Ngfr* gene as well as several related genes, such as *neurotrophin receptor homolog 1* (*Nrh1*) in Xenopus (Bromley *et al*., 2004) and *neurotrophin receptor like death domain protein* (*Nradd*, also known as *Nrh2*) in mammals (Kanning *et al*., 2003; Murray *et al*., 2004; Wang *et al*., 2003); *Nradd* (*Nrh2*) encodes a severely truncated extracellular protein domain in rodents but has become a pseudogene (*NRADDP*) in human.

Zebrafish has become increasingly popular as a model organism for the study of human neurodegenerative diseases, adult neurogenesis and neuronal regeneration (see e.g., Kabashi *et al*. (2010), and Kaslin *et al*. (2008)). The teleost *ngfr* gene family is poorly characterized at present and expression data are available for only one of several zebrafish *ngfr*-related genes (Brosamle and Halpern, 2009). Thus, for connectivity of the fish model to humans and other tetrapods, it is essential to identify orthologs of tetrapod *Ngfr* gene family members in zebrafish and other teleosts unequivocally.

The zebrafish genome (Zv9) includes four *ngfr*-related genes: *ngfr* located on zebrafish chromosome 16 (Dre16); *ngfra* on Dre3; *ngfrl* on Dre12; and ENSDARG00000089223, which is currently not listed in ZFIN and is likely to have arisen in a tandem duplication event because it resides adjacent to *ngfrl* and is transcribed in the same orientation. Other teleost genomes (stickleback, medaka, fugu, and the green spotted pufferfish) show a similar situation, with the exception that *ngfrl* lacks a neighbor related to the zebrafish ENSDARG00000089223 gene; thus, the *ngrfl*-ENSDARG00000089223 tandem duplication event occurred in the zebrafish lineage after it diverged from the percomorph lineage.

A phylogenetic tree rooted on related members of the human TNFR superfamily shows that the vertebrate Ngfr protein family consists of two major clades (Fig. 4A). One clade includes tetrapod Ngfr and two paralogs in teleosts (Ngfra and Ngfrl). The second clade includes Nradd/Nrh2 in mammals, Nrh1 in non-mammalian tetrapods, and the third Ngfr protein in teleosts. The following discussion illustrates the use of the Synteny Database (Catchen *et al*., 2009) to investigate relationships and origins of vertebrate *Ngfr* family members in detail with particular focus on understanding the origin of the zebrafish genes and the identification of their human orthologs.

B. How are Zebrafish and Teleost NGFR Genes Related to Other Vertebrate NGFR Genes?

In the Synteny Database (http://teleost.cs.uoregon.edu/synteny_db/), we select "*Danio rerio*" as source and "*Homo sapiens*" as outgroup, type in "ngfr" as the "gene to search for", and chose a sliding window size of 100 genes. The Synteny

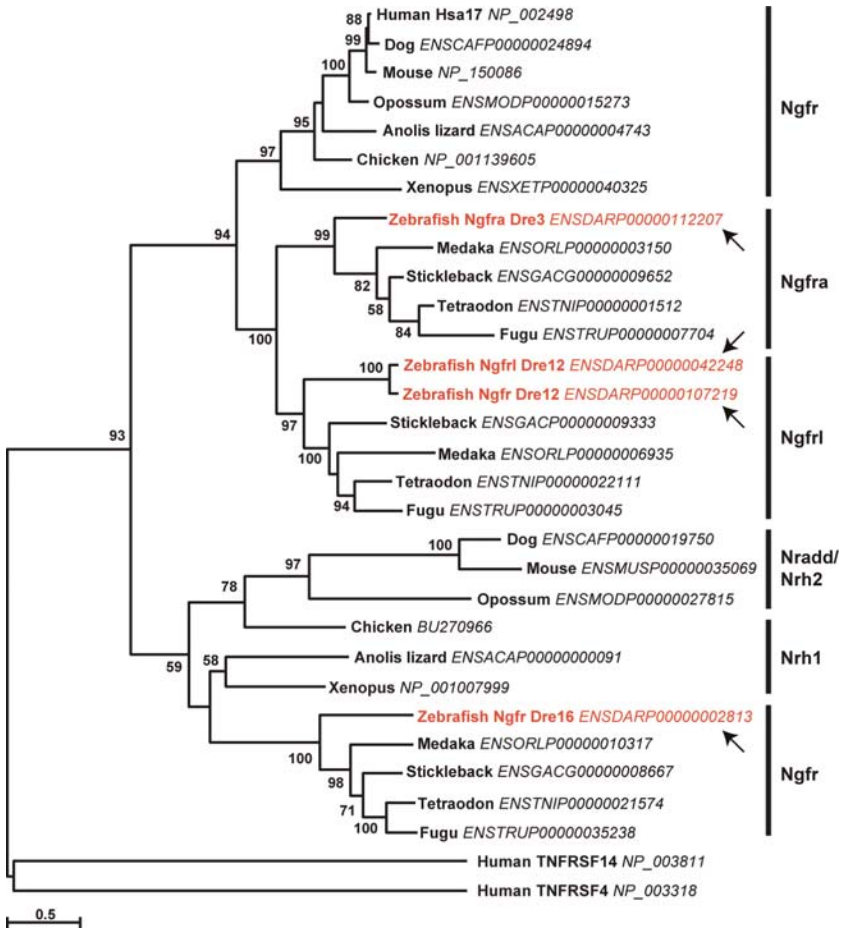


Fig. 4 Maximum likelihood phylogeny of vertebrate Ngfr-related proteins. Zebrafish has four different proteins (arrows). Clade designation includes ENSEMBL or GenBank accession numbers as well as chromosomal locations of zebrafish (Dre) and human (Hsa) genes. Current nomenclature appears to the right of the diagram with teleost proteins annotated according to ZFIN nomenclature. The tree is rooted on human TNFRSF4 and TNFRSF4 proteins. Bootstrap support values above 50 are shown (100 pseudoreplicates). (See color plate.)

Database returns, among other items, three orthologous pairwise clusters that compare the human *NGFR* region on Hsa17 to regions containing *ngfr*-related genes in the zebrafish genome, including parts of Dre3 (NGFR: #148114), Dre12 (NGFR: #145024), and Dre16 (NGFR: #144872) (Fig. 5). The human *NGFR* region on Hsa17 shares extensive conserved synteny with the zebrafish *ngfra* region on Dre3 that includes 205 gene pairs (Fig. 5A) and with the *ngfrl-ngfr* tandem duplication region on Dre12 that includes 150 gene pairs (Fig. 5B). In contrast, the orthologous pairwise cluster between Hsa17 and Dre16 shows a weakly conserved synteny that involves

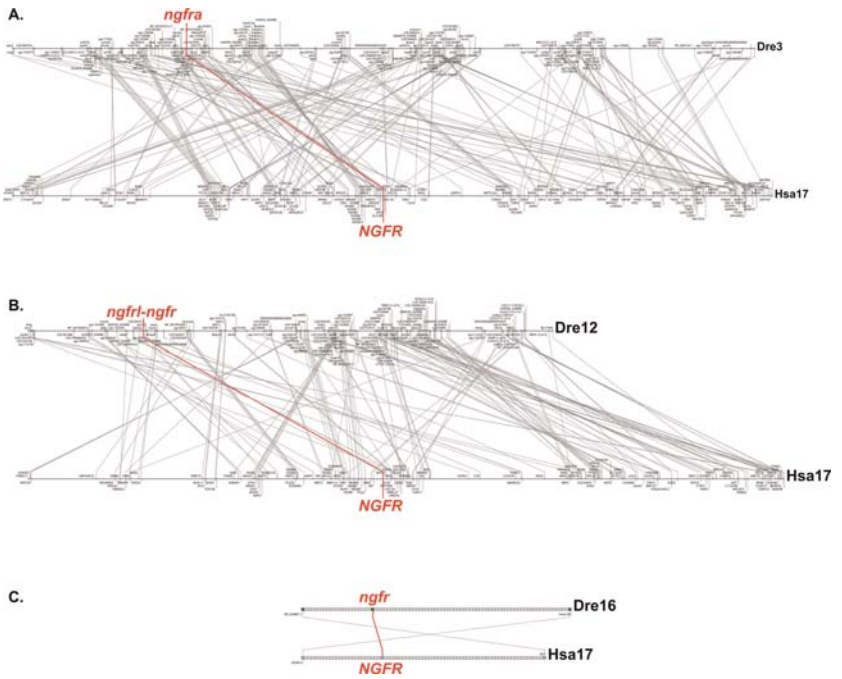


Fig. 5 *Ngfr* orthologous pairwise clusters from the Synteny Database (zebrafish vs. human). A. The portion of human chromosome 17 (Hsa17) surrounding *NGFR* shares extensive conserved synteny (205 gene pairs) with the *ngfra* region on zebrafish chromosomes Dre3. B. The *NGFR* region of Hsa17 also shares substantial conserved synteny with *ngfrl-ngfr* region of zebrafish on Dre12 (150 gene pairs). C. Conserved synteny between Hsa17 and the third zebrafish *ngfr* region, the one on Dre16, consists of only three gene pairs, including *ngfr* itself. (See color plate.)

just three rather distantly located gene pairs (Fig. 5C). This result parallels the finding from the phylogenetic analysis that showed a topology consistent with the interpretation that *ngfra* and *ngrfl-ngfr* are co-orthologs of the human *NGFR* gene (Figs. 4, 5A, 5B). Furthermore, the large extended shared conserved syntenies in Fig. 5A,B supports the hypothesis that the *ngfra* and *ngfrl* paralogs arose during a large-scale duplication event, i.e., during the TGD. In fact, Dre3 and Dre12 are classic examples for paralogous zebrafish chromosomes being derived from a common ancestral teleost protochromosome and duplicated during the TGD (they contain, for example, the *hoxbb* and *hoxba* clusters, respectively) (Amores *et al*., 1998; Postlethwait *et al*., 1998; Woods *et al*., 2005). Evidence of relatedness from the tree and from the conserved syntenies suggests a revision of zebrafish gene nomenclature to reflect genomic relationships better. Because "a" and "b" suffixes commonly distinguish TGD co-orthologs, the correct name for *ngfra* on Dre3 should be *ngfr1a* and proper nomenclature for the two genes on Dre12 should become *ngfr1ba* and *ngfr1bb*.

But how is the zebrafish Dre16 gene *ngfr*, which we now call *ngfr2*, related to the human Hsa17 gene *NGFR*? To answer this question, we turn to the dotplot capabilities of the Synteny Database. This implementation starts with the first gene on an index chromosome, here Hsa17, and then looks for that gene's zebrafish ortholog or co-orthologs. When it finds an ortholog, it puts a point on the zebrafish chromosome directly vertical to the human gene. The system then moves to the second gene on Hsa17, places another point on the zebrafish chromosome that contains the ortholog, and then continues marking zebrafish orthologs as it moves down the human chromosome. Note that the query chromosome, in this case Hsa17, specifies gene order, not the order of genes on subject chromosomes, in this case zebrafish chromosomes 1 to 25. Users can select genes to highlight on the output graph, here *NGFR*. In the output, which can be downloaded as a PDF, paralogons (paralogous chromosome segments) appear as rows of clustered dots on the same subject chromosome, in this case, most notably Dre3 and Dre12 (Fig. 6). Two paralogous regions in the zebrafish genome corresponding to a single human chromosome region provide a signature of the TGD. As illustrated by the orthologous pairwise cluster in Fig. 5B,C as well as by the dotplot representation in Fig. 6A, syntenic relationships between the zebrafish

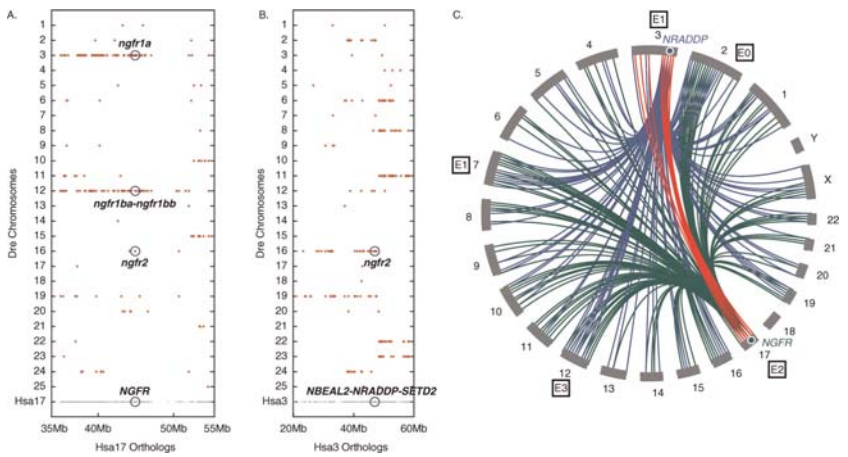


Fig. 6 Synteny dotplots from the Synteny Database. A. The *ngfr* regions on Dre3 and Dre12 share extensive conserved synteny with the human *NGFR* region on Hsa17. The *ngfr* region on Dre16, in contrast, does not share extensive conserved synteny with Hsa3. B. The *ngfr* region on Dre16 shares extensive synteny with the human *NRADD/NRH2* pseudogene region on Hsa3. Note that Dre19 contains a region of extensive conserved synteny with Hsa3 but does not contain an *ngfr*-related gene. C. Circle plot of the human genome derived from the Synteny Database. Using the amphioxus genome as outgroup, paralogies of 40 Mb surrounding the *NGFR* region on Hsa17 (green) and the *NRADD* pseudogene region on Hsa3 (blue) are plotted onto the entire human karyotype. Paralogies directly connecting Hsa17 and Hsa3 are shown in red. Hsa2, Hsa7 and Hsa12 contain additional regions of substantial paralogy. This is in agreement with a previous reconstruction of the ancestral vertebrate karyotype, which predicted that each of these human chromosomes is derived from the protochromosome E (Nakatani *et al*., 2007). Post-R2 genome duplication chromosome nomenclature given in boxes: E0-E3. (See color plate.)

ngfr2 region (Dre16) and the human *NGFR* region (Hsa17) are basically restricted to the *Ngfr* gene itself, suggesting that the chromosome segments bearing these genes (Hsa17 and Dre16) may not be closely related.

To identify the region of the human genome that contains a chromosome segment related to the *ngfr2* region in zebrafish, note that the zebrafish *ngfr2* region shares extensive conserved syntenies to human chromosome 3 (Hsa3) (Fig. 6B): On Dre16, *ngfr2* is flanked by *nbeal2* and *setd2*. Significantly, in the human genome, *NBEAL2* and *SETD2* flank the *NRADDP* pseudogene on Hsa3. This region of Hsa3 also shares many conserved syntenies with Dre19 (Fig. 6B). Large parts of Dre19 are known to be paralogous to Dre16 (they contain the *hoxaa* and *hoxab* clusters, respectively (Amores *et al*., 1998)) and they were generated during the TGD. This evidence suggests that the teleost duplicate of *ngfr2* would have originally been on the ancestor of Dre19 and was lost after the TGD from zebrafish and other sequenced teleost genomes.

A comparison of the human and mouse genomes is also instructive. Genes in the region of mouse (*Mus musculus*) chromosome Mmu9 have human orthologs in same gene order as Hsa3: *Nbeal2-Nradd*(*Nrh2*)-*Setd2*. Moreover, the *Nrh1* gene in chicken and in Xenopus are both located next to an ortholog of *Setd2*. We therefore conclude that *ngfr2* in teleosts, *Nradd/Nrh2/NRADDP* in mammals, and *Nrh1* in non-mammalian tetrapods are all orthologous, despite their varied names. Collectively, calling these genes *Ngfr2* rather than the current set of haphazard names would reflect more accurately their orthologous nature and would suggest a paralogous relationship with *Ngfr*, which probably should become *Ngfr1*.

The circle plot option of the Synteny Database provides a dramatic visualization of paralogies within a genome or orthologies between genomes (Fig. 6C). Circle plots link paralogs or orthologs with colored lines from their positions on selected chromosomes to their corresponding elements on other chromosomes arrayed about the circumference of the circle. The circle plot option can search for conserved synteny between the two *Ngfr* regions as well as relationships to other regions in the human genome. For example, Fig. 6C shows that the *NGFR1* (Hsa17) and *NGFR2* (Hsa3) regions share many conserved syntenies (red lines). In addition, these two regions share conserved syntenies with other chromosomes, especially Hsa2, Hsa7, and Hsa12. This suggests that the *NGFR* regions evolved by duplication of larger genomic blocks, probably during the vertebrate R1 and R2 genome duplications (note that Hsa2, 7, 12, and 17 contain *HOXD, HOXA, HOXC*, and *HOXB*, respectively). In fact, conserved synteny analysis has suggested that all the aforementioned regions in the human genome are derived from the ancestral vertebrate pre-R1/R2 protochromosome E (Nakatani *et al*., 2007; Sundstrom *et al*., 2008).

C. A Model for Vertebrate NGFR Gene Family Evolution

Taking results from the Synteny Database into account, we propose a new model for the evolution of vertebrate *Ngfr* genes (Fig. 7). According to the model, a single proto-*Ngfr* gene (*Ngfr1/2/3/4*) present on vertebrate protochromosome E duplicated

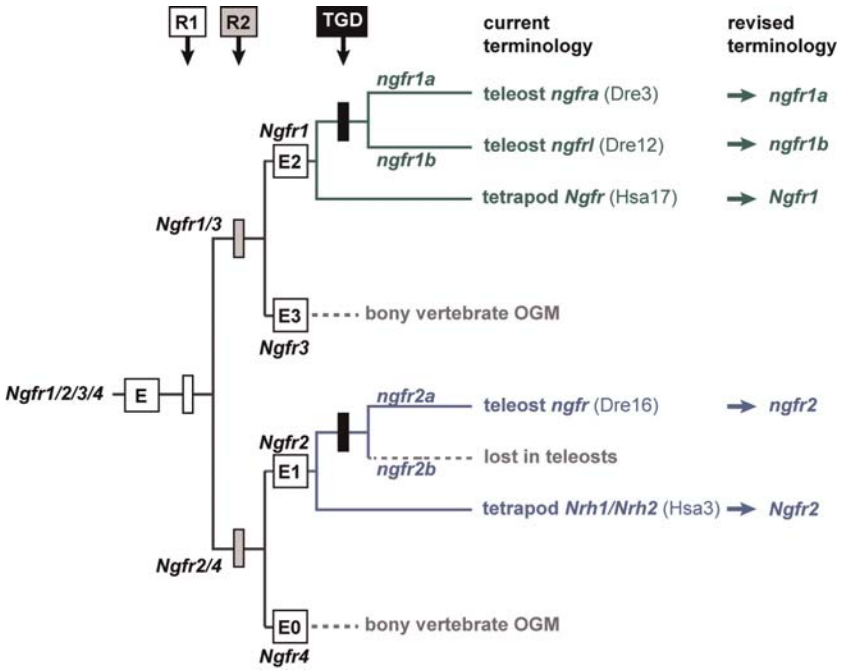


Fig. 7 Model and revised nomenclature suggestion for the evolution of vertebrate *Ngfr* genes. After three rounds of genome duplication (R1, R2, TGD), several teleost fish have three *ngfr* genes. Tetrapod genomes often contain two *Ngfr* ohnologs (*Ngfr1* and *Ngfr2*) that evolved along with other genes on the descendents of vertebrate protochromosome E. The mammalian *Ngfr2* gene suffered extensive remodeling in its extracellular domain, evolving into the unusual *Nradd/Nrh2* configuration. Abbreviation: OGM ohnolog gone missing. (See color plate.)

during the R1 genome duplication and gave rise to the ancestral genes *Ngfr1/3* and *Ngfr2/4*, the progenitors of *Ngfr1* and *Ngfr2*. The R2 genome duplication event then produced four *Ngfr* genes (*Ngfr1, Ngfr2, Ngfr3*, and *Ngfr4*), followed by the loss of *Ngfr3* from the *Ngfr1/3* clade and *Ngfr4* from the *Ngfr2/4* clade. Thus, the extant vertebrate *Ngfr1* and *Ngfr2* genes are ohnologs that share a common ancestry dating back to the R1 genome duplication, whereas the hypothesized *Ngfr3* and *Ngfr4* genes became ohnologs gone missing, with their old chromosome neighborhoods (Hsa2 and Hsa12) still detectable by their possession of numerous pairs of paralogous genes. *Ngfr3* and *Ngrf4* may still be present in more basally diverging vertebrate lineages, such as jawless and cartilaginous fish. This new nomenclature, informed by conserved synteny analysis, is sufficiently robust to integrate these genes if they are discovered.

After the divergence of the rayfin and lobefin lineage, the TGD resulted in duplication of both *ngfr1* and *ngfr2*, giving rise to *ngfr1a* and *ngfr1b*, which both persist in zebrafish, and to *ngfr2a* and *ngfr2b*, the latter of which was lost in the

teleost lineage. In the zebrafish lineage, *ngfr1b* again became duplicated, but now in tandem (not diagrammed in Fig. 7). Evolution in the mammalian lineage resulted in restructuring of *Ngfr2* by truncation of the extracellular domain to become recognized as *Nradd* in many mammals, and in the further degradation of *NGFR2* in the human genome, which finally became the pseudogene *NRADDP*.

This case study shows how to use the Synteny Database to help discover the origin of zebrafish genes and reveal their true relationships as orthologs and paralogs to genes in humans and other taxa. In this case, conserved synteny analysis shows the advantage of considering neuronal cell signaling and other biological processes in zebrafish in a rigorous phylogenetic framework. For example, studies in zebrafish will help to infer the function of the *Ngfr2* gene in the mammalian ancestor before the remodeling of the extracellular domain. Murine Ngfr2 is thought to be involved in the regulation of Ngf signaling by intracellular interaction with Ngfr and its co-receptors (Kim and Hempstead, 2009; Murray *et al*., 2004; Wong *et al*., 2008). Testing zebrafish Ngfr2 protein for such interactions will help to clarify whether or not these interactions are attributable to the mammal-specific restructuring of Ngfr2.

V. Zebrafish, Conserved Syntenies, and the Teleost Genome Duplication

Teleost genomes appear to be rearranged with respect to mammalian chromosomes (Jaillon *et al*., 2004; Nakatani *et al*., 2007; Postlethwait *et al*., 1998). One hypothesis to account for this observation is that the TGD facilitated chromosome rearrangements, which disrupt conserved syntenies; this idea, however, is controversial (Comai, 2005; Hufton *et al*., 2008; Semon and Wolfe, 2007). An alternative hypothesis is that chromosome restructuring occurred either in the long lineage leading to humans after the divergence of rayfin fish (including zebrafish) from lobefin fish (including humans). To determine whether the chromosomal rearrangements that distinguish human and zebrafish genomes occurred mostly in the zebrafish lineage after divergence from percomorphs, primarily after the TGD in the teleost lineage, or mainly before the TGD either in the rayfin stem or in the lobefin lineage, we need to examine a genome representing a lineage that diverged from the zebrafish lineage before the TGD but after the divergence of rayfin fish and lobefin fish. Although no full genome sequence represents the needed taxon, a recent genetic map of spotted gar (*Lepisosteus oculatus*) (Amores et al., 2011) contains nearly 1000 coding markers, which is sufficient to test the two hypotheses. A modified RBH-based analysis of conserved syntenies shared between human or gar with the genomes of zebrafish or stickleback revealed unexpected conservation between human and gar compared to the conservations between either human or gar and the teleosts (Fig. 8). These results are consistent with the hypothesis that chromosome rearrangements and the loss of ancestral syntenies accelerated after the TGD. This result supports the hypothesis that WGD can facilitate syntenic rearrangements.

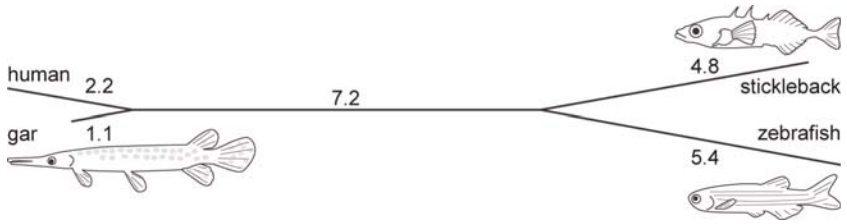


Fig. 8 Comparison of conserved syntenies between zebrafish and stickleback with respect to gar and human. Branch lengths are proportional to conserved syntenies, which were estimated from the average number of gar (or human) chromosomes that contain orthologs of genes on a teleost chromosome divided by the number of chromosomes in gar (or human) normalized to divergence times in millions of years (zebrafish/stickleback, 220 mya (million years ago); teleosts and gar, 340 mya; rayfin and lobefin fish, 440 mya (Inoue *et al*., 2005)). Branch lengths between gar and human are much shorter than those connecting either gar or human to teleost genomes. Branch lengths were calculated as the average number of chromosomes of one species that contain orthologs of a chromosome of the compared species per 100 million years of divergence time.

The acceleration of synteny-breaking rearrangements by genome duplication could occur by any of several possible mechanisms. Duplicated homologous coding elements can provide numerous substrates for illegitimate recombination between homeologous (paralogous) chromosomes; such recombination events would stimulate chromosome translocations and hence would disturb conserved syntenies (Comai, 2005). In addition, after a WGD – and especially in the case of autotetraploidy – natural selection would favor karyotypes in which one member of a pair of homeologs was rearranged relative to the other because such rearrangements would reduce the pairing of the four homeologous chromosomes in meiosis, thereby hastening rediploidization, decreasing the production of aneuploid offspring, and improving fitness (Comai, 2005).

VI. Summary

This chapter discussed principles behind the analysis of conserved syntenies to help identify gene histories and hence to identify orthologies and paralogies between zebrafish and other organisms. Discussion focused on describing algorithms underlying the Synteny Database, and then utilized tools provided at the Synteny Database web site to improve our understanding of the evolution of the Ngf receptor gene family. Finally, the work showed that chromosome rearrangements in zebrafish and other teleosts likely accelerated in the few million years after the TGD but before the divergence of zebrafish and stickleback genomes. Application of the analysis of conserved syntenies to enigmatic zebrafish genes can help to improve the connectivity of zebrafish and human genomes and thus can facilitate the translation of biomedical information from zebrafish models of human disease to human biology.

Acknowledgments

For support we thank NIH/NCRR grant R01RR020833 to J.H.P., an NIH NRSA Ruth L. Kirschstein postdoctoral fellowship 1F32GM095213-01 to J.M.C., and the Alexander von Humboldt Foundation (IB).

References

Akimenko, M. -A., Johnson, S. L., Westerfield, M., and Ekker, M. (1995). Differential induction of four *msx* homeobox genes during fin development and regeneration in zebrafish. *Development* **121**, 347–357.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711–1714.

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., and Postlethwait, J. H. (2011). Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**, 1–10.

Berglund, A. C., Sjolund, E., Ostlund, G., and Sonnhammer, E. L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.* **36**, D263–D266.

Bridgham, J. T., Brown, J. E., Rodriguez-Mari, A., Catchen, J. M., and Thornton, J. W. (2008). Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet.* **4**, e1000191.

Bromley, E., Knapp, D., Wardle, F. C., Sun, B. I., Collins-Racie, L., LaVallie, E., Smith, J. C., and Sive, H. L. (2004). Identification and characterisation of the posteriorly-expressed Xenopus neurotrophin receptor homolog genes fullback and fullback-like. *Gene Expr. Patterns* **5**, 135–140.

Brosamle, C., and Halpern, M. E. (2009). Nogo-Nogo receptor signalling in PNS axon outgrowth and pathfinding. *Mol Cell Neurosci.* **40**, 401–409.

Byrne, K. P., and Wolfe, K. H. (2007). Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics* **175**, 1341–1350.

Canestro, C., Catchen, J. M., Rodriguez-Mari, A., Yokoi, H., and Postlethwait, J. H. (2009). Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet.* **5**, e1000496.

Catchen, J. M., Conery, J. S., and Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome Res.* **19**, 1497–1505.

Chao, M. V. (2003). Neurotrophins and their receptors: a convergence point for many signalling pathways. *Nat. Rev. Neurosci.* **4**, 299–309.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846.

Courcelle, E., Beausse, Y., Letort, S., Stahl, O., Fremez, R., Ngom-Bru, C., Gouzy, J., and Faraut, T. (2008). Narcisse: a mirror view of conserved syntenies. *Nucleic Acids Res.* **36**, D485–D490.

Dehal, P., and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314.

Derrien, T., Andre, C., Galibert, F., and Hitte, C. (2007). AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps. *Bioinformatics* **23**, 498–499.

Des Marais, D. L., and Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765.

Dong, X., Fredman, D., and Lenhard, B. (2009). Synorth: exploring the evolution of synteny and long-range regulatory interactions in vertebrate genomes. *Genome Biol.* **10**, R86.

Donovan, A., Brownlie, A., Zhou, Y., Shepard, J., Pratt, S. J., Moynihan, J., Paw, B. H., Drejer, A Barut, B., Zapata, A., Law, T. C., Brugnara, C., Lux, S. E., Pinkus, G. S., Pinkus, J. L., Kingsley, P. D., Palis, J., Fleming, M. D., Andrews, N. C., and Zon, L. I. (2000). Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* **403**, 776–781.

Eberhart, J. K., He, X., Swartz, M. E., Yan, Y. L., Song, H., Boling, T. C., Kunerth, A. K., Walker, M. B., Kimmel, C. B., and Postlethwait, J. H. (2008). MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis. *Nat. Genet.* **40**, 290–298.

Ebermann, I., Phillips, J. B., Liebau, M. C., Koenekoop, R. K., Schermer, B., Lopez, I., Schafer, E Roux, A. F., Dafinger, C., Bernd, A., Zrenner, E., Claustres, M., Blanco, B., Nurnberg, G., Nurnberg, P., Ruland, R., Westerfield, M., Benzing, T., and Bolz, H. J. (2010). PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J. Clin Invest.* **120**, 1812–1823.

Ekker, M., Akimenko, M., Allende, M., Smith, R., Drouin, G., Langille, R., Weinberg, E., and Westerfield, M. (1997). Relationships among *msx* gene structure and function in zebrafish and other vertebrates. *Mol. Biol. Evol.* **14**, 1008–1022.

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584.

Fares, M. A., Byrne, K. P., and Wolfe, K. H. (2006). Rate asymmetry after genome duplication causes substantial long-branch attraction artifacts in the phylogeny of Saccharomyces species. *Mol. Biol. Evol.* **23**, 245–253.

Finnerty, J. R., Mazza, M. E., and Jezewski, P. A. (2009). Domain duplication, divergence, and loss events in vertebrate Msx paralogs reveal phylogenomically informed disease markers. *BMC Evol. Biol.* **9**, 18.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.

Force, A., Amores, A., and Postlethwait, J. H. (2002). *Hox* cluster organization in the jawless vertebrate *Petromyzon marinus*. *J. Exp. Zool.* **294**, 30–46.

Fu, B., Chen, M., Zou, M., Long, M., and He, S. (2010). The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genomics* **11**, 657.

Garcia-Fernandez, J., and Holland, P. W. (1994). Archetypal organization of the amphioxus Hox gene cluster. *Nature* **370**, 563–566.

Guo, B., Gan, X., and He, S. (2010). Hox genes of the Japanese eel Anguilla japonica and Hox cluster evolution in teleosts. *J. Exp. Zool B Mol. Dev. Evol.* **314**, 135–147.

Hahn, M. (2007). Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. *Genome Biol.* **8**, R141.

Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10915–10919.

Hufton, A. L., Groth, D., Vingron, M., Lehrach, H., Poustka, A. J., and Panopoulou, G. (2008). Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. *Genome Res.* **18**, 1582–1591.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. R Soc. Lond. B Biol. Sci.* **256**, 119–124.

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97–108.

Inoue, J. G., Miya, M., Tsukamoto, K., and Nishida, M. (2003). Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". *Mol Phylogenet Evol.* **26**, 110–120.

Inoue, J. G., Miya, M., Venkatesh, B., and Nishida, M. (2005). The mitochondrial genome of Indonesian coelacanth Latimeria menadoensis (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* **349**, 227–235.

Jabs, E. W., Muller, U., Li, X., Ma, L., Luo, W., Haworth, I. S., Klisak, I., Sparkes, R., Warman, M. L., and Mulliken, J. B., *et al*. (1993). A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell* **75**, 443–450.

Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B.,

Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigo, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., and Roest Crollius, H. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957.

Jarinova, O., Hatch, G., Poitras, L., Prudhomme, C., Grzyb, M., Aubin, J., Berube-Simard, F. A Jeannotte, L., and Ekker, M. (2008). Functional resolution of duplicated hoxb5 genes in teleosts. *Development* **135**, 3543–3553.

Jezewski, P. A., Vieira, A. R., Nishimura, C., Ludwig, B., Johnson, M., O'Brien, S. E., Daack-Hirsch, S., Schultz, R. E., Weber, A., Nepomucena, B., Romitti, P. A., Christensen, K., Orioli, I. M., Castilla, E. E., Machida, J., Natsume, N., and Murray, J. C. (2003). Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J. Med. Genet.* **40**, 399–407.

Jovelin, R., He, X., Amores, A., Yan, Y. L., Shi, R., Qin, B., Roe, B., Cresko, W. A., and Postlethwait, J. H. (2007). Duplication and divergence of fgf8 functions in teleost development and evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 730–743.

Kabashi, E., Champagne, N., Brustein, E., and Drapeau, P. (2010). In the swim of things: recent insights to neurogenetic disorders from zebrafish. *Trends Genet.* **26**, 373–381.

Kanning, K. C., Hudson, M., Amieux, P. S., Wiley, J. C., Bothwell, M., and Schecterson, L. C. (2003). Proteolytic processing of the p75 neurotrophin receptor and two homologs generates C-terminal fragments with signaling capability. *J. Neurosci.* **23**, 5425–5436.

Kaslin, J., Ganz, J., and Brand, M. (2008). Proliferation, neurogenesis and regeneration in the non-mammalian vertebrate brain. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **363**, 101–122.

Kikuta, H., Fredman, D., Rinkwitz, S., Lenhard, B., and Becker, T. S. (2007). Retroviral enhancer detection insertions in zebrafish combined with comparative genomics reveal genomic regulatory blocks – a fundamental feature of vertebrate genomes. *Genome Biol.* **8**(Suppl 1), S4.

Kim, T., and Hempstead, B. L. (2009). NRH2 is a trafficking switch to regulate sortilin localization and permit proneurotrophin-induced cell death. *EMBO J.* **28**, 1612–1623.

Kuraku, S., Meyer, A., and Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* **26**, 47–59.

Kurosawa, G., Takamatsu, N., Takahashi, M., Sumitomo, M., Sanaka, E., Yamada, K., Nishii, K., Matsuda, M., Asakawa, S., Ishiguro, H., Miura, K., Kurosawa, Y., Shimizu, N., Kohara, Y., and Hori, H. (2006). Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* **370**, 75–82.

Lamason, R., Mohideen, M., Mest, J., Wong, A., Norton, H., Aros, M., Jurynec, M., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P. M., O'Donnell, D., Kittles, R., Parra, E. J., Mangini, N. J Grunwald, D. J., Shriver, M. D., Canfield, V. A., and Cheng, K. C. (2005). SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. *Science* **310**, 1782–1786.

Leong, J. S., Jantzen, S. G., von Schalburg, K. R., Cooper, G. A., Messmer, A. M., Liao, N. Y., Munro, S., Moore, R., Holt, R. A., Jones, S. J., Davidson, W. S., and Koop, B. F. (2010). Salmo salar and Esox lucius full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics* **11**, 279.

Li, L., Meng, T., Jia, Z., Zhu, G., and Shi, B. (2010). Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. *Am. J. Med. Genet. A* **152A**, 856–862.

Li, W. H., Gu, Z., Cavalcanti, A. R., and Nekrutenko, A. (2003). Detection of gene duplications and block duplications in eukaryotic genomes. *J. Struct. Funct. Genomics* **3**, 27–34.

Lynch, M., and Conery, J. (2000). The evolutionary fate and consequences of gene duplication. *Science* **290**, 1151–1155.

Micera, A., Lambiase, A., Stampachiacchiere, B., Bonini, S., and Levi-Schaffer, F. (2007). Nerve growth factor and tissue repair remodeling: trkA(NGFR) and p75(NTR), two receptors one fate. *Cytokine Growth Factor Rev.* **18**, 245–256.

Montosi, G., Donovan, A., Totaro, A., Garuti, C., Pignatti, E., Cassanelli, S., Trenor, C. C., Gasparini, P., Andrews, N. C., and Pietrangelo, A. (2001). Autosomal-dominant hemochromatosis is associated with a mutation in the ferroportin (SLC11A3) gene. *J. Clin Invest.* **108**, 619–623.

Muffato, M., Louis, A., Poisnel, C., and Crollius, H. R. (2010). Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics* **26**, 1119–1121.

Murray, S. S., Perez, P., Lee, R., Hempstead, B. L., and Chao, M. V. (2004). A novel p75 neurotrophin receptor-related protein, NRH2, regulates nerve growth factor binding to the TrkA receptor. *J. Neurosci.* **24**, 2742–2749.

Nakatani, Y., Takeda, H., Kohara, Y., and Morishita, S. (2007). Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* **17**, 1254–1265.

Nykjaer, A., Willnow, T. E., and Petersen, C. M. (2005). p75NTR—live or let die. *Curr. Opin. Neurobiol.* **15**, 49–57.

Ohno, S. (1970). *Evolution by Gene Duplication.* Springer-Verlag, New York.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y. L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **20**, 481–490.

Postlethwait, J. H., Yan, Y. -L., Gates, M., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D Singer, A., Thomson, M., Abduljabbar, T. S., Yelick, P., Beier, D., Joly, J. -S., Larhammar, D., and Talbot, W. S., *et al*. (1998). Vertebrate genome evolution and the zebrafish gene map. *Nat Genet.* **18**, 345–349.

Prohaska, S. J., and Stadler, P. F. (2004). The duplication of the Hox gene clusters in teleost fishes. *Theory Biosci.* **123**, 89–110.

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K., Benito-Gutierrez, E. L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J. J., Grigoriev, I. V., Horton, A. C., de Jong, P. J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L. A., Salamov, A. A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I. T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L. Z., Holland, P. W., Satoh, N., and Rokhsar, D. S. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071.

Remm, M., Storm, C. E., and Sonnhammer, E. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052.

Ruddle, F. H. (1972). Linkage analysis using somatic cell hybrids. *Adv. Hum. Genet.* **30**, 173–235.

Semon, M., and Wolfe, K. H. (2007). Rearrangement rate following the whole-genome duplication in teleosts. *Mol. Biol. Evol.* **24**, 860–867.

Shimeld, S. M., McKay, I. J., and Sharpe, P. T. (1996). The murine homeobox gene Msx-3 shows highly restricted expression in the developing neural tube. *Mech. Dev.* **55**, 201–210.

Sinha, A. U., and Meller, J. (2007). Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. *BMC Bioinformatics* **8**, 82.

Spring, J. (1997). Vertebrate evolution by interspecific hybridization — are we polyploid? *Fed. Eur. Biol. Soc. Lett.* **400**, 2–8.

Steinke, D., Salzburger, W., Braasch, I., and Meyer, A. (2006). Many genes in fish have species-specific asymmetric rates of molecular evolution. *BMC Genomics* **7**, 20.

Sundstrom, G., Larsson, T. A., and Larhammar, D. (2008). Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters. *BMC Evol Biol.* **8**, 254.

Taylor, J., Braasch, I., Frickey, T., Meyer, A., and Van De Peer, Y. (2003). Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**, 382–390.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. Royal Stat. Soc. B.* **63**, 411–423.

Turunen, O., Seelke, R., and Macosko, J. (2009). In silico evidence for functional specialization after genome duplication in yeast. *FEMS Yeast Res.* **9**, 16–31.

Underwood, C. K., and Coulson, E. J. (2008). The p75 neurotrophin receptor. *Int. J. Biochem. Cell Biol.* **40**, 1664–1668.

Van de Peer, Y. (2004). Computational approaches to unveiling ancient genome duplications. *Nat. Rev. Genet.* **5**, 752–763.

van Hoof, A. (2005). Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**, 1455–1461.

Volkman, H. E., Pozos, T. C., Zheng, J., Davis, J. M., Rawls, J. F., and Ramakrishnan, L. (2010). Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. *Science* **327**, 466–469.

Wall, D. P., Fraser, H. B., and Hirsh, A. E. (2003). Detecting putative orthologs. *Bioinformatics* **19**, 1710–1711.

Wang, X., Shao, Z., Zetoune, F. S., Zeidler, M. G., Gowrishankar, K., and Vincenz, C. (2003). NRADD, a novel membrane protein with a death domain involved in mediating apoptosis in response to ER stress. *Cell Death Differ.* **10**, 580–591.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54–61.

Watterson, G. A. (1983). On the time for gene silencing at duplicate loci. *Genetics* **105**, 745–766.

Winkler, C., Schafer, M., Duschl, J., Schartl, M., and Volff, J. N. (2003). Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication. *Genome Res.* **13**, 1067–1081.

Wolfe, K. (2000). Robustness — it's not where you think it is. *Nat. Genetics* **25**, 3–4.

Wong, A. W., Willingham, M., Xiao, J., Kilpatrick, T. J., and Murray, S. S. (2008). Neurotrophin receptor homolog-2 regulates nerve growth factor signaling. *J. Neurochem.* **106**, 1964–1976.

Woods, I. G., Kelly, P. D., Chu, F., Ngo-Hazelett, P., Yan, Y. L., Huang, H., Postlethwait, J. H., and Talbot, W. S. (2000). A comparative map of the zebrafish genome. *Genome Res.* **10**, 1903–1914.

Woods, I. G., Wilson, C., Friedlander, B., Chang, P., Reyes, D. K., Nix, R., Kelly, P. D., Chu, F., Postlethwait, J. H., and Talbot, W. S. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* **15**, 1307–1314.

Yampolsky, L. Y., and Bouzinier, M. A. (2010). Evolutionary patterns of amino acid substitutions in 12 Drosophila genomes. *BMC Genomics* **11**(Suppl 4), S10.

CHAPTER 16

The Zon Laboratory Guide to Positional Cloning in Zebrafish

Yi Zhou* **and Leonard I. Zon†**

*Division of Hematology/Oncology, Children's Hospital Boston, Boston, Massachusetts, USA

†Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA

METHODS IN CELL BIOLOGY, VOL 104

0091-679X/10 $35.00
DOI 10.1016/B978-0-12-374814-0.00016-1

Abstract

Zebrafish genome sequencing project has improved efficiency of positional cloning in zebrafish and traditional chromosome walking by isolating large insert genomic libraries has become a past. However, the genetic principles underlying the positional cloning still form the foundation for current chromosome walking using the genome sequence assemblies and related genomic sequence and clone information. This guide intends to summarize our accumulated experience in positional cloning using the current genomic resources and tools, and provide a practical guide to positional and/or candidate cloning of mutants of interest.

I. Introduction

Zebrafish is a vertebrate genetic system. Its forward genetics offers a powerful tool to study gene functions in vertebrate during development in embryos and adults. More targeted forward genetic screens can be designed to learn *in vivo* gene functions in different cell types and tissues. Several genome wide diploid screens have been executed using a variety of assays. At the same time, many smaller scale mutagenesis screens, either in diploid or haploid, have taken place in laboratories worldwide. Although many of these mutants with interesting phenotypes have been isolated, a large number of these mutants are still waiting to be cloned.

Mutant mapping and cloning are based on chromosome recombination events in a diploid genome, when paternal and maternal genetic materials from different (polymorphic) genetic backgrounds exchange during oogenesis and spermatogenesis. These recombination events can be followed by background specific polymorphism, represented by molecular markers throughout a genome. Recombination frequencies between any two loci on the genome are used to measure genetic distances between them in one or multiple families that possess polymorphism at these loci, monitored by recombination of polymorphic genomic markers represent these two loci. Polymorphism is defined by unique repeat length or DNA sequence of a locus in a diploid genome. Recombination frequency is proportional to their physical distances between any two loci. Two Loci from parental genomes on separate chromosomes should be randomly segregated into genomes of next generation at a 50% ratio, a Mendel segregation. In theory, any segregation rates between two loci that are smaller than 50% indicate that these two loci may locate on the same chromosome or genetic linkage group. This forms the foundation for the initial step in mapping a mutant.

The goal of positional cloning is to find the link between a biological phenotype and genomic sequence alterations that can be point mutations or sequence insertions or deletions. This is achieved by first linking known markers, genomic landmarks that have a known location on the genome, to a phenotype causing mutation. These

markers are different in nucleotide sequences. These differences can be assayed and visualized by gel electrophoresis (Simple Sequence Length Polymorphism markers (SSLP)) – measurement of the difference in number of simple sequence repeats in the same genomic location on different genetic backgrounds; Restriction Fragment Length Polymorphism (RFLP) – detection of different genomic fragmentation patterns from different genetic backgrounds using region specific probes; a differential hybridization using ASOs (Allele Specific Oligonucleotides) or SNP chip; allele-specific real-time PCR using KASPAR platform from Biosciences, or direct sequencing of DNA fragment from the polymorphic region. In the following positional cloning process, a mutated genomic region is narrowed to an interval of a few genes by closer and closer linked polymorphic genetic markers. Eventually, the mutant is cloned after a functional sequence alteration is identified and confirmed by detection of sequence variation between mutant and wild-type siblings.

The process of positional cloning involves unique issues for all organisms. Success is usually based on experience. Because the Zon laboratory at the Children's Hospital of Boston has used positional cloning to isolate tens of zebrafish genes in lab and collaboration projects, we as a team have accumulated significant experience in the process. This guide summarizes this experience and shares practical and helpful tips with researchers, who study and work with zebrafish and needs to clone their mutants of interest. We hope it can increase mutant mapping and cloning efficiency and avoid unnecessary difficulties in the process and allow more resources to focus on learning interesting biology from mutagenesis screens.

This guide was first written at the time when the zebrafish genome sequence assemblies were improving. Many positional cloning projects were in regions where no reliable sequence assemblies were available for researchers to do chromosome walk electronically. With the latest genome assembly (Zv9), many of the chromosome walk discussion use large insert genomic clones are no longer necessary unless a mutant is mapped to an assembly region that is less reliable and did not agree with one's genetic mapping data. We decided to keep most of the chromosome walk discussions and would like to advise readers to use this guide selectively based on specific needs of individual cloning projects.

II. Mapping Strains

Many of the problematic issues concerning positional cloning in zebrafish (*Danio rerio*) arise from the genetic polymorphisms in the individual strains of zebrafish. In the Zon Laboratory, we have typically used the AB or TU (Tubingen) strains for mutagenesis screens. Mutants should he maintained in the laboratory strains that were used to create the mutants. These and all other widely available zebrafish strains are not entirely inbred. Given genetic polymorphisms may or may not be present in a given family of fish. No one should assume that zebrafish lines are as

isogenic as inbred strains of mice. Therefore, it is very important to track polymorphisms in a mutant family by obtaining DNA samples usually from tail clips of grandparents and parents in mapcrosses (see cross definitions in Section IV). This tracking will enable an examination of all possible polymorphisms in progenies, which subsequently become important in selection of close polymorphic markers to genetically locate a mutant gene.

WIK and SJD, two zebrafish strains, are more polymorphic with respect to the AB and TU strains and can therefore be used for genetic mapping of mutants on AB and TU backgrounds. Strains used in mutagenesis and mapping can he interchanged; that is, mutations can be created on WIK or SJD, and mapping can be done using AB or TU strains. For our mutations on the AB background, a heterozygote carrier on the AB strain is mated to WIK, and a mapping family is generated.

In the last phase of the zebrafish genome sequencing project, the Sanger Center created a new strain, named SAT (Sanger AB Tubingen hybrid strain), for generating a SNP-based high density genetic map. The parent genomes, a double haploid AB genome and a double haploid Tu genome, were fully sequenced using Solexa high throughput sequencing each with about 40× genome coverage. This strain is an excellent candidate strain for both mutagenesis screening and positional cloning. The Sanger Center has made the strain available at ZIRC for distribution. The field will need to optimize the maintenance of this unique strain.

III. Families and Genetic Markers

Once mapcrosses have been created and heterozygote F1 mapping pairs have been identified from these mapping families, it is important to tail clip and store the DNA from these mapping pairs (parents) and their parents (grandparents) (e.g., the AB mutant heterozygote and WIK wildtype that created the mapcrosses are the grandparents and the AB/WIK heterozygotes are the parents). Parent DNA and grandparent DNA samples are helpful for analyzing usable polymorphisms in subsequent mapping. Once flanking markers have been found in the initial mapping stage, it is important to test the parents and the grandparents to determine which mapping families (assuming multiple mapcrosses of different pairs of grandparents) and/or parent pairs (from single or different mapcrosses) have usable polymorphism that segregate the flanking markers in an easily interpretable manner. In other words, it is best to collect embryos with suitable allele systems for intermediate- and high-resolution mapping to narrow the genetic interval of the mutation loci. Once polymorphism has been determined, the mapping heterozygotes that are polymorphic can be selectively used to create large mapping panels that are polymorphic for the flanking markers. Very frequently, if a mutant does not map easily because of lacking good polymorphic markers in one mapcross family, examining another mapcross family is very helpful. It is recommended to create multiple mapcross families using different grandparent pairs at early stage of cloning to save time (at least 6 months) on fish breeding.

IV. Crosses for Line Maintenance and Mapping

For the purposes of this discussion, we should note that mutagenesis is performed with AB, and WIK is the polymorphic strain used for mapping. In addition, the reader should assume that the mutation of interest is homozygous embryonic lethal, and lines must be maintained as heterozygotes.

Definitions
Incross: Sibling cross
Outcross: AB (mutant)/AB (wildtype) heterozygote × AB/AB wildtype
Mapcross: AB (mutant)/AB (wildtype) heterozygote × WIK/WIK wildtype
Backcross: AB (mutant)/WIK (wildtype) heterozygote × WIK/WIK wildtype

For long-term line maintenance, we keep the mutation on the same strain in which mutagenesis was originally performed (in this case AB) so that we do not jeopardize future mapping efforts by introducing additional polymorphism into the mutant genetic background. For this purpose, one could perform an AB (mutant)/AB heterozygote incross or an outcross. Outcrosses are generally preferable because they help dilute out other recessive mutations that are acquired during the ENU mutagenesis and are not linked to the phenotype of interest. Using mapcrosses or backcrosses for line maintenance is problematic because recombination can result in a polymorphism loss of distant linked markers that could be critical for future low-resolution mapping (see Section VI.A).

For mapping, polymorphic hybrid strains must be created through a mapcross. The offspring are raised, half of which will be heterozygotes. Heterozygotes are identified by multiple random paired incrosses, and identified individual heterozygote pairs are set aside in isolation tanks for ongoing collection of homozygous mutant and sibling wildtype (homozygous and heterozygous) embryos from these paired incrosses. If we cannot collect enough embryos in that generation and/or the close flanking markers are not found, we turn to the newest generation of the heterozygotes on original strain background [AB (mutant)/AB (wildtype)] and perform another mapcross to repeat the process. We do not raise embryos from an AB (mutant)/WIK (wildtype) incross because recombination in this generation renders the next generation of embryos useless for mapping. If, however, close markers flanking the mutation are available, a backcross (ideally the same WIK (wildtype) is used in this cross and see the next paragraphs on backcrosses) can be performed to generate a large number of parents with identical polymorphisms for fast collection of a large mapping panel. If the markers are agarose scorable, meaning that a length polymorphism can be resolved and visualized using high resolution agarose gel electrophoresis), the potential exists to identify large numbers of parent heterozygotes with a good allele system by tail clipping instead of mating. However, we have to keep in mind that one can only learn about polymorphism of markers tested and cannot predict polymorphism of new markers that have not been tested in a given family and are closer to mutation loci. Thus, it is critical to keep track of family history of individual embryos in the panels. In cases where no informative polymorphism at

closer markers in a particular family during mapping, we will not use this family on the mapping panel for estimating genetic distances.

Using a high-throughput PCR format, we have identified up to 50 heterozygote pairs from a single generation of a mapping cross in a week. It is critical to tail clip the parents used in the mapping incrosses as well as tail clip the grandparents used for the mapcrosses so the allele system can be accurately followed in the next generation. To facilitate this process and backcrosses, it is useful for a laboratory to have isolated several (about 10) tail-clipped wild-type fish (in this example, WIK is used). Before the backcross is performed, the tail-clipped WIK wildtypes are screened with close flanking markers along with tail-clip DNA from the AB (mutant)/WIK heterozygote to establish which WIK wildtype has the best allele system for following the mutation. These WIK wildtypes are then used for the backcross. It is recommended to double check the allele systems of every parent pair used to produce the mapping panels.

When identifying heterozygotes for backcrosses by tail-clip DNA, it is important to remember that there will be a defined error rate due to recombination. The magnitude of the error depends on the distance between the marker and the mutation. The recombination rate in males is several fold lower than in females; therefore, if flanking markers are still somewhat far from the mutation, one should consider doing a backcross in which the AB (mutant)/WIK is male and the WIK wildtype is female. The goal here is to lower the chance for possible recommendations that take place between the identified marker and your mutation of interest in any backcrosses.

A. Choosing Grandparents and Parents for Better High-resolution Mapping

Usually after low-resolution mapping, our laboratory segregates individual WIK fish and then genotypes them. The goal is to select alleles of known markers in the system that represents the best advantage for our intermediate- and high-resolution mapping purposes (Section VI).

B. Microsatellite Marker – Agarose Scorable

We have undertaken a large-scale approach to evaluate microsatellites (available from the Fishman laboratory) for their ability to be scorable on an agarose gel (Massachusetts General Hospital, 2001) (http://zebrafish.mgh.harvard.edu/).

Through this analysis, we realized that SJD fishes were mostly isogenic; however, the strain has some regions with polymorphisms. SJD allows for easier mapping than WIK, but the strain is difficult to maintain and cannot be propagated in our laboratory. SJD males can be obtained and used to fertilize eggs *in vitro* from heterozygote females carrying the mutation for creating mapping families. The WIK strain works well, but it is very polymorphic between individual fish (not inbred). We use WIK in most of our mapping. We have put together a new mapping panel (288 CAs) for

genome-wide scan. A list of these markers can be downloaded at http://zfrhmaps.tch.harvard.edu/ZonRHmapper/positionCloningGuide/288LowResScanCAsMay2006.htm.

C. SNP Markers

Single nucleotide polymorphism (SNP) is the most common sequence polymorphism in genomes. It was estimated that there was 1 SNP in every 145 bp nucleotide sequence between C32 and SJD strains. By comparing genomic sequences from a double haploid AB and a double haploid Tu, larger than 10 million SNPs were observed. Given the quality of current genome sequence assembly (Zv9), researchers can feel confident that majority mutants can be mapped and cloned using known and unknown SNP markers that are present in a giving mapping family. However, a selected mapcross family will make the mapping and cloning process faster and easier. Genotyping SNP markers in general can be achieved by allele specific real-time PCR analysis, SNP microarray chip hybridization, direct sequencing of a SNP region, and even random whole-genome sequencing using next generation sequencing technologies. Different genetic backgrounds will likely be polymorphic at a subset of SNP sites found in public sequences. Your mapping families could have their unique set of SNPs. Design of new SNP assays is necessary. We listed three most commonly used SNP genotyping methods.

1. Allele Specific Quantitative PCR

Allele specific quantitative PCR is limited in bulk mapping since the required fluorescent reagents can be expensive and have short shelf life. The primer design and testing and requirement of using fluorescent primers also limit its application during the initial mapping of mutants. However, KBioscience in England, Britain has developed a zebrafish SNP panel containing 1073 SNPs, that can be used to map and genotype mutants. Additional information is available at http://www.kbioscience.co.uk/. Established SNP assays can be ordered and custom assays can be designed by submitting SNP sequences to KBioscience. SNP genotyping services and/or ready-to-use assays can then be ordered.

KBioscience has decided to discourage SNP assay design by individual researcher using their online tool, Primer Picker. Currently, researchers have to register a user account online at http://www.kbioscience.co.uk/application form.html before submitting sequences for custom SNP assay design using SNP genotyping service. Users can choose to use KBioscience genotyping service for genotyping the samples or order the assay kits for running genotyping assay in their own labs. In both cases, ones will need to follow detailed instruction on how to conduct SNP assay using the KASPar system. The detailed manual including data analysis can be found at http://www.kbioscience.co.uk/download/KASP%20manual.pdf and at http://www.kbioscience.co.uk/reagents/KASPar_SNP_Genotyping_System_Leafletv6.3.pdf.

2. SNP Chip Hybridization

There have not been any commercial zebrafish SNP genotype microarray platforms available. A SNP genotyping microarray containing 2000 SNPs derived from 4000 on the zebrafish heat shock genetic mapping was developed in Talbot's lab (Stickney *et al*., 2002). Guryev *et al*. (2005, 2006) identified more than 50,000 SNPs by examination of sequence differences between ESTs and whole genome random shotgun sequences. Later, Bradley *et al.* (2007) at Vanderbilt University has found 645,088 SNPS from public zebrafish sequences and established a custom SNP genotyping slide array containing about 200,000 validated SNPs. Recently in the genome sequencing project at the Sanger Center, over 10 million SNPs were discovered by whole genome sequencing of two double haploid fish, one AB and one TU. 200,000 SNPs were used to design a custom SNP Affymetrix genotyping chip. This valuable tool will be available for zebrafish genotyping and mapping projects.

Affymetrix provide a general protocol for using Affymetrix SNP genotyping arrays. It is recommended to follow the protocol at http://gegf.net/web/wp-content/uploads/2008/03/genomewidesnp6_manual.pdf closely when using the zebrafish SNP array. DNA samples used for the array needs to be free of PCR inhibitors and not highly degraded. Any DNA contamination from other fish DNA or non-fish DNA should be avoided. Affymetrix SNP chip can work with amplified DNA using whole genome amplification kits. RpliG from Qiagen has been tested by Affymetrix and can work well with the Affymetrix SNP array system. It is also a good idea to discuss genotyping details with an Affymetrix core, where the prehybridization process of genomic DNA and hybridization will take place, to make sure all requirements are met and informative genotyping information can be obtained since SNP chip hybridization is still very expensive when it is compared to traditional agarose and polyacrylamide gel analysis.

One important consideration is that only 20–30% polymorphic markers are expected to remain polymorphic in a given family. The number of SNPs on a SNP panel needs to be large enough to include all SNP haplotypes at each genomic locus and to have multiple markers representing a genomic region. This can insure even coverage of the genome with polymorphic markers in any mapping projects. The limitation of the real time PCR and SNP chip array approaches is that the mapping resolution is still dependent on the frequency of genetic recombination. High-resolution mapping can only be achieved by increased genotyping of individuals in a family. However, it can get expensive when a large number of genomes need to be genotyped.

3. Direct Sequencing of the SNP Containing Region

Very often, candidate genes in a genetic interval will be sequenced at cDNA or genomic levels in the final stage of positional cloning. We have found family specific SNPs will be discovered between mutant and wild-type backgrounds. These SNPs are useful to further narrow down the genetic region. Here, direct sequencing can be used especially after the high resolution mapping when small-size recombinant

panels from both sides of a mutation have been established. The SNP regions are amplified by PCR using about 20 ng genomic DNA. Amplified PCR products are purified to remove excess primers and submitted for Sanger sequencing. Sequences from wildtype and mutant genomes are then aligned using sequence alignment programs and SNP genotypes are called. In heterozygous situation, SNP sites can only be scored using sequencing electropherogram and heterozygous SNPs should appear as double peaks of both wildtype and mutant nucleotides.

SNPs can also be assayed by allele specific oligonucleotide hybridization (ASO), (Farr *et al*., 1988; Wood *et al*., 1985) SSCP, and direct sequencing. These methods have been discussed in previous edition of the protocol and will not be discussed in details here

V. Preparation of the DNA

If the embryos are not hatched, chorion is removed before preparing DNA from these embryos. We place the embryos individually into wells of a 96-well plate. We then remove excess buffer and store them dry or in methanol. These samples should be kept at −20 °C. Embryos can also be stored together at −20 °C after dehydrated in methanol and arrayed into a 96-well plate when DNA preparation is needed.

When working with embryos frozen or methanol preserved, plates should be kept on ice unless otherwise noted. To prepare the embryos, remove all methanol or excess liquid from the wells. All of the following incubation steps can be carried out in a PCR machine:

1. Add 50 μL of lysis buffer (composition follows) to each well and incubate at 98 °C for 10 min to lyse cells. Quench on ice or 4 °C in the PCR machine.
2. Add 5 μL of Proteinase K (10 mg/mL stock) to degrade proteins.
3. Incubate at 55 °C for at least 2 h. We recommend at least one mixing during the PK incubation to increase consistency of the DNA preparations. This incubation can also be left to run overnight. The longer the incubation time, the cleaner the DNA tends to be.
4. Incubate at 98 °C for 10 min after a 55 °C incubation to destroy Proteinase K. Quench on ice or a 4 °C sink in PCR machine (a 10 °C sink in PCR machine is better for PCR machines especially in an overnight procedure).
5. Spin down lysed embryo debris at 4000 rpm for 10 min.
6. Draw off supernatant into a clean 96-well plate.
7. Dilute as necessary.

Embryo Lysis Buffer

Solution: 1 × PCR buffer made to 0.3% Tween 20 (10% stock) and 0.3% NP40 (10% stock)

For 10 mL of embryo lysis buffer, use the following:
10 mL PCR buffer (see following composition)

300 μL NP40, 10% stock
300 μL Tween 20, 10% stock

PCR Buffer
Solution: 10 mM Tris-HCl, pH 8.3, and 50 mM KCl

For 50 mL of buffer, use the following:
500 μL 1 M Tris, pH 8.3 (autoclaved)
2.5 mL 1 M KCl
47 mL sterile ddH_zO

Commercial kits can be used to isolate high quality DNA samples for real-time PCR and SNP chip analysis. Please follow the instructions provided by KBioscience and Affymetrix. When sample numbers are small, proteinase K digestion followed by careful phenol–chloroform extractions should yield high quality DNA samples for these applications.

VI. Mapping Genes

A. Low-resolution Mapping – Bulk Segregant Linkage Analysis

The preferred method for low-resolution mapping of a mutant to a particular chromosome is the bulk-segregant analysis. This method makes use of scanning microsatellite markers throughout the genome (Knapik *et al.*, 1998; Shimoda *et al.*, 1999). The microsatellite markers on the genome wide scanning panel are pre-selected for their agarose scorabilities due to large differences in the number of simple repeats between different genetic backgrounds. We have successfully mapped many recessive and dominant mutants using this approach.

The first step in mapping a recessive mutation to a chromosome is the generation of mapping hybrids (AB (mutant)/WIK (wildtype)). To make this mapping cross, a heterozygous AB (mutant)/AB (wildtype)) is mated to a wild-type WIK fish, and the resulting F1 generation is raised. Practically, we generate 2–4 mapcross families with different WIK founders to ensure an informative allele system is obtained (because AB and WIK may share common alleles at certain genetic loci and WIK is a polymorphic strain). To identify heterozygous F1 individuals, AB/WIK hybrids are mated to each other (incrossed), and their clutches scored for the mutant phenotype and the mutant/wildtype ratio (1:3 mutant to wildtype). Once a pair of heterozygous hybrids is found, they are mated, and their wild-type and mutant progenies are both collected.

In addition, we tail clip the mapping heterozygotes to obtain DNA for tracking the alleles carried in a family. For the initial low-resolution mapping, we use 40 mutants and 40 wildtypes from a single clutch. Two mutant pools of 20 and two wild-type pools of 20 are made from these stocks. To make the pools, we take 8 μL of the individual original stock DNA from each of the 20 individuals (mutant

embryos for the mutant pool or wild-type embryos for the wild-type pool) to give 160 μL, and we dilute the pool DNA to 1.6 mL with water and use 4 μL of the dilution for each PCR reaction. Lastly, we use the agarose-scorable microsatellite markers in PCR reactions to scan the genome for linkage to the phenotype in a bulk segregant analysis.

A bulk segregant analysis uses 288 agarose-scorable microsatellites that are typed on a set of four DNA samples from wild-type and mutant embryos. Each set contains two pools of 20 wild-type embryos each and two pools of 20 mutant embryos each. PCR products are run on 2% agarose gels (1% regular resolution and 1% of high resolution agarose) at 200 V for 2 h to separate bands. Thus far, most polymorphisms encountered have been subtle; hence, running the gels longer than necessary is always better. Linkage is assumed when a band present in the wild-type pools is absent in the mutant pools. The mutant band(s) may also have a size shift when compared with the wildtypes. This may also indicate a linkage. It is the best that embryos used for this stage and the intermediate mapping stage (see next section) come from the same family. Occasionally, our laboratory is not able to map a mutation in one family, and we have to test another family. It is critical that the 80 embryos come from the same family. Introducing individuals with a different set of alleles might lead to false positives, confusing allele segregations, and inaccurate estimation of genetic distances. One might confirm linkage to a particular chromosome based on the pattern obtained with additional families. Very frequently, we find linkage to more than one chromosome, but only one of these linkages is real. To determine which microsatellite is truly linked, each positive marker must be tested on individuals that are used to create the pools. By testing individuals, chromosomal linkage is confirmed, and the distance between markers can be evaluated.

B. Intermediate-resolution Mapping

The purpose of intermediate-resolution mapping is to position mutants between close flanking markers that are scorable on an agarose gel. This technique allows us to do high-resolution mapping with 1500 embryos with relative ease. Mapping with this number of embryos is not always possible, but it is a goal worth striving for.

We use 8 wildtypes and 88 mutants for intermediate-resolution mapping. To clone a mutant gene, flanking markers should ideally be less than 10 cM apart. We scan microsatellites on the chromosome by ordering roughly six microsatellites on the linked chromosome arm. If these microsatellites are not polymorphic, we test another six markers until the mutation is linked to two flanking polymorphic microsatellites on the chromosome arm. Based on this recombination mapping strategy, it should be possible to determine the close flanking microsatellite markers. Markers that are far away from the mutation should yield more recombinants than markers that are closer to the mutation. Markers that are on opposite sides of the mutation should have different sets of recombinants, and markers that

are on the same side of the mutation should share recombinants. We narrow down the region by analyzing more microsatellites until we have markers that are 10 cM (or less) apart or as close as possible. When we define microsatellite markers that are polymorphic on an agarose gel and that are close enough to use as flanking markers, we set up mapping crosses with zebrafish that have this allele system. These new stocks of fish and validated flanking markers are then used for the high-resolution mapping.

C. Fish Husbandry

The number of tanks needed to map a mutant varies based on zebrafish sex ratio in the tanks and on the ease of scoring the phenotype. We will typically generate eight mapcross tanks or backcross tanks for a genetic mapping. Ultimately, we will sacrifice most of these fish, but the goal is to have at least five pairs of fish with an advantageous allele system so that genetic mapping can be conducted very quickly.

D. Tempo

While working out the flanking polymorphisms and genotyping, it is important to continue collecting mutant embryos. We find that between 1500 and 2000 embryos are required for positional cloning. Assuming an interval of 600 kb/cM and a meiotic recombination frequency of one per embryo (this is true for haploid individuals; diploids have an average of 1.3 to 1.5 meioses/individual – one from the mother and 0.3–0.5 from the father) will give a resolution of close to 30 kb per meiosis event on a 2000 mutant embryo panel. This resolution allows the positioning of the mutant gene on a sequenced BAC or PAC clones on the genome sequence assemblies. The following discussions are based on the genome sequence assemblies. The goal is to identify closer polymorphic markers between the existing or new flanking markers on the mutant mapping panel. Ideally, the mutant locus is located to a defined genomic region with high quality sequences from a genomic clone (BAC or PAC) and a few or single genes are in the interval. Please refer to chromosome walking protocols in previous editions of this guide when the mutant region is mapped to a gap or mis-assembled region on the genome sequence assembly.

The characterization of markers between the flanking polymorphic markers is important. The number of recombinants obtained with each of the flanking markers (which should be identified on the genome assemblies by sequence BLAT on the UCSC genome browser) from intermediate-resolution mapping is considered a guide for estimating the position of the mutated gene. The sequence assemblies are annotated with a multitude of expressed sequence tags (ESTs), genes, BAC ends, and genomic sequences in the region, which can he used as markers on the walk toward the mutated gene. Typically, our laboratory will pick 3–4 known sequences in

proximity of the estimated gene location and first check if these sequences contain polymorphic SNPs by direct sequencing in our mapcross. In our experience, one out of four such sequences is polymorphic in a given mapcross, especially in non-protein coding regions. If a polymorphism with the chosen sequences is not found, more sequences can be tested for polymorphism. We have found that BLAST analysis of the Sanger Institute genome sequence (The Sanger Institute, 2004a) with sequences in the region will give more extensive sequence variation information. These sequences can be searched for polymorphisms. Introns have a higher rate of polymorphism than 3′-UTRs. In addition, CA repeat regions or contigs are also good for designing primers to find polymorphic markers. Once new polymorphic markers are identified, the panel of recombinants identified at each of the flanking markers is tested to see if new markers are closer to mutated sites.

1. Example

Problem: A total of 30 recombinants were identified with the left flanking marker and 35 with the right flanking marker. If 1500 mutant embryos have been collected as part of the mapping panel, the distance to the mutant gene is estimated to be 2 cM from the left flanking marker and 2.3 cM from the right flanking marker (both are too far to initiate a walk). Two markers are identified in the estimated region of the mutated gene. Both are polymorphic. Test the panels of recombinants from either side with these markers as shown in Fig. 1.

Solution: The mutated gene is situated between marker 1 and 2. The estimated distance from the closer marker 1 is 0.26 cM. This distance is sufficiently small to locate the mutant to a BAC, PAC, or sequence assembly scaffold. If the interval is in low quality or in a gap, a tradition chromosome walk can be initiated and the local map of the region needs to be established using ESTs and whole genome shotgun sequences.

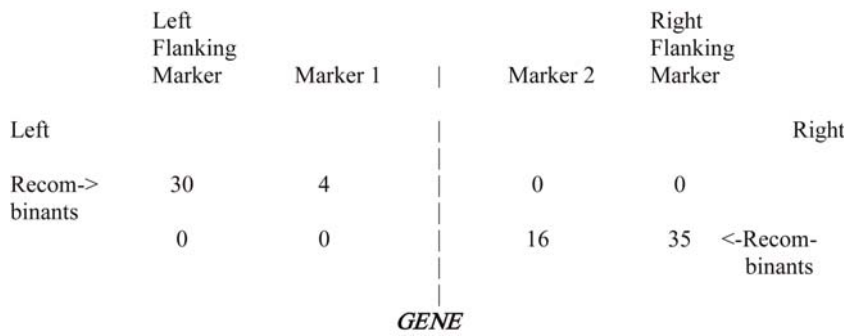


Fig. 1 Mapping of mutant genes.

E. High-resolution Mapping

We have traditionally collected between 1500 and 2000 mutant embryos in an effort to clone mutants positionally. In high-resolution mapping, these mutant embryos, arrayed in a 96-well format, are tested with the flanking markers found in the low and intermediate stages of mapping. It is critical that every recombination event is scored in this step, assuming a two-allele system. If a three- or four-allele system is used for mapping, some recombination events will be missed. Therefore, it is recommended that the families used for collecting the embryos be chosen according to the most useful allele system as well as the correct polymorphic mapping strain. Furthermore, it is advantageous to limit the number of families used to collect the mutant embryos. Although collecting embryos from only two or three pairs of fish may lengthen the time needed to reach the target number of embryos, this step will simplify allele systems of markers in downstream steps in positional cloning. We recently moved our high-resolution mapping to the ABI 3730 for quick scanning through entire mapping panels using the closest flanking SNP markers. We directly sequence PCR products of a SNP region amplified from genomic DNA samples isolated from different genetic backgrounds to determine the SNP genotype. Directly sequencing is also the ultimate way for researchers to confirm point mutations and deletion mutations. KASP SNP genotyping system can also be used here to narrow the intervals and clone mutants.

F. Three-allele Systems *versus* Four-allele Systems

Typing mapcross grandparents and parents is essential to evade the "allele traps" encountered by this type of bulk segregant analysis. Once flanking markers have been found, mapcross grandparents and parents should be tested for the proper allele segregation. Three-allele systems are quite common in the AB/WIK crosses used in our laboratory. We often see that the flanking markers do not segregate similarly between different families. This becomes a problem with the high-resolution scan that includes individual mutant embryos from all mapcrosses. One of the wild-type alleles sometimes migrates the same as the mutant alleles in agarose gels, so when a high-resolution scan is performed, recombinants can be missed because their banding patterns look like ones from a mutant embryo. A simple way to avoid these had-allele systems is to type all the parents and grandparents before collecting all 1500 embryos and discard families with bad allele systems from the mapping panel. While the low and intermediate stages of mapping are being performed, mutant embryos can be collected. It is very important to remove crosses that have bad allele migration from the mapping collection. Four-allele systems can be just as confusing if not fully investigated before the high-resolution phase. Generally, four alleles can he tracked with ease. Problems, however, can arise when an AB allele segregates with the mutant WIK allele. Heterozygotes would he counted as homozygous mutants because similar allele size of both AB and WIK and not as recombinants. A similar situation is shown with a three-allele system in Fig. 2.

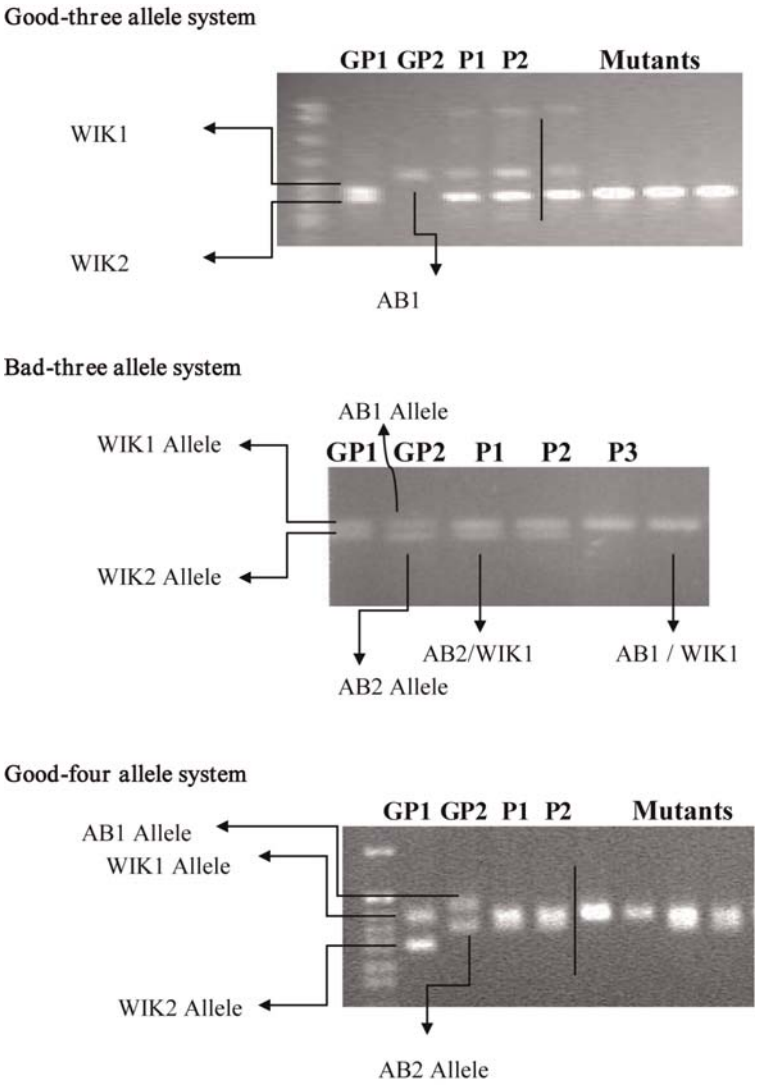


Fig. 2 Allele systems in zebrafish.

G. Collecting Mutant Embryos

With the conclusion of high-resolution mapping, a number of recombinants have been identified. DNA samples from these recombinants are re-arrayed onto two new 96-well plates to create the recombinant panels from each direction. During this process, it is advisable to continue collecting mutant embryos from the mapping strain beyond the initial collection number of 1500–2000 because more embryos

may be required later in the process. The recombinant panel is now used in the final stage of a positional cloning project, with the number of recombination events lessening as the mutation is neared by polymorphic genetic markers.

H. Chromosomal Walking

After having the recombinant panels made from 1500 to 2000 collected mutant embryo mapping panel and the closest flanking marker (s) within 1 cM distance, a chromosomal walking can be initiated (Fig. 3). The flanking markers, that are polymorphic by SSLP or SNP and linked to your mutation, are used to find the next internal markers between these two flanking markers. Based on a flanking marker sequence, they can be located on the genome sequence assemblies and a genomic interval is defined. Typically, marker sequences are derived from scaffold sequences in this interval. Ends of the sequence assembly scaffold are tested on the mapping panels and new polymorphism markers are identified. Using these internal polymorphic markers, orientation of the walk relative to the mutated gene is established by studying the recombinant panel. By taking the marker that is closer to the gene, more known sequences in the new interval are identified and sequenced from both of their ends. Polymorphism of these ends is tested on the recombinant panel again and the orientation of the walk is again established. These end sequences are used to generate new markers. These markers are typed on the mapping panel again, and therefore, a "walk" towards the mutant gene has been established. When starting a chromosomal walk, keep in mind that the narrower the genetic interval, the faster the positional cloning will go.

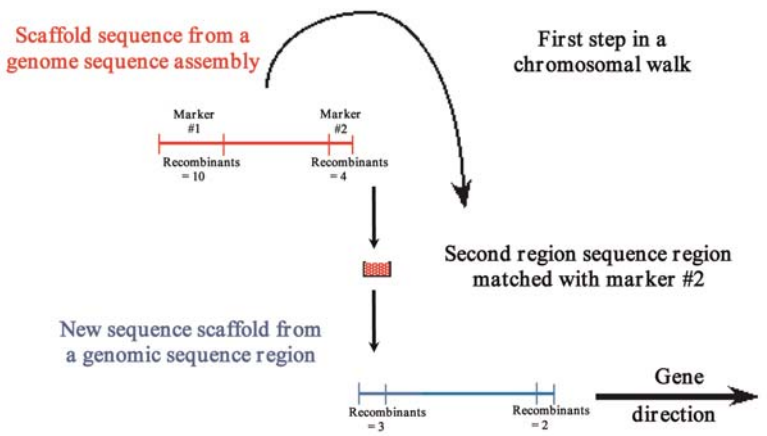


Fig. 3 Chromosome walking. (See color plate.)

VII. General Flow of Information from the Zebrafish Genome Sequence Assemblies

The Sanger Institute has released version 9 of the zebrafish genome assembly, called Zv9. This is the final version assembled by the zebrafish genome sequencing project, which has reached a maintenance stage. This zebrafish genome sequence assembly together with human and mouse ones are now in the hands of the Genome Reference Consortium. Zv9 assembly consists of 1.41 Gbp in 4560 scaffolds. It was generated starting with 1.18 Gbp from 11,099 BAC clones ordered with SATMAP, a new high-density meiotic map. The resulting genome sequence was supplemented with a new whole genome shotgun assembly WGS31, based on a combination of Illumina and capillary reads from double-haploid Tubingen individuals. The large contigs presented in the sequence assembly provide an incredibly useful start point to looking at your interval. Our laboratory traditionally examines flanking markers and markers that are in the critical genetic interval on the genome assembly and other genome maps. Very often, we will find that the same contig contains two individual makers from genome maps in the region. This process establishes that the contig does in fact represent a sequence between the flanking markers.

The Zon Laboratory has also developed two independent programs, available on the laboratory Web site, that evaluate the contigs. The first program involves BLASTing a human gene query to the assembly contigs (Zon Laboratory, 2004a). This procedure will allow any human gene in an interval to be BLASTed in order to find a contig that represents that human gene. Using BLAST in this manner is very useful for examining conserved synteny relationships in zebrafish and human genomes because other genes in the interval may be present. In addition, for our second program, we are using a reverse BLAST in which we have examined the Sanger Institute Zv7 and Zv8 assemblies BLASTed to known human proteins (Zon Laboratory, 2004b). The same can be done between zebrafish and other fish species such as Japanese pufferfish *(Fugu rubripes)* and Spotted green pufferfish (*Tetraodon nigrovirdis*) genomes as well as between human and other fish genomes. Reverse BLASTing allows an investigator to see if other genes are present on the contig and can be used again to identify a synteny. In addition, the Ensembl database at the Sanger Institute and the UCSC genome browser has gene annotation in individual contigs, which is a tremendously helpful resource (Sanger Institute, 2004b). We find that in our laboratory, we need to use both Ensemble and our own databases because even though the Ensemble database provides genes, the Web site is difficult to use for directly finding human proteins. Our Web site compares sequences by a BLAST statistical number and allows this type of analysis to occur very rapidly by providing information that is preBLASTed for the entire assembly against the human protein database. From this information, we can determine if a reliable contig exists within our critical genetic interval.

Our laboratory studies contigs that are very close to the mutated gene based on meiotic recombination. We use the assembly sequence, find simple repeats at http://

danio.mgh.harvard.edu/markers/ssr.html of greater than 12 nucleotides, and use flanking primers in single-stranded length polymorphism (SSLP) analysis to determine whether linkage is available. This process is much better than looking for SNPs that are present within the interval (although in important regions or genes, we do use SNPs especially when SSLP markers are not available). By finding simple repeats on a contig in the region, we can quickly study the critical interval and then genotype more embryos that will narrow this interval. Candidate genes in contigs can be tested by rescue assays or by morphilino analysis.

VIII. Synteny between Human, Zebrafish, Fugu, and Tetraodon Genomes

Japanese pufferfish (*Fugu rubripes*) genome sequencing project at the Department of Energy (DOE) Joint Genome Institute Web site and Spotted green pufferfish (*Tetraodon nigroviridis*) genome sequencing project at France are wonderful resources for those involved in positional cloning and comparative genomics (DOE Joint Genome Institute, 2004; and Genoscope and the Broad Institute of MIT and Harvard, 2004). The *Fugu* genome has been shotgun sequenced at a 4× the genome coverage level. This allows assembly of 10- to100-kb scaffolds of genomic sequence. The *Tetraodon* genome was sequenced at 8× coverage with an average size of supercontig at 7.62 Mb. Because of the relatively small size of the *Fugu* and *Tetraodon* genomes, synteny is extremely helpful for positional cloning in zebrafish. More recently, Medaka (10.6× genome coverage) and Stickleback (9.0× genome coverage) genomes have been sequenced and assembled into large genomic regions (University of Tokyo Genome Browser and Stickleback Genome Project). All these genome assemblies and automatic comparative genomic analysis are available on both UCSC and ENSMEBL genome browser sites. We make use of these sites by first finding a zebrafish genes and EST near a mutant gene. We then look at the human genome at browser sites set up by the University of California at Santa Cruz (University of California at Santa Cruz, 2004) and ENSEMBL, and identify human ortholog of this zebrafish gene. We hope to establish if there is zebrafish and human synteny. We look at all the adjacent genes in the human and see if there are zebrafish orthologs. We then use TBLASTN analysis of the *Fugu* and *Tetraodon* genome using the predicted peptide sequence of either human or zebrafish genes. Synteny can usually be found on a scaffold. Each scaffold is nicely annotated on the browser sites and individual project sites, and the sequences can be downloaded and used for BLAST analysis.

Alternatively, one can use gene prediction programs, such as BLAST P and GeneScan. These programs allow investigators to probe the syntenic relationships. When one clicks on the BLAST P or GeneScan program, individual predicted exons are revealed at the bottom of the page. Clicking on the exons gives the predicted homology between *Fugu*, *Tetraodon*, and human genomes, and this result can be

cross-referenced to the zebrafish. Clicking on the upper bar gives the predicted peptide of the *Fugu* and *Tetraodon* sequences. Using the predicted peptide of the *Fugu* and *Tetraodon* sequences, we use our TBLASTN server and establish if the Sanger Institute sequence has the zebrafish ortholog of the *Fugu* and *Tetraodon* gene. When putative orthologous genes are not available on the zebrafish genome assembly, they are mapped using the zebrafish radiation hybrid panels to confirm if syntenic relationships occur. We have found that for most chromosomal walks, synteny does exist. These sites are also extremely useful for isolating zebrafish orthologs of human or *Fugu* genes. The reader can browse the Zon laboratory's comparative genomics data (e.g., for human, zebrafish, and *Fugu* genomes) for more information (Zon Laboratory, 2004c).

With the newly finished zebrafish genome sequence assembly (Zv9), it has become easier to conduct comparative genomic analysis using both the UCSC and Ensembl genome browser. On the UCSC browser page, a user can turn on all Comparative Genomics tracks, including Fish Chain/Net, Vertebrate Chain/Net especially Human Proteins track in Genes and Gene Prediction Tracks group, and Chimp Chain/Net. The Fish Chain/Net track group includes four other sequenced fish genomes, Fugu, Medaka, Stickleback, and Tetraodon. The Vertebrate Genomes, Chain and Net Alignments, include human, mouse, and frog, *X. tropicalis*. It is helpful to select "Color track based on chromosome" so the chromosomal location of homologous genes are apparent on the browser page. You will need to click on individual tracks in group track setting to turn on and off the function "color track based on chromosome." Since these homology analyses were done using an automated workflow, it is important to verify by manual homology examinations as described above, especially when gene synteny gaps or inserions are observed in a region. ENSEMBL genome browser offers similar pre-calculated analysis but it tends to work slower in USA. It is helpful to check both browser sites for confirmation of observed gene synteny.

A number of sites have been created to help with positional cloning. These sites are listed on the ZFIN database at The Zebrafish Model Organism Database Web site (ZFIN, 2004). We have provided ZFIN with a short description of each of the Web sites.

In a positional cloning project, we find a close polymorphic marker after the high resolution map, and we proceed to evaluate the fingerprinted BAC contigs and available end sequences of BACs in contigs. We then derive markers from those end sequences and determine where on the chromosomal walk the markers may lie relative to the mutant gene. This process will be used increasingly in the future as chromosomal walking by hybridizations will only become necessary if mutants are mapped to a gap or less known region on genome sequence assemblies.

The Genome Reference Consortium is willing to extend contigs for individual projects. An interested individual can simply contact zfish-help@sanger.ac.uk and provide information regarding a chromosomal walk. The assembly team can lower the stringency of the FPC software and thereby extend walks over many more contigs. The fingerprinting software is very specific for matches of Hind III

fragments. By relaxing the stringency (where not all Hind III fragments must match), one can extend contigs, but one can also encounter less certainty that the contigs are truly forming a physical linear assembly. Nevertheless, genetic or genomic locations of these BACs within these regions can be confirmed by mappings of markers derived from these BACs on the mutant mapping panel or on the RH mapping panels.

IX. Proving a Candidate Gene is Responsible for the Mutant Phenotype

What is termed the *rescue* of the mutant phenotype is the gold standard for confirming that a candidate gene is responsible for the mutant phenotype. One can demonstrate that the wild-type gene can rescue the genetic defect (assuming that the defect has compromised the function of the gene rather than generated a gain-of-function mutation). To do this, both wild-type and mutant cDNAs should be subcloned into a vector that is suitable for the synthesis of capped mRNA (i.e., an RNA polymerase site at the 5′-end of the cDNA, a stable 3′-untranslated region containing a polyadenylation signal, and unique restriction enzyme sites for linearization of the plasmid). Commonly used vectors include pCS2+, pSP64T, and pXT7. Both wild-type and mutant cDNA plasmids should be used to generate an *in vitro* translated protein (i.e., ^{35}S-met labeled), which can be resolved on acrylamide minigels and detected by autoradiography. This assay provides an assessment of protein size, stability, and translational efficacy of the wild-type and mutant constructs.

"Run-off" synthetic mRNA transcripts can be generated from the linearized template using commercially available kits, purified, and then resuspended in either RNase-free distilled water or 1× Danieau's solution (typically at a stock concentration of 1–3 μg/μL). Diluted mRNA (typically 5–500 pg/μL, depending on the potency and toxicity of the encoded protein) is then injected into 1–4-cell stage embryos produced by parents carrying the mutation. Amelioration of the phenotype (either by morphology or biochemical assay) is then assessed at the appropriate developmental stage. Generally, the later the onset of the mutant phenotype, the more difficult it is to *rescue,* owing to the instability of the injected mRNA and protein during development. Thus, it is unreasonable to expect every mutant embryo to be rescued. In addition, many proteins, particularly transcriptional regulators when over-expressed early, can have severe effects on early embryonic patterning events, thereby making it impossible to rescue later developmental pathways. Injection of mRNA synthesized from the mutant cDNA plasmid permits an assessment of the severity of the mutation [i.e., no *rescue* (or morphological effect on development) would suggest a complete loss of function, whereas *partial rescue* may indicate a hypomorphic mutation].

As alternatives, wild-type and mutant cDNA constructs can be used to generate expression plasmid construct driven under powerful ubiquitous promoters and

contain fluorescent protein tags, such as β-actin and CMV promoters, and tissue/cell specific promoters if the cell and tissue types and timing of the promoter activation are known. The constructs need to be fully sequenced to make sure there are no critical mutations in the sequence especially if cDNA constructs are from a PCR reaction. The plasmid constructs are then injected into mutant embryos at one-cell stage and scored for transient rescue. If necessary, stable transgenic fish may need to be raised and analyze rescue.

Genomic clones are sequenced and provide abundant coverage of genomic regions of protein coding genes. It is now possible to modify large genomic clones (containing native cis-regulatory elements) to include color fluorescent protein tags. Tagged and untagged genomic clones can be injected into fertilized eggs and score for mutant phenotype rescue. Again, stable lines may need to be raise to observe the rescue. Fluorescent tagged transgenics are much easier to select for stable lines. In large genomic transgenics, native gene expression timing and pattern can be preserved and result in better rescue. On the other hand, the gene expression level and timing may be influenced by the genomic environment of an integration site, RNA *in situ* hybridization analysis may be needed to make sure proper expression of a transgenic construct before evaluation of rescue results.

X. Morpholinos

To evaluate candidate genes, one morpholino is used against its ATG region to prevent translation, and another morpholino is designed against a splice site. Typically, we use the splice donor because this seems to create aberrant splicing. An RT-PCR analysis of the target mRNA products can be extremely useful in showing that there is no normal splice form or a small percentage of normal splice product in morpholino injected embryos. Correlation between absence or reduction of normal mRNA splicing products and mutant phenotype provides a proof that the candidate gene is the mutated gene in the genetic mutant.

XI. Future Technologies

A. Region-specific Sequence Capture

The high quality genome sequence assembly and the second generation sequencing technologies have offered new approaches in mapping and cloning of genetic mutants. After the initial mapping of mutant to a fine region, for example, about 100 kb region, one can design allele specific oligonucleotides to capture the mutation containing region and corresponding wild-type region from mutant and wild-type genomes. The captured DNA fragments can be make into sequencing libraries that are specific for different second generation sequencing platforms and sequenced at many times coverage. These short read sequences were then assembled and

compared between the mutant and wild-type genome. The point mutation, deletions, and insertion can be identified. The detected sequence changes can then be confirmed by the Sanger sequencing of the mutated sequences. A couple labs have worked with Generation Biotech at http://www.generationbiotech.com/ (Gupta *et al.*, 2010; Zon lab, unpublished). Currently, this platform is still very project specific and no general experimental protocols have been developed yet.

B. Exome-capture and Next Generation Sequencing (NGS)

Alternatively, all exons (exome), known and predicted, based on genome assembly can be captured using SureSelect from Agilent and Sequence Capture Arrays or SeqCap libraries from NimbleGen, Roche. NGS sequencing platforms are then used to sequence all known exons in one sequencing run. Sequence changes in any exons can be quickly detected without going through traditional mapping and cloning process. The obvious limit of this approach is that non-exon mutations cannot be discovered by these approaches.

C. Whole Genome Sequencing by NGS

Attempts have been made to directly sequence genomes of mutants and wild-type siblings and look for SNPs that co-segregate with mutant phenotypes at 100% ratio. There is no doubt that linkages will be established and mutants be mapped to chromosomes. However, because of limited recombination events (mutant and wild-type genomes) examined, the mutation region may be defined to a large genomic interval. It will be exciting to see if the high-throughput sequencing can significantly reduce the amount of time and effort required for cloning mutants.

These new approaches are powerful but can be expensive at current market pricing. The workflow has not been tested sufficiently. More importantly, this approach requires significant amount of bioinformatic support, at least at the beginning phase, to analyze obtained sequences for making SNP calls. SNP genotyping methods described earlier in this chapter are useful for genotyping a large number of individual fish and confirm causal sequence changes in mutant genomes.

References

Bradley, K. M., Elmore, J. B., Breyer, J. P., Yaspan, B. L., Jessen, J. R., Knapik, E. W., and Smith, J. R. (2007). A major zebrafish polymorphism resource for genetic mapping. *Genome Biol.* **8**(4), R55.

DOE Joint Genome Institute (2004). Japanese pufferfish *Fugu rubripes*. http://genome.jgi-psf.orgIfugu6/fugu6.home.htm1.

Farr, C. I., Saiki, R. K., Erlich, H. A., McCormick, F., and Marshall, C. J. (1988). Analysis of RAS gene mutations in acute myeloid leukemia by polymerase chain reaction and oligonucleotide probes. *Proc. Natl Acad. Sci. U. S. A.* **85**, 1629–1633.

Genoscope and the Broad Institute of MIT and Harvard (2004). Spotted Green pufferfish *Tetraodon nigroviridis*. http://www.genoscope.cns.fr/externe/tetranew/.

Gupta, T., Marlow, F. L., Ferriola, D., Mackiewicz, K., Dapprich, J., Monos, D., and Mullins, M. C. (2010). Microtubule actin crosslinking factor 1 regulates the Balbiani body and animal-vegetal polarity of the zebrafish oocyte. *PLoS Genet.* **6**(8), e1001073.

Guryev, V., Berezikov, E., and Cuppen, E. (2005). CASCAD: a database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. *BMC Genomics* **27**(6), 10.

Guryev, V., Koudijs, M. J., Berezikov, E., Johnson, S. L., Plasterk, R. H., van Eeden, F. J., and Cuppen, E. (2006). Genetic variation in the zebrafish. *Genome Res.* **16**(4), 491–497.

Knapik, E. W., Goodman, A., Ekker, M., Chevrette, M., Delgado, I., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M. C., and Jacob, H. J. (1998). A microsatellite genetic linkage map for zebrafish (Danio rerio). *Nat. Genet.* **18**(4), 338–343.

Massachusetts General Hospital (2001). Zebrafish microsatellite map server. http://zehralish.mgh.harvard.edu/mapping/ssr_map_index.html.

Shimoda, N., Knapik, E. W., Ziniti, I., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., and Fishman, M. C. (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**(3), 219–232.

Stickleback Genome Project. http://www.broadinstitute.org/models/stickleback.

Stickney, H. L., Schmutz, J., Woods, I. G., Holtzer, C. C., Dickson, M. C., Kelly, P. D., Myers, R. M., and Talbot, W. S. (2002). Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Res.* **12**(12), 1929–1934.

The Sanger Institute (2004a). Sequencing projects BLAST search services. http://iwww.sanger.ac.uk/DataSearch/.

The Sanger Institute (2004b). Pre! Project Ensembl. http://pre.ensembl.org.

University of California at Santa Cruz (2004) UCSC Genome Bioinformatics. http://genome.ucsc.edu/index.html.

University of Tokyo Medaka Genome Browser. http://utgenome.org/UTGBMedaka/.

Wood, W. I., Gitschier, I., Lasky, L. A., and Lawn, R. M. (1985). Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl Acad. Sci. U. S. A.* **82**, 1585–1588.

ZFIN (2004). The zebrafish model organism database. http://zfin.org/zfin_info/dbase/db.html.

Human BLAST against zebrafish assembly Zv8. http://zfrhmaps.tch.harvard.edu/Hg18BlastZv8/.

Zon Laboratory (2004b). Zebrafish assembly Zv8 BLAST human gene. http://zfrhmaps.tch.harvard.edu/Zv8BlastHg18/.

Zon Laboratory (2004c). Zon lab comparative genomics. http://zfrhmaps.tch.harvard.edu/compGenomics.htm.