Gabriel D. Dakubo

Cancer Biomarkers in Body Fluids

Principles

Springer

Cancer Biomarkers in Body Fluids

Gabriel D. Dakubo

Cancer Biomarkers in Body Fluids

Principles

Springer

Gabriel D. Dakubo
Medical Sciences Division
Northern Ontario School of Medicine
West Campus, Lakehead University
Thunder Bay, Ontario
Canada

ISBN 978-3-319-01579-8 ISBN 978-3-319-01580-4 (eBook)
DOI 10.1007/978-3-319-01580-4

Library of Congress Control Number: 2016950197

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To Luciana and Patricia
In memory of Cuthbert and Augustine

Preface

Cancer is a complex, multifaceted, dynamic, and progressive disease that continues to plague the world. With improved quality of life and increasing life expectancy, and yet our inability to completely curtail the intricate oncogenic networks, the incidence and prevalence of cancer is likely to continue to increase. Obviously, this trend will reflect adversely on health economics and quality of life. Even more worrisome is the fact that the burden of cancer currently rests heavily on resource-poor regions of the world.

Primary prevention is the ideal approach to cancer control. However, the challenges with successful implementation of cancer prevention programs are obvious, and not all cancers are easily amenable to such control measures. Therefore, the utility of biomarkers offers a positive dimension in this effort. These objectively measurable entities that are reflective of normal and pathologic states offer solutions to early detection and prevention of cancer. Additionally, the presence of affordable efficacious pharmaceutical agents should help with early effective disease control and improved quality of life. Thus, the ability to develop valid, safe, and effective biomarkers is commendable. Such valid biomarkers are certifiable by the Clinical and Laboratory Standards Institute for clinical applications and are very useful cost-effective means for accelerating the drug development process.

The word biomarker has a very broad meaning. However, with the completion of the human genome project and development of advanced technologies, the importance of the molecular genetic landscape of cancer is being appreciated at a much higher resolution. The ability to identify, validate, and fit these molecular biomarkers into routine laboratory medicine and molecular imaging holds tremendous value. The liquid biopsy concept ushers in a new era of practice. Not only does liquid biopsy fit into routine clinical practice, but it also overcomes tumor heterogeneity, as well as enabling the study of tumor evolution that is critical to developing novel effective therapeutic agents.

An important challenge to partly resolving the global cancer disparities is the ability to translate and impact the underserved with all these discoveries and developments. Advances in technology, including microfluidics and nanofluidics,

enable the development of miniaturized devices (the lab-on-a-chip concept) that can be deployed at remote settings. Coupling the use of such devices at the point of care with telemedicine should partly bridge the gap in global cancer statistics between the resource-rich and resource-poor communities.

This book covers the general principles of cancer biomarkers in body fluids and their potential clinical applications. Chapter 1 is the synthesis of the molecular pathology of cancer as the "treasure chest" of cancer biomarkers. Chapter 2 details the liquid biopsy phenomenon and cancer biomarkers. Chapter 3 provides an overview of some advanced technologies in cancer biomarker analysis. Chapter 4 addresses the general applications and analytical principles and challenges of assaying methylated circulating tumor genomes. Chapter 5 details the concept of circulating tumor DNA and the longitudinal analysis of cancer mutations in patient management. Chapter 6 critically examines the potential of cancer mitochondrial genome alterations in body fluids. Chapter 7 examines the principles of miRNA as noninvasive cancer biomarkers. The roles of circulating tumor cells and tumor-derived extracellular vesicles as cancer biomarkers that are being developed for clinical applications are detailed in Chapters 8 and 9, respectively. Finally, Chapter 10 examines the challenges to biomarker development and provides guidance on how to navigate the regulatory path to successful biomarker evaluation and use in the drug development process.

This book could not have been successfully completed without the help of many people that I must acknowledge. The diligent work of Mr. Srinivasan Manavalan and Ms. Sulata Kumari Nayak through the production process are gratefully appreciated.

I still wonder how my project coordinators, Ursula Gramm and Martina Himberger, managed to be so patient with me, for which I am so grateful. Finally, I will like to thank my loving and caring wife Crescentia and children Collins, Ethan, Bernard, and Zaneta for supporting me throughout this work.

Thunder Bay, Ontario, Canada Gabriel D. Dakubo
April 2016

Contents

1 Molecular Pathology of Cancer 1
1.1 Introduction 2
1.2 Functional Anatomy of the Cancer Cell 3
1.3 History of Cancer Molecular Pathology 4
1.4 Causes and Risk Factors of Cancer 5
1.4.1 Infectious Causes of Cancer and Circulating Biomarkers 6
1.5 The Biology of Cancer and Biomarkers 7
1.6 Mutations as Cancer Biomarkers 8
1.7 DNA Repair Mechanisms 9
1.7.1 Direct Repair System 9
1.7.2 Excision Repair 10
1.7.3 Repair of Double-Strand Breaks 11
1.8 Oncogene and Tumor Suppressor Gene Alterations as Biomarkers 12
1.9 Cancer Signaling Pathway Alterations as Biomarkers 13
1.9.1 The Receptor Tyrosine Kinase (RTK) Signaling Pathway 13
1.9.2 The Phosphoinositide 3-Kinase (PI3K) Signaling Pathway 14
1.9.3 The Mitogen-Activated Protein Kinase (MAPK) Signaling Pathway 16
1.9.4 The Mesenchymal-Epithelial Transition (MET) Factor Signaling Pathway 16
1.9.5 The Janus Kinase (JAK) Signaling Pathway 17
1.9.6 The Transforming Growth Factor-Beta (TGFβ) Signaling Pathway 19
1.9.7 The Nuclear Factor-kB (NF-kB) Signaling Pathway . . . 21
1.9.8 Src Signaling in Cancer Invasion 21
1.9.9 Control of the Cell Cycle 22

1.9.10 Hypoxia-Inducible Factor Signaling 23
1.9.11 The WNT/β-Catenin Signaling Pathway 25
1.9.12 The Hedgehog Signaling Pathway 27
1.9.13 The Notch Signaling Pathway 29
1.10 Cell Death Processes and Cancer 30
1.10.1 Nomenclature for Cell Death Processes 31
1.10.2 Apoptosis 31
1.10.3 Cancer Initiation as Defects in Apoptosis 33
1.11 Multistep Carcinogenesis and Early Detection Cancer Biomarkers 33
1.12 Personalized Cancer Care 34
1.12.1 Nomenclature of Targeted Therapies 37
1.12.2 Tests of Interest in Personalized Oncology 37
1.13 Epigenetic Modifications and Cancer 38
1.13.1 DNA Methylation 39
1.13.2 Histone Modification 39
1.13.3 The Cancer Cell Epigenome 40
1.14 Genetic Alterations in Cancer 42
1.14.1 Chromosomal Alterations 42
1.14.2 Gene Mutations 44
1.15 The Cancer Cell Transcriptome 46
1.16 The Proteome or Peptidome 48
1.16.1 Serum Peptidome 48
1.16.2 Serum vs. Plasma Proteomics 49
1.16.3 Issues with Proteomic Biomarker Studies 50
1.17 The Cancer Cell Metabolome 50
1.17.1 Body Fluid Metabolomics 52
1.18 Summary 52
References 53

2 Advanced Technologies for Body Fluid Biomarker Analyses 55
2.1 Introduction 56
2.2 High-Throughput Technologies for Biomarker Analyses 57
2.3 Next-Generation Sequencing 58
2.4 Single-Cell Analysis 60
2.4.1 Whole Genome Amplification 60
2.4.2 Whole Transcriptome Amplification 62
2.4.3 Whole Genome Analysis of Circulating Tumor Cells 63
2.5 Proteomic Technologies for Body Fluid Analysis 64
2.6 Gel Based (2D and DIGE) 65
2.7 Non-gel-based Proteomic Techniques 65
2.7.1 Mass Spectrometry 67
2.7.2 Labeling Techniques 68

2.8 Protein Microarrays . . . 69
2.9 Metabolomic Technologies . . . 70
2.9.1 NMR Spectroscopy . . . 70
2.9.2 Commercial Plasma Metabolomic Kits . . . 71
2.10 Lab-on-a-Chip . . . 71
2.11 Summary . . . 73
References . . . 74

3 Liquid Biopsy and Cancer Biomarkers . . . 75
3.1 Introduction . . . 76
3.2 Liquid Biopsy . . . 77
3.3 Biomarkers . . . 79
3.4 Cancer Biomarkers . . . 80
3.4.1 Features of an Ideal Cancer Biomarker . . . 81
3.4.2 Classification of Cancer Biomarkers . . . 82
3.4.3 Cancer Biomarker Development . . . 85
3.4.4 Challenges to Cancer Biomarker Development . . . 87
3.5 Evidence-Based Medicine . . . 88
3.6 Body Fluids . . . 89
3.6.1 Historical Perspective on Body Fluids and Cancer . . . 90
3.6.2 Issues with Cancer Screening . . . 91
3.6.3 Mechanism of Cancer Biomarker Release into Body Fluids . . . 93
3.6.4 Body Fluid Handling to Preserve Biomarker Integrity . . . 94
3.6.5 Precautionary Measures for Handling Body Fluids . . . 95
3.7 The Peripheral Circulation as a Repository of Cancer Biomarkers . . . 97
3.7.1 Clinical Utility of Biomarkers in Circulation . . . 98
3.8 Body Fluid Cancer Biomarkers and Point-of-Care Patient Management . . . 99
3.9 Traditional Circulating Tumor Markers in Clinical Practice . . . 100
3.10 Summary . . . 102
References . . . 102

4 Methylated DNA as Cancer Biomarkers in Circulation . . . 103
4.1 Introduction . . . 104
4.2 DNA Methylation and Cancer . . . 105
4.3 DNA Methylation and Molecular Field Cancerization . . . 106
4.4 Circulating MetDNA as Diagnostic Biomarkers . . . 108
4.5 Circulating MetDNA as Prognostic Biomarkers . . . 109
4.6 Circulating MetDNA as Treatment Response Biomarkers . . . 110
4.7 DNA Methylation and CTCs . . . 112
4.8 MetDNA in Other Body Fluids . . . 114
4.9 Methods for MetDNA Analysis . . . 115
4.10 Challenges with MetDNA Analysis in Body Fluids . . . 118

4.11 Commercial Products Based on MetDNA 119
4.12 Summary 120
References 121

5 Nucleic Acids as Cancer Biomarkers in Circulation 125
5.1 Introduction 126
5.2 Historic Recap of ccfDNA in Cancer 126
5.3 Features of Circulating Tumor Nucleic Acids 128
5.3.1 Mechanisms of Nucleic Acid Release 130
5.4 Features of Cell Surface-Associated Circulating Nucleic Acids 131
5.5 Features of Circulating Nucleosomes 132
5.5.1 Mode of Nucleosomal Release into and Elimination from the Circulation 132
5.5.2 Circulating Nucleosomes Mirror Caspase Activity . . . 133
5.6 Circulating Tumor Cells Contribute to Cell-Free Nucleic Acids 134
5.7 Pathogenicity of Circulating Cancer Nucleic Acids 134
5.8 Circulating Microbial Nucleic Acids 135
5.9 Pre-analytical and Analytical Issues with ccfDNA 136
5.9.1 Recommendations for Pre-analytical Sample Handling for ccfDNA Analysis 136
5.10 Translational Potential of Circulating Nucleosomes in Cancer . . . 138
5.10.1 Diagnosis 138
5.10.2 Prognosis 139
5.10.3 Therapy and Disease Monitoring 139
5.11 Translational Potential of Circulating Cell-Free (Unbound) DNA 140
5.11.1 Diagnosis 140
5.11.2 Prognosis 141
5.11.3 Therapy and Disease Monitoring 141
5.12 Translational Potential of DNA Integrity Index in Cancer 142
5.13 Translational Potential of Circulating Tumor-Specific DNA . . . 143
5.13.1 Circulating Tumor DNA Detection Methods 143
5.13.2 Early Detection 144
5.13.3 Prognosis 145
5.13.4 Treatment Prediction 145
5.14 Digital PCR Technology 146
5.15 BEAMing Technology 147
5.16 Summary 148
References 149

6 Noncoding Transcripts as Cancer Biomarkers in Circulation 153
6.1 Introduction 154
6.2 “Dark Matter” RNA 155

6.3 Biology of Noncoding Transcripts 155
6.3.1 Noncoding Nuclear Transcripts 155
6.3.2 Noncoding Mitochondrial Transcripts 156
6.4 MicroRNA 159
6.4.1 Historic Recap of MicroRNAs 160
6.4.2 MicroRNA and Cancer 160
6.4.3 Genomics of MicroRNAs 161
6.4.4 MicroRNA Biogenesis 162
6.4.5 Defects in MicroRNA Biogenesis in Cancer 164
6.4.6 MicroRNA Release and Uptake 165
6.4.7 MicroRNA Target Recognition and Gene Silencing . . . 167
6.4.8 MicroRNA Stability in Body Fluids 167
6.4.9 Physiologic and Pathologic Variations in Circulating MicroRNA 168
6.4.10 MicroRNA Alterations in Cancer 168
6.4.11 Oncogenic and Tumor Suppressor Functions of MicroRNAs 169
6.4.12 Precautionary Measures in MicroRNA Analyses . . . 171
6.5 Nonhuman MicroRNA in Circulation 172
6.6 Other Small ncRNAs in Circulation 173
6.7 Long Noncoding RNA in Cancer 173
6.7.1 Genomic Structure of LncRNAs 173
6.7.2 Classification of LncRNAs 174
6.7.3 Functions of LncRNAs 175
6.7.4 Expression of LncRNAs in Cancer 176
6.7.5 LncRNA Mutations in Cancer 177
6.7.6 LncRNAs in Cancer Biology 178
6.7.7 Circulating LncRNA in Cancer Patients 179
6.7.8 LncRNA in Cancer Tissue-Associated Body Fluids . . . 180
6.7.9 LncRNA in Cancer-Derived Extracellular Vesicles . . . 180
6.8 Summary 181
References 182

7 Mitochondrial Genome Changes as Cancer Biomarkers in Circulation 187
7.1 Introduction 187
7.2 The Mitochondrial Genome 188
7.3 The Electron Transport Chain and ROS Production 189
7.4 Mitochondrial Genetics and Cancer 192
7.4.1 Mitochondrial Dysfunction and Cancer Metastasis . . . 194
7.5 Mitochondrial Genome Changes in Body Fluids of Cancer Patients 194
7.5.1 Tumor mtDNA in Circulation 195
7.5.2 Tumor mtDNA in Tissue-Associated Body Fluids . . . 197
7.6 Summary 201
References 201

8 Circulating Tumor Cells and Cancer Stem Cells 205
8.1 Introduction 206
8.2 Thomas Ashworth and Circulating Tumor Cells 207
8.3 Metastasis and Formation of Circulating Tumor Cells 207
8.3.1 Basement Membrane Disruption 209
8.3.2 Invasiveness 210
8.3.3 Molecular Genetic Control of Metastasis 210
8.3.4 Epithelial-to-Mesenchymal Transition 213
8.3.5 Circulating Tumor Cell Formation Independent of EMT 215
8.4 Distant Site Selection: The Anatomic vs. Seed and Soil Hypotheses 215
8.5 Enrichment and Isolation of Circulating Tumor Cells 217
8.5.1 The Filtration Principle 218
8.5.2 Centrifugation Principle 220
8.5.3 Immunomagnetic Bead Principle 221
8.5.4 Other CTC Separation Methods 225
8.5.5 Detection and Characterization of CTCs 226
8.6 Circulating Cancer Stem Cells 227
8.7 Clinical Applications of CTCs 228
8.8 Summary 229
References 230

9 Extracellular Vesicles as Cancer Biomarkers in Circulation 233
9.1 Introduction 234
9.2 Cancer-Derived Extracellular Vesicles 235
9.3 Types of Extracellular Vesicles 235
9.4 Exosomes 237
9.4.1 Exosome Biogenesis 237
9.4.2 Exosomal Content 239
9.5 Cancer-Derived Microvesicles 239
9.5.1 Microvesicular Biogenesis 240
9.6 Mechanisms of Extracellular Vesicle Uptake by Target Cell . . . 240
9.7 Biological Functions of Cancer-Derived Extracellular Vesicles 242
9.7.1 Immunosuppressive Effects of Extracellular Vesicles 242
9.7.2 Horizontal Transfer of Mutated Genes or Proteins to Other Cells 243
9.7.3 Modulation of Tumor Microenvironment and Premetastatic Niche 244
9.7.4 Extracellular Vesicles as Circulating Cancer Biomarkers 246
9.7.5 Extracellular Vesicles as *Vehicles* and *Targets* for Cancer Treatment 247

9.7.6 Extracellular Vesicles as Mediators of Cancer Treatment Resistance . . . 249
9.7.7 Extracellular Vesicles Promote Cancer Cell Migration and Invasion . . . 249
9.7.8 Extracellular Vesicles in Cancer Stem Cell Biology . . . 250
9.7.9 Blockage of Extracellular Vesicle Release in Cancer Management . . . 251
9.8 Apoptotic Bodies . . . 251
9.9 Oncosomes . . . 252
9.10 Quality Standards in Analysis of Extracellular Vesicles . . . 253
9.11 Methods for EV Isolation . . . 254
9.11.1 Centrifugation . . . 254
9.11.2 Filtration . . . 255
9.11.3 Immunoaffinity . . . 255
9.12 Analyses of EVs . . . 255
9.13 Commercial Applications . . . 256
9.14 Summary . . . 257
References . . . 258

10 The Critical Path to Successful Biomarker Development . . . 261
10.1 Introduction . . . 262
10.2 Biomarker Development Challenges . . . 263
10.3 Biomarker Method Development and Evaluation . . . 267
10.3.1 Fit-for-Purpose Approach to Biomarker Method Validation . . . 267
10.4 Biomarkers and Surrogate Endpoints . . . 269
10.5 FDA and Biomarker Development . . . 269
10.6 The Problem with Successful Biomarker Development . . . 270
10.7 The Critical Path Initiative . . . 271
10.7.1 Product Development Tools . . . 272
10.8 Biomarkers in the Drug Development Process . . . 274
10.8.1 Utility of Biomarkers in Drug Development Programs . . . 275
10.8.2 Regulatory Requirements for Biomarker Use in Drug Development . . . 275
10.8.3 Use of Diagnostic Devices for Qualified Biomarkers . . . 276
10.9 Biomarkers and Qualification Process for Drug Development Tools . . . 277
10.9.1 Qualification . . . 277
10.9.2 The Qualification Process . . . 279
10.10 Evaluation of Biomarker Clinical Utility . . . 280
10.11 Biomarker Adoption into Clinical Practice . . . 281
10.12 Summary . . . 282
References . . . 283

11 Ethical Considerations in Biomarker Development 285
11.1 Introduction 286
11.2 Bioethics 286
11.3 The Cancer Patient and Research Participants 287
11.4 Biobanks 290
11.4.1 Custodianship 292
11.4.2 Privacy Protection 293
11.5 Ethical Concerns of Cancer Screening 294
11.6 Ethical Concerns of Cancer Diagnosis 295
11.7 Industry-Academia Relationships 296
11.8 Summary 298
References 299

Chapter 1
Molecular Pathology of Cancer

Key Topics

- Causes of cancer
- Molecular biology of cancer
- Mutations
- DNA repair mechanisms
- Oncogenes and tumor suppressor genes
- Signaling pathways in cancer
- Cell death processes
- Multistep carcinogenesis
- Personalized oncology
- Cancer cell epigenome, genome, transcriptome, and metabolome

Key Points

- The molecular pathology of cancer is a "treasure chest" of cancer biomarkers.
- Intricate molecular alterations of a cell orchestrate desired signaling networks to promote the development and spread of cancer. Thus, molecular pathogenesis of cancer offers a depth of knowledge on cancer biology that is paving the way for the identification and development of targeted therapies for personalized oncology.
- The "omics" sciences provide a global molecular view of the cancer cell. Coupled with advanced bioinformatics, this approach uncovers molecular profiles for cancer risk assessment, early detection, subclassification, and patient management.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_1

1.1 Introduction

Globally, cancer remains a primary cause of morbidity and mortality. For every four-recorded deaths, cancer accounts for one. Over half of the population will be afflicted by invasive cancer at some point in their lifetime. As an age-associated disease, this trend is likely to increase mirroring the increase longevity, if curtailment measures are ignored. In spite of improvements in detection and management, lung (1.59 million deaths/year), liver (745,000 deaths/year), stomach (723,000 deaths/year), colorectal (694,000 deaths/year), breast (521,000 deaths/year), and esophageal (400,000 deaths/year) cancers still account for a major proportion of cancer-related deaths.

The discipline of molecular pathology deals with the study of biomolecules in organs, tissues, cells, and body fluids of the infirm or diseased individual. Molecular pathology of cancer therefore focuses on the differential levels of biomolecules between cancer patients and control healthy individuals. The discipline employs molecular, genetic, and biochemical approaches in the study of biomolecules in clinical samples. In contrast to anatomic pathology, molecular pathology of cancer is a powerful approach to cancer studies that enables the discovery of altered molecules that impinge on signaling networks, which perpetuate the malignant phenotype. Molecular pathology has several clinical applications that range from disease prevention and companion diagnostics to personalized medicine and beyond. Various molecular approaches including the polymerase chain reaction (PCR), microarrays, mass spectrometry, biochips, nanotechnology, and microfluidics are used to detect, measure, and quantify biomolecules such as cells, subcellular particles, DNA and its epigenetic modifications, RNA, proteins, and metabolites (including lipids and carbohydrates) in disease states. The discipline of molecular pathology will revolutionize the field of medicine because it will enable early disease detection and chemoprevention, accurate diagnosis, and myriads of other clinical applications beyond what anatomic pathology can offer.

Well acknowledged and accepted by cancer researchers is the fact that cancer is an extremely diverse and complex disease. By and large, both exogenous and endogenous carcinogens target and alter the epigenome and genome of normal cells culminating in neoplastic growth. Recent advanced epigenomic, genomic, transcriptomic, proteomic, metabolomic (including lipidomics, glycomics), and nanoscale technologies, coupled with refined bioinformatics and systems biologic approaches, are deciphering a high-resolution map of the myriads of pathways evoked by the cancer cell. In addition, these endeavors are identifying subclasses of cancers that respond to specific therapies, thus leading the path toward individualized care or personalized medicine.

No specific individual cancer harbors monogenic alterations but often carries multiple epigenetic and genetic aberrations that interfere with multiple pathways. Thus, valid and broadly encompassing biomarkers of a particular disease may more likely involve gene or biomarker panels targeting multiple pathways.

The cancer cell must overcome multiple hurdles to sustain its existence. For example, it must evade proliferation and cell cycle control and apoptosis, exhibit abnormal DNA repair mechanisms, and overcome the paucity of blood vessels, poor cellular differentiation (remain progenitor-like), senescence, immune control, and various therapeutic interventions. Thus, Douglas Hanahan and Robert Weinberg [1] identified the six hallmarks of cancer to include the following:

- Lack of reliance on growth signals.
- Insensitivity or lack of response to antitumor growth signals.
- Refractoriness to apoptotic signals.
- Lack of response to antiproliferative signals (i.e., acquired limitless replicative potential phenotype).
- The ability to build new blood vessels (sustained angiogenesis).
- The capability to invade and metastasize without anoikis.

These changes are also influenced by metabolic alterations of the cancer cell. This chapter provides a synopsis of the numerous molecular alterations in a cell that culminate in its transformation into a cancer cell and eventual progression to acquire invasive phenotypes in the multistep carcinogenesis process. The numerous molecular and genetic aberrations serve as biomarker repositories for discovery, evaluation, and product development for clinical applications.

1.2 Functional Anatomy of the Cancer Cell

The cancer cell is a novel cell in the body, having acquired anatomic, physiologic, and biochemical alterations at various levels. These changes range from its membrane to organellar structure, as well as biomolecules including nucleic acids, proteins, lipids, carbohydrates, and other metabolites.

Cancer cells have distinct morphology that enables histologic diagnosis. These include the presence of a large nucleus that may be irregular in shape and size, with a prominent nucleolus. The large nucleus in association with a scare cytoplasm leads to a decreased cytoplasm-to-nuclear ratio. The cytoplasm may be basophilic due to increased levels of rRNA and mRNA. There are numerous organellar changes associated with malignant transformation. Mitochondria show variable shapes and volumes, with some being very large. The Golgi apparatus lacks development as a result of the poor differentiation state of the cancer cell. The rough endoplasmic reticulum may be fragmented with degradation. The decreased rough endoplasmic reticulum is associated with increased polyribosomes and free ribosomes needed for synthesis of proteins to be used for cancer cell proliferation and growth. The smooth endoplasmic reticulum may be hyperplastic initially, but then becomes reduced with tumor progression. The levels of glycogen are high initially, but are subsequently used up by the glycolytic cancer cell. Thus, low glycogen in association with increased lipid levels is a cancer cell phenotype. A few other examples of the altered cancer cell anatomy are illustrated here.

There are numerous alterations to lipid membranes of cancer cells. In hepatoma cells, membrane lipids and fatty acid composition of phospholipids are altered compared to normal liver cells. Exemplary alterations include the following:

- Cholesterol levels are low in cancer cells, but high in preneoplastic hyperplastic nodular cell membranes.
- The relative amounts of choline phospholipids are low in membranes of cancer cells.
- Total amounts of short fatty acyl chains (<18 carbons) are increased in preneoplastic nodules and hepatomas compared with normal hepatocytes. Consistently, long fatty acyl (>20 carbons) chains are reduced in these lesions.
- Saturated fatty acid levels are reduced, while unsaturated fatty acid levels are high in membranes from hepatomas and preneoplastic lesions.
- Arachidonic acid levels are lower in cancer than in normal cell membranes.

These lipid profiles reflect on plasma membrane fluidity that is determined by the lipid composition and levels of cholesterol. The high levels of short fatty acyl chains and unsaturated fatty acids in association with low cholesterol content increase the fluidity of cancer cell membranes. However, the high cholesterol content in preneoplastic nodules rather increases the rigidity of their membranes.

There are numerous protein changes as well with malignant transformation. Changes in numerical densities and topographic distribution of integral membrane proteins (IMPs) are associated with malignant transformation as demonstrable in both tissue culture models and solid tumors. The amount of IMPs is increased in noninvasive tumors, but reduced in invasive cancers. Topographical distribution of IMPs is also altered in these cancers. The proteins are mildly aggregated in nonneoplastic epithelial cells. However, aggregation is increased in noninvasive cancer cells. But invasive cancers have random distribution of IMPs. The possible explanation for the randomness of IMPs in the cancer cell membrane includes loss of IMP interactions with cytoskeletal elements (maybe due to loss of E-cadherin junctions), enabling free lateral motility to achieve random distribution. Also, plasma membrane fluidity is increased in cancer cells, which could enhance free and random movement of IMPs.

Many of these features are responsible for the histopathologic alterations associated with the cancer cell. The "cancer anatomic" features also reflect on the biomolecular changes that give rise to cancer biomarkers.

1.3 History of Cancer Molecular Pathology

Carcinogenesis involves acquisition by a cell of heritable epigenetic and genetic traits, followed by expansion and natural selection of some cell populations. Some clones perish from harboring deleterious mutations, while others with alterations that favor proliferation, growth, and survival or evasion from apoptosis remain and acquire several more mutations to fully manifest the malignant phenotype.

Currently, over 100,000 somatic mutations in cancer genomes have been reported and catalogued. Efforts by a number of multinational initiatives including the International Cancer Genome Consortium (ICGC), The Cancer Genome Atlas (TCGA), and OncoArray Network over the next few years should expand this number. The ICGC aims to characterize extensively, the somatic mutational spectrum in over 50 classes of cancer, and integrate this data with epigenetic and expressional changes, as well as clinical information of the patients. Efforts are also directed, albeit with some challenges, at teasing out cancer causing or "driver" from inert or "passenger" somatic mutations in cancer genes.

These current endeavors make it very exciting to envision the future of cancer genomics, but how we got here is sobering indeed. The disease cancer has been known since 3000 BC, with the description of the Edwin Smith Papyrus in an ancient Egyptian textbook on trauma and surgery. However, until the late nineteenth and early twentieth centuries, cancer as a genetic disease was unknown. Pioneering works by Theodor Boveri, Peter Nowell, and David Hungerford first revealed the genetic nature of cancer. Microscopic analysis by these pioneers revealed chromosomal aberrations in dividing cancer cells. Consequently, genomic damage as a cause of cancer was proposed.

That DNA damage underlies carcinogenic processes came after the elucidation of the DNA structure and as the genetic material. Shortly after the terms mitosis and chromosomes were first introduced, von Hansemann, a German pathologist, in 1890 observed mitotic abnormalities in cancer tissue sections. A few decades later, in 1914, Boveri began formulating hypotheses about how somatic genetic changes could cause cancer [2]. He hypothesized that tumors might be initiated by mitotic abnormalities that led to numerical chromosomal alterations (aneuploidy). This thesis was subsequently proven with the hard work of Peter Nowell and David Hungerford on leukemic cells [3] and named by Tough as the Philadelphia chromosome (associated with chronic myeloid leukemia).

Classic demonstration that genetic alteration caused cancer in normal NIH3T3 cells also led to the historic identification of the first cancer single nucleotide mutation in codon 12 of the HRAS gene in 1982 [4]. Today, several cancer mutations are being deciphered and catalogued. The search continues for more. The implications of these mutations for personalized medicine are enormous. Some applications of these developments are already being realized in the clinic.

1.4 Causes and Risk Factors of Cancer

Cancer is a genetic disease, induced either by endogenous or exogenous carcinogens. In a minority of cases, these genetic changes are in the germline and passed on from generation to generation. These germline mutations constitute the heritable cancer syndromes. Also well known is that cancer is a disease of the aging process, thus making aging an unmodifiable risk factor for the development of cancer. One explanation for the age association could be the adequacy of time available for the

Table 1.1 Agents classified by the International Agency for Research on Cancer (IARC) as posing carcinogenic risk to humans

Group	Risk level	Number	Examples
1	Carcinogenic to humans	118	Tobacco smoke, wood dust, alcohol, UV radiation, solar radiation, X- and gamma radiation, radioiodine, soot, estrogen-progestogen oral contraceptives
2A	Probably carcinogenic to humans	70	Red meat, petroleum refining, biomass fuel, anabolic steroids
2B	Possibly carcinogenic to humans	290	Pickled traditional Asian vegetables, gasoline, coffee, progestogen-only contraceptives
3	Not classifiable as to its carcinogenicity to humans	501	Magnetic fields, electric fields, fluorescent lightning, coal dust, chloroquine, prednisone
4	Probably not carcinogenic to humans	1	–

Source: IARC monographs on the evaluation of carcinogenic risks to humans

cell to acquire the necessary complement of mutations to be transformed. Lifestyle exposures including tobacco use, alcohol abuse, radiation exposure, consumption of unhealthy diets such as those heavy in saturated fat or red meat, excessive weight gain and obesity, physical inactivity, and sedentary lifestyle are all established risk factors for the development of cancer. Exogenous environmental exposures including occupational carcinogens (e.g., toxic chemicals and radiation exposure) and even pharmaceuticals can all be triggers of cellular transformation (Table 1.1). Finally, certain infections (see below) are causative agents of cancer.

1.4.1 Infectious Causes of Cancer and Circulating Biomarkers

Among the most important discoveries in medicine is the demonstration that some infectious agents can cause cancer. In fact, infectious agents account for a considerable proportion (~20 %) of all human cancers. Of note, a large percentage of these cancers occur in the less developed world. Also noteworthy is the fact that the majority of these cancers are preventable through early diagnosis and simple antibiotic treatment of the infection or immunization of targeted population at risk. While these infectious agents are mostly viruses, bacterial and parasitic agents account for some cancers as well (Table 1.2). Viruses appear to use molecular genetic mechanisms, such as deregulation of cell cycle through, for example, HPV oncoproteins, to initiate cellular transformation. However, chronic inflammation associated with increased redox species production may underlie cancer initiation by some other organisms. Cancer preventive measures, such as screening and treatment of infections, vaccination, and other public health interventions (e.g.,

Table 1.2 Infectious agents that cause cancer

Infectious agents	Associated cancers
Viruses	
HPV	Cervical cancer, head and neck cancer Also vulva, vaginal, penile, and anal cancers
HBV	Liver cancer
EBV	Nasopharyngeal cancer, lymphomas (Burkitt's lymphoma, Hodgkin lymphoma)
HIV	Kaposi sarcoma, cervical cancer, non-Hodgkin's lymphoma
HHV8	Kaposi sarcoma
HTLV1	Adult T-cell leukemia/lymphoma (ATL)
MCV	Merkel cell carcinoma of the skin
Bacteria	
Helicobacter pylori	Gastric cancer, gastric lymphoma
Chlamydia trachomatis	Cervical cancer
Parasites	
Schistosoma haematobium	Urothelial bladder cancer
Opisthorchis viverrini, *Clonorchis sinensis*	Cholangiocarcinoma

HPV human papillomavirus, *HBV* hepatitis B virus, *EBV* Epstein-Barr virus, *HIV* human immunodeficiency virus, *HHV8* human herpes virus 8, *HTLV1* human T-lymphotropic virus 1, *MCV* molluscum contagiosum virus

safe sex), have been instituted to help curtail the incidence and prevalence of such cancers in the community. These infections are associated with defined biomarkers in circulation and other body fluids that can be assayed noninvasively for early detection. For example, Epstein-Barr virus (EBV) infection is associated with circulating exosomes loaded with viral associated miRNA that can easily be detected.

1.5 The Biology of Cancer and Biomarkers

Understanding of basic cancer biology and genetics is important to appreciating biomarkers and how they influence cancer development and management. All the molecular and genetic lesions that a cell sustains, which impinge on alterations of many signaling networks leading to the eventual transformation of the normal cell, and subsequent maintenance of the malignant and invasive phenotype, are associated with a trail of imprints and footprints that constitute cancer biomarkers. The functional roles of these biomarkers are paving the way for deeper understanding of cancer biology and hence the deployment of effective therapeutic approaches.

1.6 Mutations as Cancer Biomarkers

Somatic and germline mutations are causative factors of cancer. Gene mutations and genomic instability are easily measured or assayed and are offering new insights into cancer development, management, therapy, and recurrence. At the epidemiologic level, genome-wide association studies (GWAS) enable interrogation of numerous genetic variants in the population to tease out risk association alleles for specific diseases. Primarily focused on SNP loci, GWAS is a powerful approach to risk identification for effective primary cancer preventive measures.

Mutations are bedrock of carcinogenesis because they cause changes in normal gene functions. For example, mutations in proto-oncogenes, tumor suppressor genes, and/or DNA repair genes can lead to genomic instability and further accumulation of more mutations. Mutations in a single cell if not repaired can be propagated through cell division cycles to amplify such clones. Some daughter cells may acquire other mutations, such that those that accumulate sufficient enabling mutations become transformed. Mutations can cause unregulated cell divisions and/or escape from cell death. It is estimated that a cell requires 6–10 enabling mutations to become cancerous.

The mutations in cancer can be either "driver" mutations, which actively participate in the cellular transformation process, or inert "passenger" mutations ("hitchhikers") that are associated with the "driver" mutations, but can be assayed as biomarkers as well. Many cancer mutations are somatic gene alterations, but germline mutations associated with hereditary or familial cancer syndromes are of equal importance to cancer biology.

The causes and propagation of somatic cancer mutations can be internal and/or external mutagens. Mutation rate (the rate at which mutations occur in the cell over a period of time, e.g., per generation) also may differ between cell types and among different mutations. For example, the mutation rate may be lower in hepatocytes because they express enzymes and other scavenger molecules that can inactivate mutagens, than in cells such as fibroblasts that do not have such efficient ability to inactivate mutagens. Here are some features of mutations and their roles in carcinogenesis:

- A cell with continuous and efficient DNA repair mechanisms (as are many normal cells) can offset genomic instability from mutations.
- Some mutations, just a handful, are often not repaired and remain as fixed mutations.
- DNA replication can contribute to mutations in a cell due to the occurrence of small but important errors that occur during the process.
- Mutation rates in some types of cancers mirror the levels of mutagens, e.g., lung cancer and tobacco smoke carcinogens.
- Some mutagens can cause specific mutations in genes (e.g., the frequent G to T transversions in *TP53* mutational hotspots (codons 248 and 273) caused by tobacco carcinogens).

- Mutation rates can increase substantially in some cancers as a result of inefficient DNA mismatch repair.
- The "mutator" phenotype (e.g., due to *MLH1*, and *MSH2* defects from epigenetic silencing, mutations, or expressional abnormalities) is associated with increased mutation rates in some cancers such as colorectal and endometrial cancers.
- Telomere attrition can cause a burst of mutations in cells leading to what is referred to as "mutation crisis." Such mutations are sufficient to initiate cancer formation.
- Many mutations are a function of chemical and physical carcinogens. The chemical pro-carcinogens may even be in the normal diet as proposed by Bruce Ames [5].
- Indeed, some of the supposedly healthy diets such as celery contain carcinogens such as furocoumarins and psoralens [5]. Other pro-carcinogens can be converted to carcinogens by bacteria, liver enzymes, or through the cooking process. For example, cooking of red meat produces heterocyclic amines that are potentially mutagenic.

Body fluid assay of cancer mutations is potentially powerful for cancer risk assessment, early detection, diagnosis, and follow-up management by tracking mutational evolution over time, such that therapies can be adapted appropriately. These clinical applications have been explored in many cancers.

1.7 DNA Repair Mechanisms

Genomic integrity is critical to normal cell functions and survival. DNA damage is, however, an inevitable occurrence in the human genome that occurs spontaneously, during replication or induced by mutagens. To offset genomic instability, efficient systems are in place to repair any DNA damage. Irrespective of the causative factors of DNA damage, there are evolutionary conserved mechanisms for DNA repair. Failure of any part of this system can result in human diseases such as xeroderma pigmentosum and ataxia-telangiectasia, which have a high predisposition to cancer.

The many mechanisms involved in DNA repair can be broadly categorized as (1) those involved in direct repair or reversal of damage and (2) the excision repair systems, which are probably the major players in repair of most genomic damage.

1.7.1 Direct Repair System

Some DNA damage can be repaired by direct reversal of the damage. The common lesions that are repaired by direct reversal are pyrimidine (thymine) dimers (results in formation of cyclobutane rings) from UV damage and modification of alkylated

guanine by the addition of methyl group at the O^6 position of purines. Purine methylation at O^6 can result from chemotherapy with alkylating agents such as carmustine, dacarbazine, and temozolomide.

UV-induced pyrimidine dimers can be repaired by direct reversal in a photoreactivation process. This mechanism operates in prokaryotes and some eukaryotes but not in humans. A relevant direct reversal system in humans involves repair of O^6 methylation of guanine to form O^6-methylguanine. The O^6-methylguanine is recognized by DNA polymerase as adenine instead of guanine, which can lead to a G:C to A:T transition mutation. The O^6-methylguanine-DNA methyltransferase (MGMT) enzyme removes the alkyl group returning it to guanine. The mismatch excision repair system can repair the transition mutations that occur due to the failure of MGMT activity. Loss of MGMT functions from mutations or promoter methylation resulting in oncogene (e.g., *RAS*) and tumor suppressor gene (e.g., *TP53*) mutations is common in a variety of cancers including lung, head and neck, esophageal, gastric and colorectal cancers, gliomas, and diffuse large B-cell lymphomas.

1.7.2 Excision Repair

Excision repair is the major DNA repair mechanism in humans. The process involves removal of the damaged DNA as freebase or nucleotide, with subsequent closure of the gap through synthesis of new DNA by polymerase and sealed by DNA ligase. There are three excision repair mechanisms, (1) base-excision repair (BER), (2) nucleotide-excision repair (NER), and (3) mismatch repair (MMR).

1.7.2.1 Base-Excision Repair

The base-excision repair (BER) system functions to repair single-strand breaks from deamination, oxidation, or effects of alkylating agents. These commonly occur in the genome from induction by ROS, UV radiation, and chemotherapy with alkylating agents. The abnormal bases including uracil, hypoxanthine, and O^6-alkylguanine can be removed by DNA glycosylase. The removal of the base leaves an apurinic or apyrimidinic (AP) site on the DNA. AP endonuclease creates a nick in the phosphodiester backbone at the AP site, and the deoxyribose is removed. The gap is then sealed by the activities of DNA polymerase and ligase.

Expectedly, several cancers harbor alterations including mutations and promoter hypermethylation in genes involved with BER. For example, 8-oxoguanine DNA lesions caused by ROS are repaired by 8-oxoguanine glycosylase (*OGG1*), which is mutated in several cancers including breast, lung, and thyroid cancers. Other BER genes silenced in cancer include thymine DNA glycosylase (*TDG*) in multiple myeloma, X-ray repair cross-complementary gene 1 (*XRCC1*) in lung cancer, uracil

DNA glycosylase (*UNG2*) in gliomas, mediator complex subunit 1 (*MED1*) in colorectal cancer, and Werner syndrome (*WRN*) gene in multiple cancers.

1.7.2.2 Nucleotide-Excision Repair

Nucleotide-excision repair (NER) targets carcinogenic effects on widespread or large-scale nucleotides. It is also involved in the repair of pyrimidine dimers in humans. The damaged DNA is removed as an oligonucleotide. The NER system has been extensively studied in xeroderma pigmentosa. Mechanistically it involves cleavage on the 3′ and 5′ sides of the lesion with removal of 12 or 13 bases by a helicase. The gap is then filled and sealed by DNA polymerase and ligase.

Several genes involved in NER are inactivated in a number of cancers. There are seven xeroderma pigmentosa genes (*XPA* to *XPG*). Several of these genes are mutated in xeroderma pigmentosa and in some cases of Cockayne's syndrome and trichothiodystrophy. Additionally, *XPC* is inactivated in lung, bladder, and ovarian cancers. Other NER genes silenced in cancer include the *RAD23 homolog B* (*hHR23B*) in multiple myeloma and excision repair cross-complementation group 1 (*ERCC1*) in gliomas.

1.7.2.3 Mismatch Repair

Base incorporation errors can occur during DNA synthesis. Similarly base mispairing can occur through effects of alkylating agents. Mismatched bases that occur during replication are normally recognized and removed by DNA polymerase using its proofreading activity. However, other mismatched bases escape the recognition by the polymerase, and others occur via other chemical reactions. These mispaired bases in the genome are recognized by the scanning activities of the mismatch repair system, excised, and removed such that the correct base can then be incorporated. The MMR system includes *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *PMS1*, and *PMS2*. Failure of this system can lead to accelerated mutation rate up to 1000-fold, which can cause microsatellite instability in the cell. Mutations and associated microsatellite instabilities are common in colorectal cancer. Also, inactivation of MLH1 is associated with breast, lung, head and neck, esophageal, gastric, pancreatic, and ovarian cancers.

1.7.3 Repair of Double-Strand Breaks

Although rare, double-strand breaks (DSBs) which are simultaneous damage to both strands of DNA occur. Agents such as the topoisomerase inhibitor, etoposide, and ironizing radiation can cause DSBs. Unlike single-strand breaks (SSB), DSBs are more difficult to repair, requiring a multitude of factors and conditions. DSBs

are very toxic to the cell if unrepaired and may cause chromosomal rearrangement defects if not properly repaired. Two different mechanisms, homologous recombination (HR) and non-homologous end-joining (NHEJ), mediate the repair of DSBs. Homologous recombination repair mechanisms require RAD52/BRAC2 proteins, as well as RAD51 and its paralogs (RAD51BCD, XRCC2, XRCC3). Regarding cancer, the relevant repair proteins include the Fanconi anemia (FA), BRCA, and XRCC5 proteins, all of which show some frequencies of inactivation in a wide variety of cancers. Thirteen FA genes are involved in the HR repair mechanisms, of which *BRCA2* (*FANCD1*) is a component. Germline inactivation of *BRCA1* and *BRCA2* coffers a lifetime breast cancer and ovarian cancer risks of 60–80 % and 20–40 %, respectively. Somatic *BRCA1* and *BRCA2* mutations are associated with sporadic breast, ovarian, lung, gastric, pancreatic, colorectal, and bladder cancers. *FANCF* is inactivated at various frequencies in breast, lung, head and neck, gastric, ovarian, and cervical cancers. *XRCC5* is inactivated in NSCLC, while *FANCC* and *FANCL* are inactivated in leukemia.

1.8 Oncogene and Tumor Suppressor Gene Alterations as Biomarkers

Robert Huebner and George Todaro introduced the term "oncogene" in 1969, and the first oncogene, *src*, was uncovered by G. Steve Martin and sequenced by P. Czernilofsky in 1980. Then, in 1976, J. Michael Bishop and Harold E. Varmus, the 1989 Nobel Laureates in Medicine, discovered proto-oncogenes. Proto-oncogenes are normal human genes, but, when mutated, usually produce proteins referred to as oncogenes that promote tumor formation and progression. These mutations or changes in genes cause their overexpression in many cancers. They target oncogenic pathways to drive cancer progression. While their exact numbers are not established, it appears there are dozens in the human genome. Some of the well-known proto-oncogenes are *V-src*, *RAS*, *MYC*, *Bcr-Abl*, *WNT*, *ERK*, *RTK*, and several growth factors. *V-src* oncogene and its oncoprotein, Src, were first to be cloned and demonstrated to function as a tyrosine kinase.

Tumor suppressor genes (TSGs), on the contrary, function to prevent the development of cancer. Hence, their functions are often lost through epigenetic silencing, mutations, loss of heterozygosity (LOH), or microsatellite alterations and are therefore often downregulated in many cancers. The lost of TSG functions may lead to cancer development and progression. There are ~716 human TSGs, including well-known ones such as *TP53*, *RB*, *APC*, *VHL*, *BRCA1*, and *BRCA*2. Many of these proteins function to inhibit cell cycle progression and/or promote cell death to curtail propagation of deleterious mutations. Specifically, they serve as cell cycle checkpoint controllers. They can repress the expression of genes involved in driving the cell cycle. Following the S-phase of the cell cycle (DNA replication or synthesis phase), DNA mutations may be introduced into the newly synthesized

molecules. Should this be the case, TSGs, specifically p53, can halt the cell cycle to allow for efficient DNA repair before cycle progression. However, should the repair mechanisms fail, p53 can instead induce genes to kill the cell. This serves to prevent propagation of cells with deleterious mutated genes that are more likely to progress to cancer. Some TSGs such as *CDH1* prevent cancer cell invasion and spreading by enhancing cell adhesion and inhibiting contact inhibition by these cells.

The phrase "oncogene addiction" introduced by Bernard Weinstein in 2008 [6] suggests that cancer cell growth, survival, and metastasis are under the critical control of oncogenic activity and/or loss of tumor suppressor gene functions. This established phenomenon of cancer is ushering in the era of targeted biotherapies. While laudable, the intricate signaling networks perpetuating "oncogene addiction" sometimes make it formidable to effectively target a particular oncogenic pathway. The myriads of oncogenes and TSG alterations in cancer are assayed in cancer tissues, cells, and body fluids as cancer biomarkers.

1.9 Cancer Signaling Pathway Alterations as Biomarkers

In normal growth control of the metazoan, cells communicate akin to quorum sensing in bacterial colonies, via growth factors that are interpreted as to whether cell growth and proliferation vs. differentiation control is appropriate for the organism. The cancer cell as a consequence of altered expression, and induction of growth factor signaling pathways, abrogates this tight-knit growth control to sustain its selfish existence. A synopsis of some important signaling pathways altered in cancer cells and that present opportunities for biomarkers and therapeutic targets are provided herein.

1.9.1 The Receptor Tyrosine Kinase (RTK) Signaling Pathway

Growth factors play important roles in cancer biology (Fig. 1.1). Many growth and survival factors including FGF, EGF, PDGF, FGF, HGF, IGF, TGF, VEGF, and cytokines are overexpressed in a number of cancer cells. Several of these growth factors interact with their cognate receptors, which are transmembrane proteins, composed of extracellular, transmembrane, and cytoplasmic tyrosine kinase domains. Ligand and receptor interaction leads to cytoplasmic signaling cascade that impinges on multiple oncogenic signaling pathways including the MEK/ERK/MAPK, PI3K/AKT, PLCγ, and JAK/STAT pathways, among several others. The output of these effector signals involves altered cellular functions including increased cell cycle progression, cellular proliferation, growth and survival, as well as cell migration, invasion, metastasis, and angiogenesis.

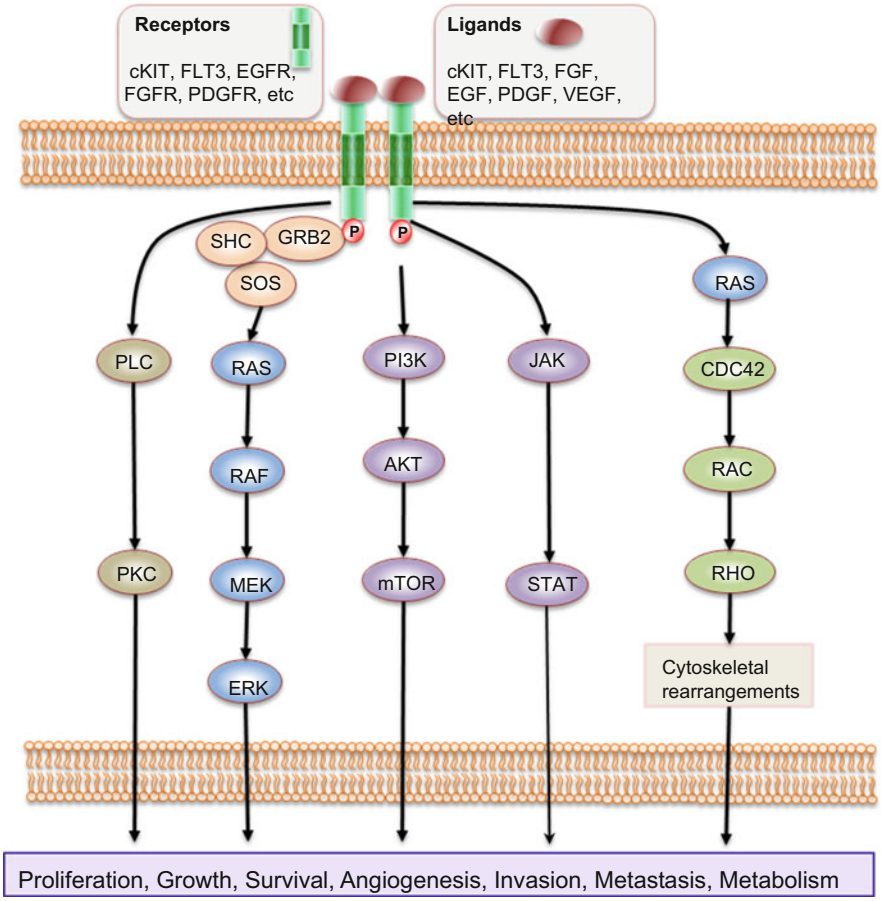

Fig. 1.1 The RTK signaling pathway. Receptor-ligand interaction causes RTK autophosphorylation leading to RAS and a number of signaling pathway activation. RAS activation involves the recruitment of SOS and GRB2 to tyrosine phosphate docking sites on the receptor. Activated RAS induces RAF-MAPK signaling, activated AKT induces mTOR, and PLCγ activation mobilizes $Ca2^{+}$ to activate PKC

1.9.2 The Phosphoinositide 3-Kinase (PI3K) Signaling Pathway

The PI3K and protein kinase B (PKB/AKT) signaling pathway regulates important cancer biologic processes, such as cellular growth, proliferation, survival, and metabolism (Fig. 1.2). The pathway is activated by myriads of factors including growth factors (e.g., via RTK signaling), cytokines, G-protein-coupled receptors, as well as loss of tumor suppressor gene, *PTEN*, which usually dephosphorylates an important pathway molecule, phosphatidylinositol-3-4-5-triphosphate (PIP3), thereby inhibiting signal transduction.

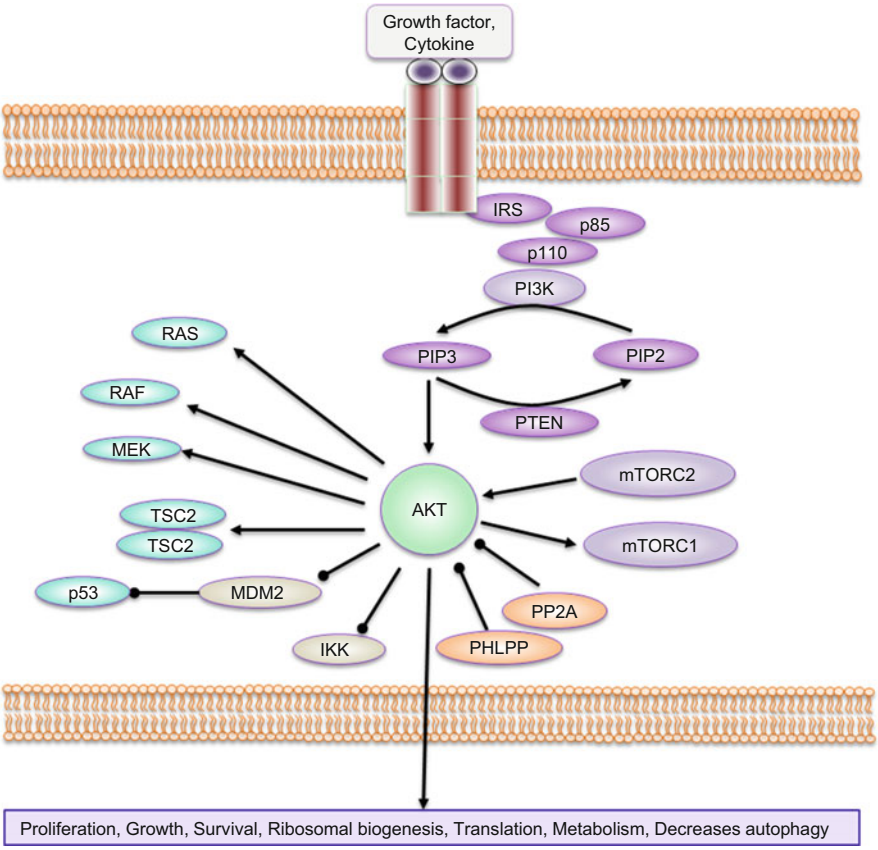

Fig. 1.2 The PI3K-AKT signaling pathway. Upon receptor (RTK, cytokine receptors, and others) activation, PI3K class IA isoforms (p110α, p110β, and p110δ) cause phosphorylation of phosphatidylinositol-3-4-biphosphate (PIP2) to phosphatidylinositol-3-4-5-triphosphate (PIP3), which activates 3-phosphoinositide-dependent protein kinase 1 (PDPK1 or PDK1). Activated PDK1 then phosphorylates and activates AKT. Activated AKT modulates multiple targets including mTOR in mTORC1

Ligand-receptor interaction leads to PI3K-mediated formation of PIP3. PIP3 creates docking sites for pleckstrin homology (PH) domain-containing proteins such as 3-phosphoinositide-dependent protein kinase 1 (PDK1), which in turn partly activates AKT/PKB via threonine-308 phosphorylation. Mammalian target of rapamycin complex (mTORC) and PI3K-related kinase (PIKK) family members subsequently phosphorylate serine 473 to fully activate AKT. The pathway has many negative regulators. PTEN, for instance, antagonizes AKT by dephosphorylating PIP3, while protein phosphatase 2A (PP2A) and PH-domain leucine-rich repeat-containing protein phosphatases 1 and 2 (PHLPP1/2) can directly dephosphorylate AKT to prevent pathway progression.

1.9.3 The Mitogen-Activated Protein Kinase (MAPK) Signaling Pathway

The mitogen-activated protein kinase (MAPK) signaling pathway also controls cellular growth, proliferation, apoptosis, and differentiation. Growth factors, mitogens, stress signals, cytokines, and G-protein-coupled receptor agonists activate this pathway. There are three established pathways that have distinct and overlapping functions. However, in general, signaling through each pathway involves sequential activation of MAPK kinase kinase (MAPKKK), MAPK kinase (MAPKK), and finally MAPK (Fig. 1.3). The ERK1/2 pathway that is deregulated in many malignant cells causes increase cell growth, survival, and differentiation. Stress and cytokine stimuli induce the activation of the JNK and p38 pathway, which mediate inflammation, growth, apoptosis, and differentiation. Finally, signaling via ERK5 has identical effects on cell growth and differentiation and is physiologically important in developmental biology.

1.9.4 The Mesenchymal-Epithelial Transition (MET) Factor Signaling Pathway

The MET signaling pathway is aberrantly activated in cancer and plays an important role in cancer biology (Fig. 1.4). This pathway integrates multiple downstream oncogenic signals including the PI3K, MERK/ERK, STAT, β-catenin, EGFR, VEGFR, and the NOTCH pathways, to induce sustained cell growth, proliferation, survival, migration, differentiation, angiogenesis, and epithelial-to-mesenchymal transition. MET is a hepatocyte growth factor (HGF), also known as scatter factor (SF) receptor. It is an RTK that interacts with its only known ligand, HGF/SF, to orchestrate myriads of intracellular programs involved in carcinogenesis. HGF-MET interaction leads to receptor homodimerization and phosphorylation of tyrosine residues (Y1234 and Y1235). Later, Y1349 and Y1356 get phosphorylated as well, causing the formation of tandem SH2 recognition motif. This motif orchestrates the various downstream functions by its ability to recruit several adaptor signaling biomolecules including growth factor receptor-bound protein 2 (GRB2)-associated binding protein 1 (GAB1), Src homology-2-containing (SHC), v-crk sarcoma virus CT10 oncogene homolog (CRK), CRK-like (CRKL), PI3K, PLCγ, v-src sarcoma viral oncogene homolog (SRC), STAT3, son of sevenless (SOS), and Src homology domain-containing 5′ inositol phosphatase (SHIP2).

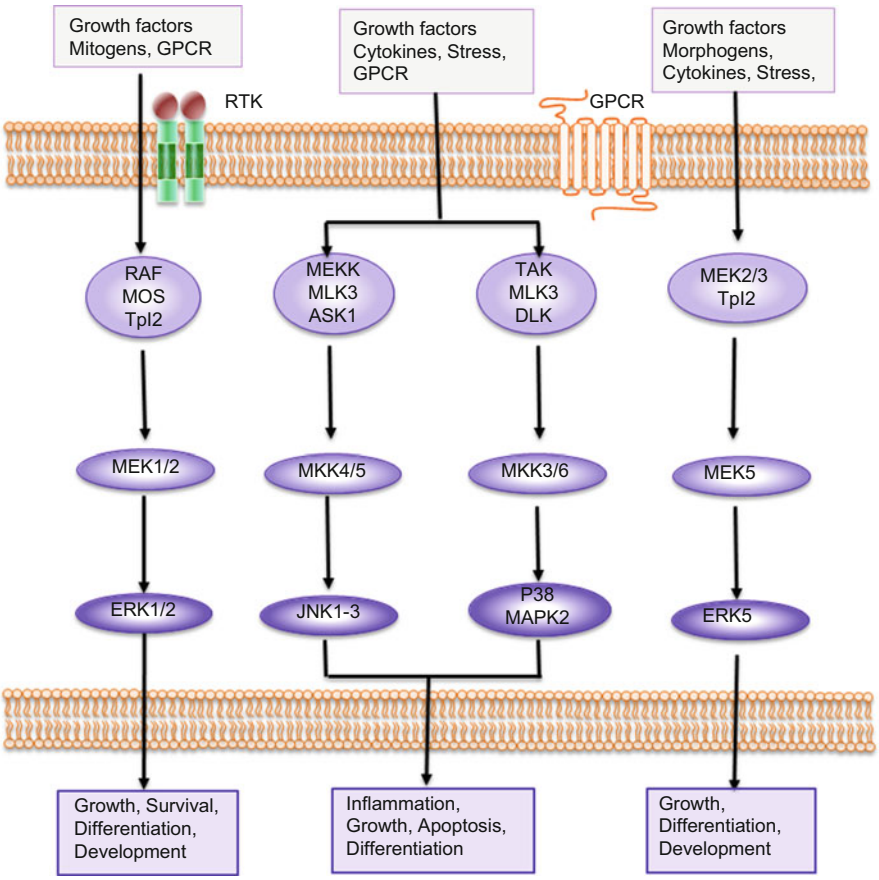

Fig. 1.3 The MAPK signaling pathway. This pathway can be activated by growth factors, inflammatory cytokines, and environmental stress such as heat, toxins, and radiation. Classic oncogenic signaling via RAS leads to induction of RAF, with phosphorylation and activation cascade involving MEK1 and MEK2, then ERK1 and ERK2. Phosphorylated and activated ERK enters the nucleus to induce expression of oncogenes and repression of tumor suppressor genes. Activated RAF can also signal through MKK3/6-p38, MKK4/7-JNK, and MEK5-ERK5 signaling pathways

1.9.5 The Janus Kinase (JAK) Signaling Pathway

The JAK and signal transducer and activator of transcription (STAT) signaling pathway is commonly deregulated in hematologic malignancies and some solid tumors. This pathway controls cell growth, survival, differentiation, and immune functions. Cytokine binding to their receptors causes receptor dimerization leading to the associated JAK activation (Fig. 1.5). Activated JAK cross phosphorylates tyrosine residues on each other, as well as those on the receptors. These phosphorylated sites enable SH2-containing molecules including STATs to dock and be

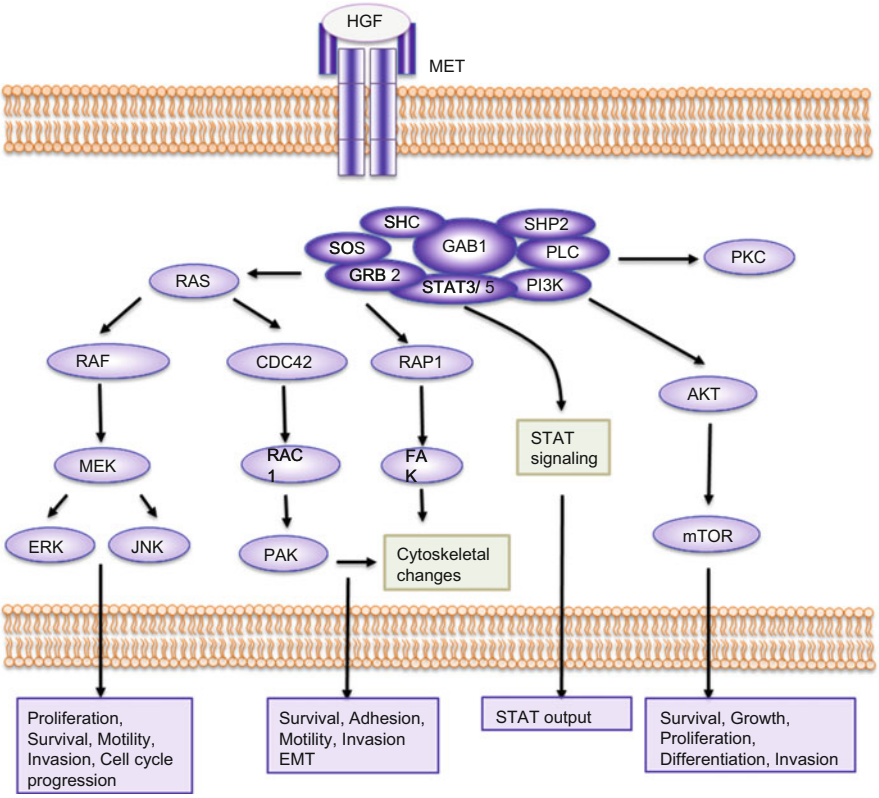

Fig. 1.4 The MET-HGF signaling pathway. Binding of HGF to MET extracellular domain causes receptor dimerization and phosphorylation of many tyrosine residues in the intracellular domain. Phosphorylation at the juxta-membrane domain mediates internalization, at the catalytic domain mediates catalytic activities, and at the cytoplasmic tail facilitates docking of regulatory molecules. Docking of adaptor molecules such as GAB1, GRB2, SHC, and cCBL activates multiple signaling pathways including ERK-MAPK, PI3K, STATs, FAK, and PLCγ

phosphorylated. The phosphorylated STATs dissociate, dimerize, and translocate into the nucleus to control target gene expression. Phosphorylated JAK can also translocate into the nucleus to regulate epigenome structure, such as histone modifications. As expected of many signaling pathways, the JAK/STAT pathway cross talks with both the PI3K/AKT and the RAS/MEK/ERK pathways.

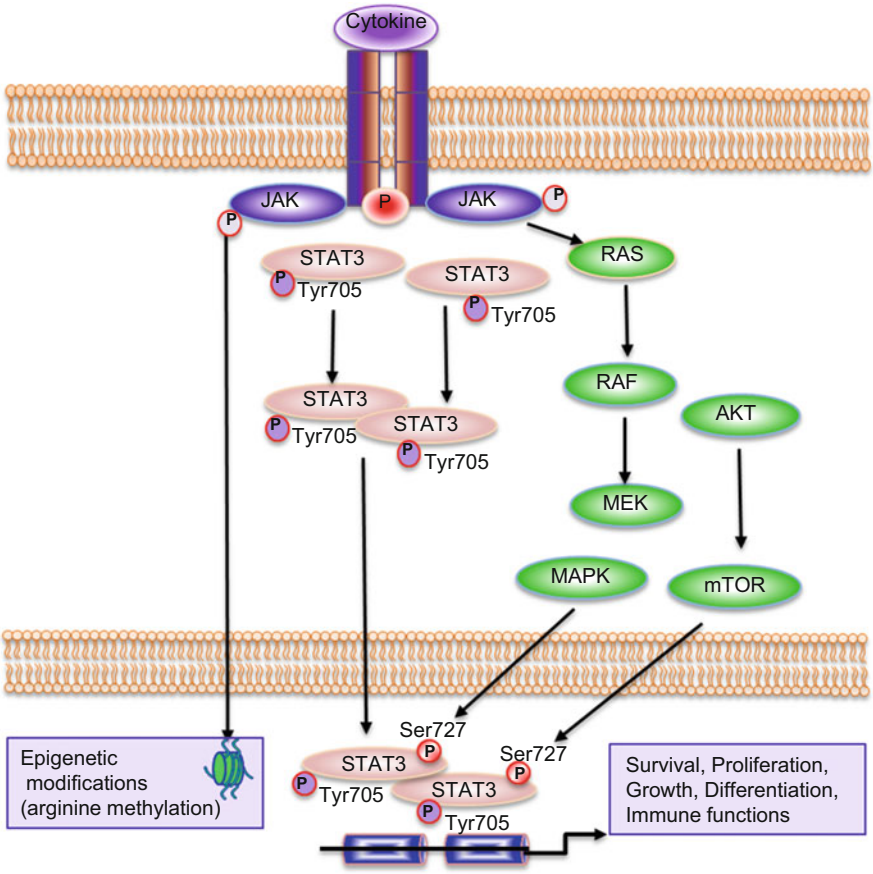

Fig. 1.5 The JAK-STAT signaling pathway. Several receptors, including GPCRs, IL-6Rs, and TLRs, mediate this pathway. Upon ligand binding, the receptors undergo conformational changes to enable binding of adaptor proteins. GPCRs catalyze the conversion of guanine nucleotides to activate JAK2. Cytokine signaling leads to receptor dimerization and the interaction of JAK1 and JAK2 with its gp130 cytoplasmic tails and phosphorylation of STAT3 at tyr705. Phosphorylated STATs dissociate and dimerize via SH2 domains and enter the nucleus to induce gene expression. Serine/threonine kinase signaling can also cause phosphorylation of STAT3 at ser727, which enhances its transcriptional activity

1.9.6 The Transforming Growth Factor-Beta (TGFβ) Signaling Pathway

The TGFβ superfamily signaling pathway controls several cellular processes including cell growth, differentiation, and cytoskeletal rearrangements associated with cell migration, adhesion, and cytokinesis. The TGFβ pathway ligands including nodal, activin, BMP2, BMP4, and BMP7 interact with type I/II receptors leading to oligomerization of serine/threonine receptor kinases and phosphorylation

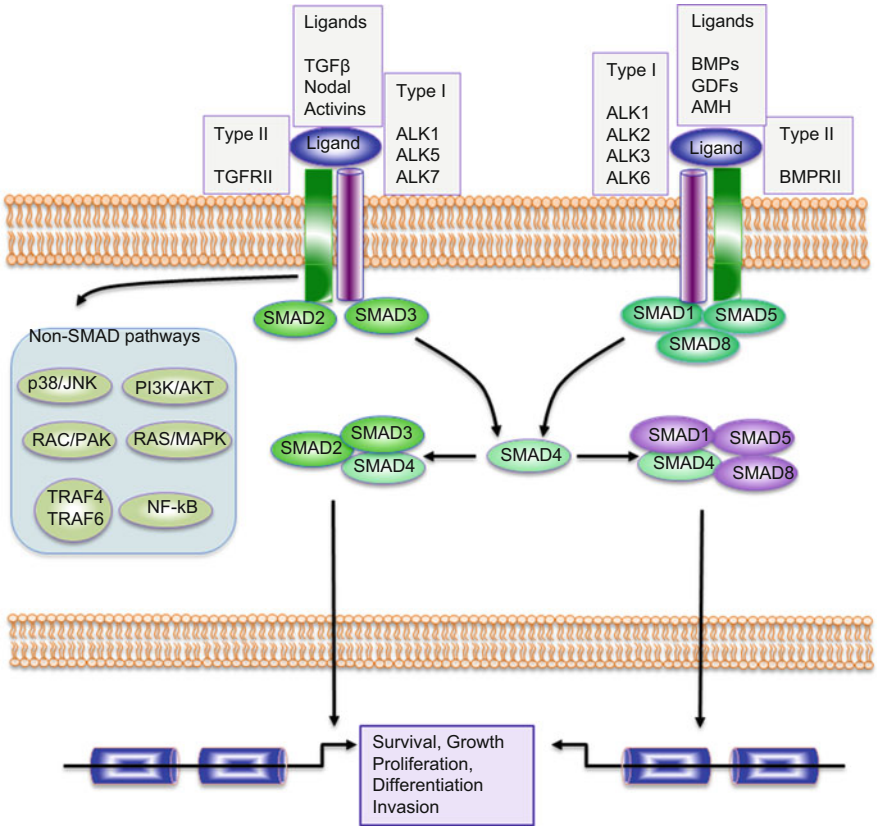

Fig. 1.6 The TGFβ signaling pathway. This pathway is dichotomized downstream of ligand interaction. There are two types of receptors, type I and II. The numerous ligands of this pathway interact differently with the receptors. The interaction of TGFβ, nodal, or activin involves recruitment of ALK4, ALK5, and ALK7 type I receptors that phosphorylate SMAD2 and SMAD3, while interactions by other ligands such as BMPs, growth and differentiation factors (GDFs), and anti-Müllerian hormone (AMH) lead to ALK1, ALK2, ALK3, and ALK6 phosphorylation of SMAD1, SMAD5, and SMAD8. In a non-canonical mode, activated receptor complex can signal via PI3K, TGFβ-associated kinase 1 (TAK1) to p38/JNK, RAS-MAPK, NF-kB, and TNF receptor-associated factor 4 (TRAF4) and 6 (TRAF6)

of cytoplasmic signaling molecules (Fig. 1.6). Phosphorylated SMAD2/3 mediate the TGFβ/activin pathway, while SMAD1/5/8 are signaling mediators of the BMP pathway. However, both groups of SMAD molecules form complexes with the common signal transducer, SMAD4. These complexes then translocate into the nucleus to induce transcription of specific genes. There is also a cross talk between this pathway and the PI3K, ERK, JNK, and p38 MAPK pathways.

1.9.7 The Nuclear Factor-kB (NF-kB) Signaling Pathway

The nuclear transcription factor, NF-kB, controls inflammatory responses, apoptosis, cell cycle progression, and epithelial-to-mesenchymal transition. There are five Rel/NF-kB family members, namely, p65 (RelA), RelB, c-Rel, p50/p105 (NF-kB1), and p52/p100 (NF-kB2). These proteins are characterized by the possession of the Rel homology domain (RHD) located at the N-terminal regions of the proteins. The RHDs are critical to their functions. The five family members exist as inactive homodimer or heterodimer complexes via their RHD to inhibitor proteins, IKBα and IKBβ. The NF-kB-IKB complex cannot translocate into the nucleus and hence is prevented from transcriptional regulation of gene expression. NF-kB is activated in two different ways, the classical or canonical and alternative pathways. The inhibitors of IKB kinases (IKK) mediate NF-kB activation. Inhibitors of IKB kinases have two catalytic subunits, IKKα and IKKβ, and a structural subunit, IKKγ/NEMO. In the classical pathway, IKB is phosphorylated by activated IKKβ. The phosphorylated IKBs are then targeted for ubiquitination and proteasomal degradation. Thus, activated IKK frees NF-kB, which then translocates into the nucleus to activate target genes. The alternative pathway involves activated IKKα-dependent phosphorylation and processing of p100 with nuclear translocation of p52/RelB dimer. Both pathways are illustrated in Fig. 1.7.

1.9.8 Src Signaling in Cancer Invasion

Src, the first discovered oncogene, is overexpressed in many cancers. It belongs to the large conglomerate family of non-receptor tyrosine kinases including Yes, Blk, Yrk, Lyn, Fyn, Fgr, Lck, and Hck. Src is, however, the most extensively studied of this family. Activation of Src orchestrates numerous oncogenic cellular activities including cellular proliferation, survival, and importantly cytoskeletal reorganization and changes in adhesion molecules that lead to EMT, cell motility, migration, invasion, and subsequent metastasis. Src is activated by stimulation of plasma membrane receptors including integrins and RTK. Src and focal adhesion molecule (FAK) localize to the cell-extracellular matrix adhesion sites where they regulate integrin signaling. FAK-integrin interaction leads to autophosphorylation of FAK at tyrosine residue 397. Phosphorylated tyr397 serves as docking sites for SH2 domain-containing molecules including Src. Recruited Src to phosphorylated tyr397 leads to phosphorylation of multiple tyrosine residues on FAK, which in turn creates docking sites for other SH2 domain-containing molecules such as growth factor receptor-bound protein 2 (GRB2). These intracellular activities lead to loss of cadherin-dependent contacts and the assumption of mesenchymal phenotypes with eventual motility and invasion. The mechanism of this phenotypic acquisition is partly due to activation of the small GTPase, Rac, which in turn activates JNK to induce expression of matrix metalloproteinase 2 and 9.

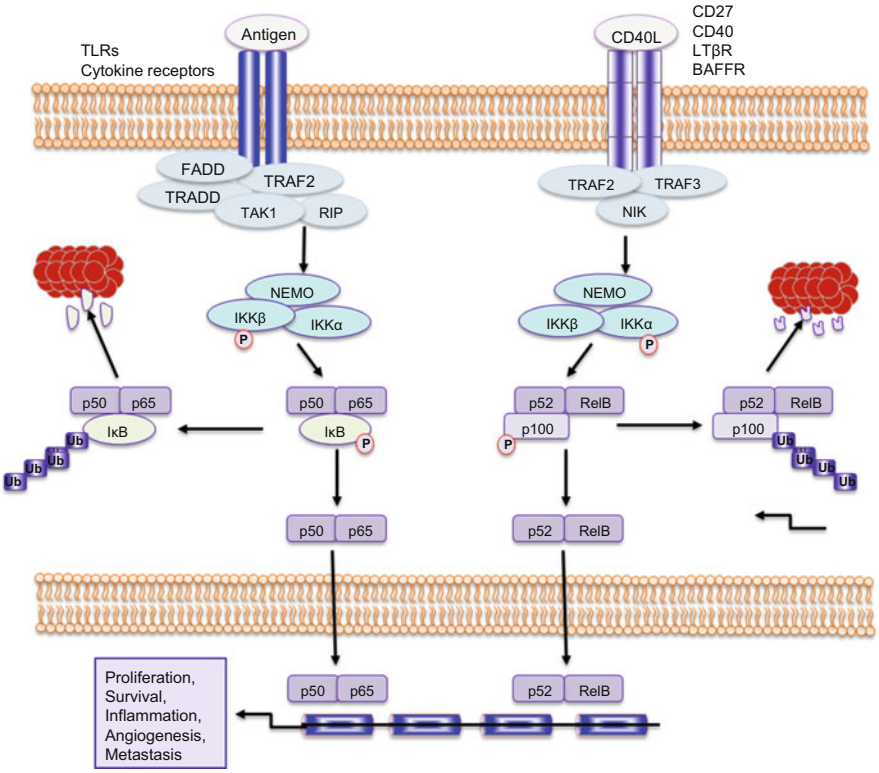

Fig. 1.7 The NF-kB signaling pathway. TLR ligands, cytokines (e.g., TNF), and antigens interact with their receptors to activate canonical NF-kB signaling (*left*). Using a number of signaling adaptors, IKKβ in the IKK complex gets activated, which then phosphorylates Ikβ for ubiquitination and degradation. This releases p50-p65 NF-kB dimers to enter the nucleus to induce gene transcription. The non-canonical pathway, which controls expression of cytokines and genes for lymphoid development, signals via NIK to phosphorylate IKKα. Phosphorylated IKKα then phosphorylates p100, which exists in a complex with RelB, for ubiquitination and proteasomal processing to release p52-RelB dimer that enters the nucleus to mediate transcriptional activity

1.9.9 Control of the Cell Cycle

The cell cycle involves a complex interplay between signaling molecules and pathways that control the induction of genes required for cell growth, DNA synthesis and repair, and the eventual production of two new daughter cells from the parental cell. In tissues needing continuous cellular replacement (e.g., hematologic system and mucosal linings), the cell cycle is in constant progression in progenitor cells. Otherwise, postmitotic cells exit the cycle into a state of quiescence (G0), but some can be induced to reenter the cell cycle. Abnormal stimulation and hence reentry into the cell cycle can lead to unregulated growth and possible disease evolution.

Mitogens and growth factors drive cell cycle progression. Their levels and effects are regulated to ensure a cell divides only when physiologically required. However, deregulated expression and hence excessive growth factor stimulation can lead to several pathologic conditions including cancer.

Various molecules, primarily cyclins and their counterparts and cyclin-dependent kinases (CDKs), control the cell cycle at specific checkpoints. These molecules (cyclins and CDKs) form complexes to drive cell cycle progression. Consistent with every regulatory system, there are several negative regulators of these complexes. The retinoblastoma (RB) gene products (p107/RBL1 and P130/RBL2) tightly regulate G1 checkpoint, which controls transition of the cycle into DNA synthesis or S-phase. In nondividing cells, RB sequesters E2F family of transcription factors involved in the induction of genes required for S-phase transition, thereby halting cell cycle progression. The actions of mitogens and growth factors overcome this inhibitory effect of RB. Mitogen signaling activates cyclin D1-CDK4 and cyclin D1-CDK6 complexes with the phosphorylation of RB leading to the release and activation of E2Fs. Freed E2Fs then enter the nucleus to induce genes with TTTCCCGC sequence motifs in their promoters, including cyclin E, which is needed for S-phase transition. The cyclin E-CDK2 complex propels S-phase activity and hence DNA synthesis. The cyclin D1-CDK4 and cyclin D1-CDK6 complexes are under negative control by p16INK4A, which is usually expressed to cause senescence and usher in cellular differentiation.

Even after DNA synthesis, the cell is not necessarily destined to divide (except for those with deranged regulatory mechanisms such as the cancer cell). The "guardian of the genome," p53, halts the cell cycle at the G2 checkpoint to ensure any DNA copying mistakes are efficiently repaired. P53 levels are usually low in the nondividing cell, through the E3 ubiquitin protein ligase or mouse double minute 2 (MDM2) homolog-mediated degradation. During cell division, however, p53 levels and activity are elevated through the actions of DNA damage sensors, including ATR and ATM. These proteins phosphorylate CHK1 and CHDK2, which in turn phosphorylate and increase p53 activity. Phosphorylated p53 forms tetramers, which act as transcription factors to induce expression of CDKN1A/p21CIP. The expressed protein, p21CIP, inhibits the cyclin-CDK complexes to temporarily stop the cell cycle, in order to facilitate the repair of any possible DNA damage. If the DNA damage cannot be repaired, p53 rather induces expression of genes involved in apoptosis. Figure 1.8 illustrates the basic control mechanisms of the cell cycle.

1.9.10 Hypoxia-Inducible Factor Signaling

The growth of cancer cells is associated with different levels of oxygen deficiencies. Because solid tumors grow rapidly, cells at the periphery will have access to surrounding vasculature and therefore are more likely to be adequately oxygenated (>5 %). However, progressively deep into the tumor will be cells exposed to

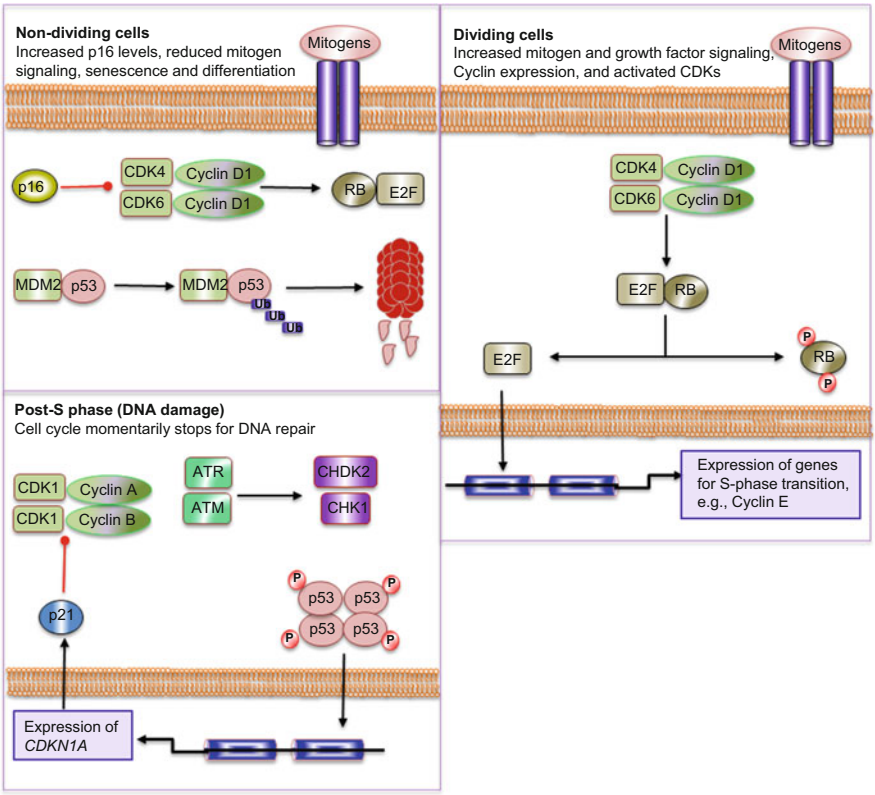

Fig. 1.8 Cell-cycle regulation. The cell cycle is tightly controlled to prevent unwanted cell proliferation. In nondividing cells, negative regulators enable the RB pocket proteins to sequester E2F transcription factors and hence reduce the expression of genes required for S-phase transition. Cell division, induced by mitogen signaling, overcomes the negative controls, partly through cyclin D1 expression. Cyclin D1-CDK4 and cyclin D1-CDK6 complexes phosphorylate RB leading to the release of E2F and hence propulsion of the cell cycle through S-phase. Any possible DNA miscopying that might have occurred at S-phase is repaired prior to cell division. This is achieved by p53 activity, which in nondividing cells is kept low by its sequestration by MDM2 for degradation

decreasing levels of oxygen that could range from hypoxia (<5 %) to even anoxia. The reason for the reduced oxygen levels is primarily due to poor blood supply from rapid growth that outstrips vasculogenesis. Indeed, as little as 300 cancer cells can create hypoxic conditions as a consequence of inadequate vascularization. One mechanism to survive these adverse conditions is the induction and stabilization of hypoxia-inducible factors (HIFs), which partly restore new blood vessels to the tumor.

The HIFs are transcription factors with basic helix-loop-helix and Per/ARNT/Sim (PAS) domains. They were first identified as factors that regulate increased expression of erythropoietin in response to hypoxia and hence so named [7]. The

two family members, HIFα and HIFβ, become active upon formation of heterodimers. Hypoxia-inducible factor-α has three subunits, namely, HIF1α, HIF2α, and HIF3α. Hypoxia-inducible factor 2α is ~48 % homologous to HIF1α and is expressed and stabilized under hypoxic conditions as well. On the contrary, HIF3α is a dominant negative regulator of HIF because it dimerizes with HIF1β to form a transcriptionally inactive heterodimer, thus reducing the activity of HIFs. Hypoxia-inducible factor-β is an aryl hydrocarbon receptor nuclear translocator (HIF1β/ARNT) with two homologs (ARNT2 and ARNT3). All three HIF1β homologs are constitutively expressed and can heterodimerize with HIFα subunits. Whereas both HIF1α and HIF2α are functionally active and can interact with hypoxia-responsive elements (HREs: canonical CCATG sequences) in gene promoters, it appears HIF1α preferentially induces HREs in glycolytic gene promoters and hence plays important roles in cancer cell metabolic alterations.

Hypoxia-inducible factor 1α is stabilized under hypoxic conditions (Fig. 1.9). In normoxia, the levels of HIF1α are low in cells because of its ubiquitin-mediated proteasome degradation. In this process, the protein is first hydroxylated on proline 402 and 564 in the oxygen-dependent degradation (ODD) domains by HIF1α prolyl hydroxylases. The von Hippel-Lindau (VHL) protein products, in a complex with Cul-2, elongin B, and elongin C, recognize the hydroxylated proteins. The VHL protein is an E3 ubiquitin ligase (together with NEDD8) that mediates the ubiquitination and proteasomal degradation of hydroxylated HIF1α. Under hypoxic conditions, HIF1α cannot effectively be hydroxylated by prolyl hydroxylases since their Michaelis constant (Km) for oxygen is very high. Therefore, under such circumstances, HIF1α is stabilized and enters the nucleus to form an active transcription factor by heterodimerizing with HIF1β subunits. The heterodimeric complex then binds to hypoxia-responsive elements and induces expression of several genes including those involved with glucose metabolism, angiogenesis, tumor invasion, and survival. In normal physiology, hydroxylation is the primary mode of HIF1 regulation. However, hypoxia and pseudohypoxia are hallmarks of fast-growing solid tumors. Cancer cell microenvironment can stabilize HIF1α via hydroxylation. It has recently become evident that apart from hypoxia, the levels and activity of HIF1α can be increased under normoxic conditions by cancer cells, a condition referred to as pseudohypoxia.

1.9.11 The WNT/β-Catenin Signaling Pathway

The WNT signaling pathway is very complex and diversified. In addition to the WNT/β-catenin pathway, other pathways such as the planar cell polarity pathway control cell polarity and migration, while the WNT/calcium-NFAT pathway is involved in cell-fate determination, as well as cellular migration. The canonical WNT/β-catenin signaling pathway regulates normal embryogenesis and is deregulated in a number of cancers. It is also involved in stem cell maintenance

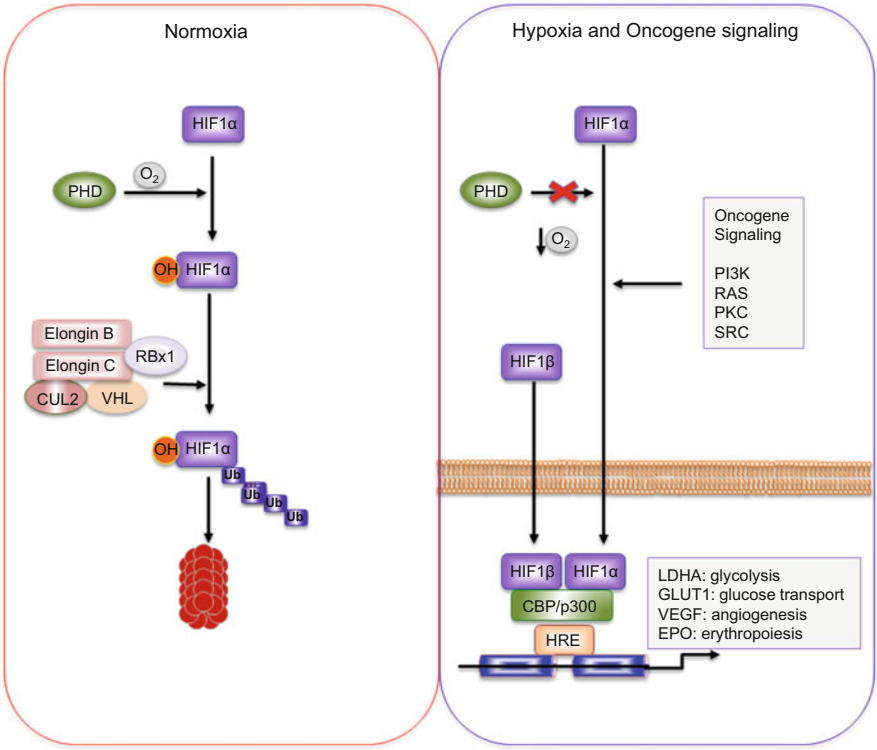

Fig. 1.9 The HIF1 signaling pathway. In normoxia, HIF1α is hydroxylated, ubiquitinated, and degraded by the proteasome. However, in hypoxic conditions, PHD activity is low; hence, VHL is ubiquitinated and degraded, leading to the stabilization of HIF1α. Stabilized HIF1α enters the nucleus and forms a complex with HIF1β and CBP-p300 to induce gene expression

and cell-fate determination, both of which are important in developmental and cancer biology.

In the absence of WNT ligand, β-catenin, the pathway transcriptional co-regulator, and E-cadherin cell-cell adhesion adaptor are sequestered in a multimeric protein complex composed of AXIN, APC, and GSK3β (Fig. 1.10). Beta-catenin is then phosphorylated by CK1 and GSK3β leading to β-TrCP-mediated ubiquitination and proteasomal degradation. This abrogates pathway activity.

When WNT ligands bind to their receptors (FRIZZLED), a complex is formed between the FRIZZLED receptor and its co-receptor, LRP5/6. This complex activates DISHEVELLED, which causes release of GSK3β from the complex, as well as destabilization of the AXIN/APC/GSK3β complex, and hence the release and stabilization of β-catenin. The freed β-catenin is transported into the nucleus, with the help of RAC1. In the nucleus, β-catenin interacts with LEF/TCF transcription factors to induce WNT/β-catenin target genes, including *CCND1*, *MYC*, *PPARγ*, *TCF1*, *AXIN2*, *CD44*, and *MMP7*. The WNT/β-catenin pathway interacts with the TGFβ, BMP, FGF, and other signaling pathways.

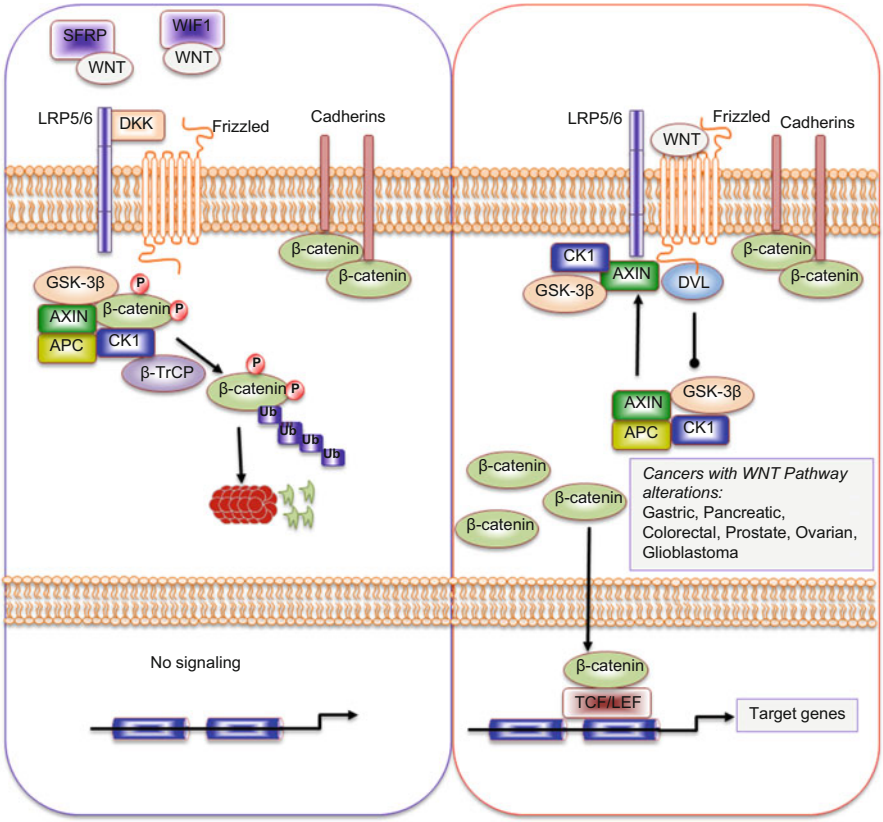

Fig. 1.10 The WNT/β-catenin signaling pathway. In the absence of WNT interactions with its FRIZZLED receptor, β-catenin is sequestered in a complex and phosphorylated by CK1 for ubiquitination and proteasomal degradation. Cytoplasmic β-catenin levels are low (found mainly in association with cadherin). WNT interaction with FRZ causes destabilization of the complex with the release of β-catenin, which then accumulates in the cytoplasm, and hence enters the nucleus to form complexes with TCF/LEF and CBP to induce target gene expression

Noncanonical β-catenin-independent WNT signaling occurs, and these include WNT/calcium and WNT/JNK pathways. The WNT signaling pathway has a number of negative regulators including RNF43 and ZNRF3 that can ubiquitinate FRIZZLED. Also, DKK and SFRPs can inhibit LRP and FRIZZLED, respectively.

1.9.12 The Hedgehog Signaling Pathway

The evolutionarily conserved hedgehog (HH) pathway consists of three ligands, sonic hedgehog (SHH), Indian hedgehog (IHH), and desert hedgehog (DHH), with distinct and overlapping functions. This signaling pathway plays important roles in

embryogenesis, stem cell and tissue maintenance, tissue regeneration, and cancer. Thus, the signals from this pathway are involved in cellular proliferation, differentiation, survival, and cell-fate determination among several others.

The HH signaling pathway is unique in a number of ways. For instance, the synthesized ligands undergo autocatalytic cleavage in association with cholesterol modification of the C-terminus and palmitoylation at the N-terminus in order for these molecules to be functional. Additionally, the secretion of these ligands requires Dispatched and Scube 2, while their transport in the extracellular milieu is facilitated by cell surface glypican family of heparin sulfate proteoglycans (GPC1-6) and LRP2.

In the absence of HH proteins, the cognate receptor, PATCHED (PTCH), represses SMOOTHENED (SMO), leading to the sequestration of GLIs, SUFU, and Kif7 (Fig. 1.11). The GLIs are then phosphorylated and processed into

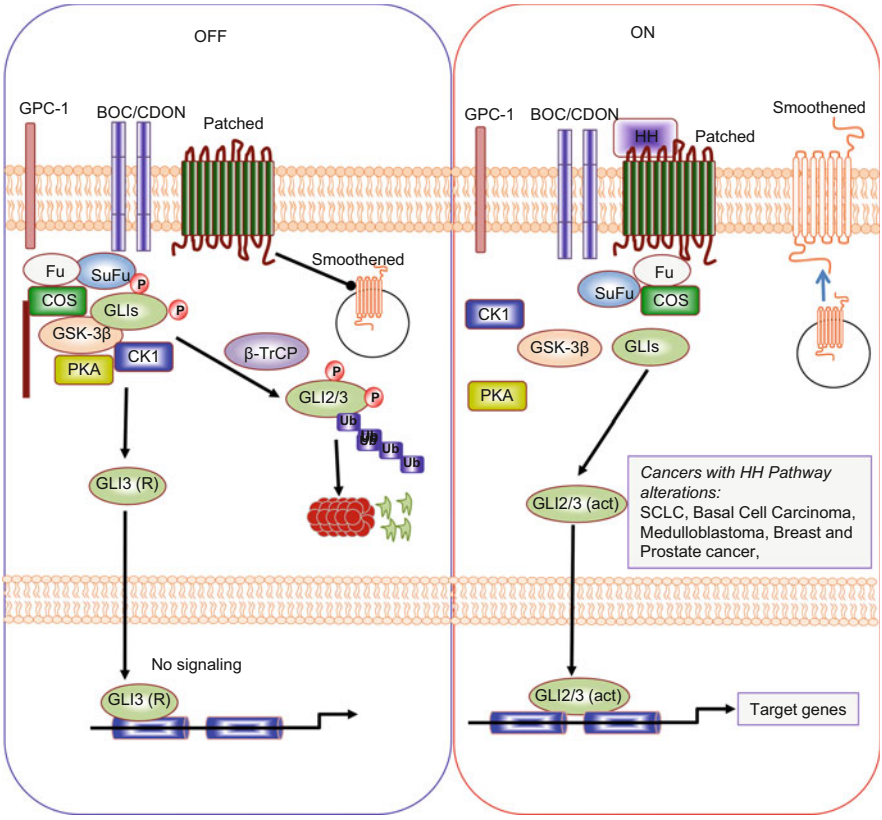

Fig. 1.11 The HH signaling pathway. In the absence of HH ligand, PATCHED represses SMO preventing its interaction with SUFU/FU complex. In this state, GLI proteins are phosphorylated by PKA to form repressors that enter the nucleus to halt target gene expression. When HH binds to PTCH, SMO is released to mediate the formation and release of GLI activators, which enter the nucleus to induce expression of target genes

transcriptional repressors (GLI3) by GSK3β, PKA, and CK1. GLI1 and GLI2 are in this state ubiquitinated for proteasomal degradation through the E3 ubiquitin ligase, β-TrCP.

When HH proteins bind to PTCH, and co-receptors BOC, CDO(N), and GAS1, this causes derepression of SMO, allowing dissociation of sequestered GLI proteins from Kif7 and SUFU. The GLIs then enter the nucleus to mediate transcription of HH target genes, which include *GLI1*, *PTCH*, *CYCLIN D/E*, *HIP*, *WNT*, *MYC*, and *BMP*.

1.9.13 The Notch Signaling Pathway

The evolutionarily conserved Notch signaling pathway is an important regulator of normal embryogenesis by exerting its effects on cell-fate decision-making. Additionally, this pathway is involved in the maintenance of adult tissues. Notch signaling is deregulated in a variety of malignancies including adult T-cell acute lymphoblastic leukemia and lymphoma.

The Notch pathway ligands, Jagged (JAG1, and JAG2) and Delta-like (DLL1, DLL3, and DLL4), and the notch receptors are all transmembrane proteins (Fig. 1.12). Therefore, the cell sending the signal and the one receiving it must be juxtaposed or in close proximity for pathway activation. The notch receptor is composed of three main domains, intracellular (NICD), transmembrane (TM), and extracellular (NECD) domains. These molecules are synthesized and processed by the signal-receiving cell via the "synthesis for export" pathway of protein synthesis. Thus, they are synthesized in the rough endoplasmic reticulum; posttranslationally modified, sorted, and packaged in the Golgi apparatus; and pinched off in an endosomal vesicle that delivers the receptor to the cell membrane via membrane fusion and integration. The posttranslational processing involves glycosylation and cleavage (S1 cleavage) to generate a calcium-stabilized heterodimer of NECD, noncovalently bound to the TM-NICD.

Ligand-receptor interaction leads to NECD cleavage (S2 cleavage) by TNFα ADAM metalloprotease-converting enzyme (TACE). The NECD together with the ligand is endocytosed and recycled by the signal-sending cell through Mib-mediated ubiquitination. Finally, NICD is released from TM by γ-secretase, enabling its nuclear translocation and interaction with CBF1/Su(H)/Lag1 (CSL) transcription factor complex to induce expression of notch target genes including *HES* family, *CDKN1A*, and *MYC*.

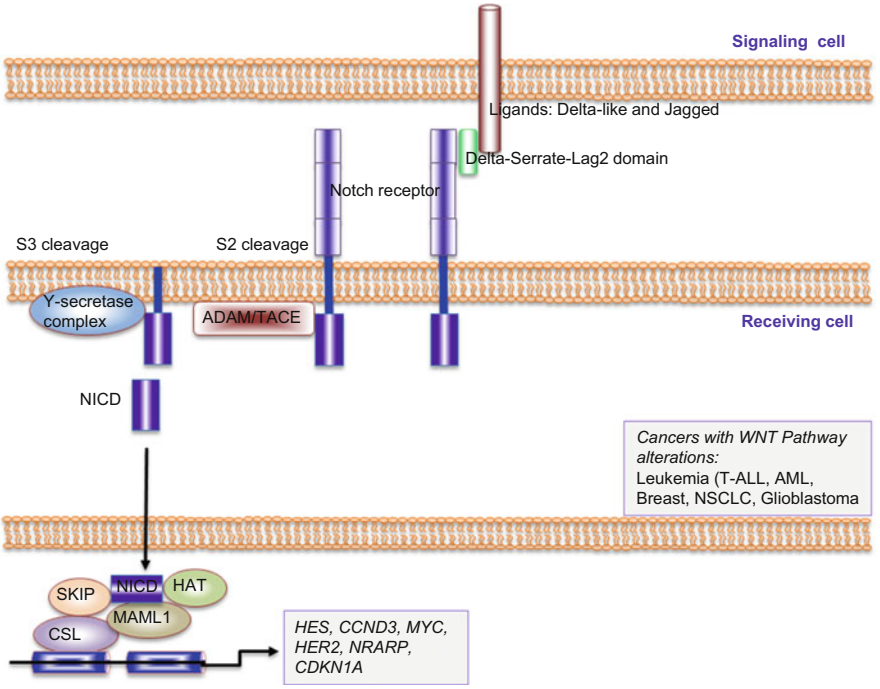

Fig. 1.12 The Notch signaling pathway. Notch signaling requires the juxtaposition of the signal-sending cell (with membrane-bound JAG and DLL ligands) and signal receiving (with membrane-bound NOTCH receptor). Ligand-receptor interaction leads to cleavage and proteolytic processing of NOTCH receptor by ADAM/TACE and γ-secretase to release the soluble NICD into the cytoplasm. NICD translocates to the nucleus to form complex with MAML1, CSL, HAT, and SKIP to induce target gene expression

1.10 Cell Death Processes and Cancer

Cell death processes are morphologically and biochemically characterized mechanisms by which various cells are killed. Apoptosis, autophagy, and necrosis have distinct structural features that define them. However, biochemically, there could be cross talk between these cell death mechanisms. Irrespective of how it is executed, cell death is a normal physiologic as well as a pathologic feature of living organisms. Apoptosis, for instance, is a normal physiologic cell death process that operates during the development of an organism to facilitate normal tissue patterning and in adult tissues to maintain normal tissue homeostasis. In pathologic conditions, apoptosis serves an important function of removing severely damaged cells safely without causing adverse effects to neighboring cells. In order for the cancer cell to survive in toxic environmental niches, they modulate apoptotic pathways using mechanisms including adaptive oncogenic signaling pathways. De novo inhibition of apoptosis can initiate neoplastic transformation, and established tumors have altered molecular signatures that enable evasion of

apoptosis, either spontaneously or when challenged with chemotherapy or radiotherapy.

1.10.1 Nomenclature for Cell Death Processes

Several cell death processes have been identified and described. Traditionally, a point of no return was used as the sine qua non of cell death. This pivotal point of cell demise was defined as caspase activation, dissipation of mitochondrial electrochemical gradient, mitochondrial outer membrane permeabilization, and exposure of phosphatidylserine moieties by the dying cells that enable neighboring cells to destroy them. Whereas these processes can lead to cell death, they are not necessarily sufficient to kill a cell. In view of this obvious difficulty, the Nomenclature Committee on Cell Death (NCCD) recommends that a cell should only be considered dead when at least one of the following biochemical or morphologic features is met: (1) the cell loses its membrane integrity, (2) the cell is completely fragmented into apoptotic bodies, and/or (3) engulfment of the remains of the cell by nearby cells has taken place [8].

Apoptosis, necrosis, autophagy, and cornification are distinct modes of cell death. Specific morphologic features are used to delineate these cell death pathways. For example, the term "apoptosis" was coined by Kerr et al. to describe a form of morphologic features associated with cell death [9]. Apoptotic cell death involves withdrawal of pseudopods leading to rounding up of cells, plasma membrane blebbing, cellular pyknosis, nuclear fragmentation (karyorrhexis), and phagocytic removal of the cell. In necrosis, cytoplasm and its organelles swell up leading to plasma membrane disintegration. This process could be associated with some nuclear condensation. The hallmark of autophagic cell death is massive cytoplasmic vacuolation with accumulation of autophagic bodies without nuclear condensation and phagocytosis [8]. The best-studied and characterized cell death process is apoptosis, and the importance of defective apoptosis in cancer is well established.

1.10.2 Apoptosis

Apoptosis or programmed cell death pathway is dichotomized into distinct but interacting pathways, namely, intrinsic and extrinsic (Fig. 1.13). This distinction is partly based on the inciting signals. Ligands that interact with FAS, TRAILR, IL-1R, or TNFR1 tend to trigger the extrinsic pathway, while intracellular stressors such as excessive genomic damage, growth factor withdrawal, radiotherapy, or chemotherapeutic agents signal mainly via the intrinsic pathway. These pathways induce target and activate different caspases; however, they eventually converge on the induction of the terminal effector caspases (3, 6, and 7) to execute cell death. Activated caspase 8 links the two pathways together. Apart from the noted inducers

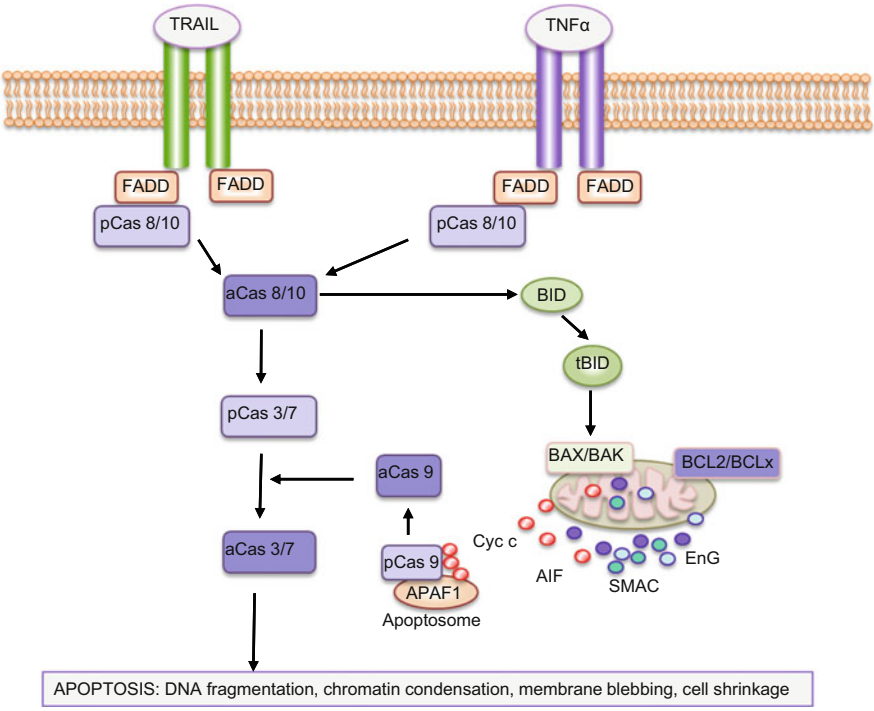

Fig. 1.13 The apoptotic pathway. Death receptors including CD95 (FAS), DR3, DR4, DR5, and TNFR interact with their respective ligands to trigger the extrinsic apoptotic pathway. TNF-related apoptosis-inducing ligand (TRAIL) or CD95L interacts with CD95, DR3, DR4, or DR5, while TNFα with TNFR. This interaction leads to recruitment of FAS-associated protein with death domain (FADD) to form the death-inducing signaling complex (DISC), which then recruits caspase 8 (and10). Homodimerization and cleavage activates caspase 8, and together with activated caspase 9 activate the executioner caspases 3 or 7. Activated caspase 3/7 mediate multiple protein cleavages resulting in the morphological features of apoptosis (this pathway occurs in type I cells). In type II cells, caspase 8 cleaves and activates BID to truncated BID (tBID), which induces mitochondrial release of proapoptotic cytochrome c and second mitochondrial-derived activator of caspases (SMAC). The BCL-2 proapoptotic (BAX, and BAK) and anti-apoptotic (BCL-2, and BCL-X_L) proteins control the mitochondrial (intrinsic) apoptotic pathway. BAX and BAK can oligodimerize to form pores in the outer mitochondrial membrane leading to the release of cytochrome c and SMAC. In the absence of an apoptotic signal, BCL-2/BCL-X_L prevents BAX/BAK oligomerization. The released cytochrome c forms a complex with apoptosis protease-activating factor 1 (APAF1) and procaspase 9 to form the apoptosome. The procaspase 9 dimerizes, cleaves, and gets activated within the complex. IAP (XIAP) can inhibit caspases 3, 7, and 9, while SMAC inhibits IAP to offset this inhibition. Apoptosis-inducing factor (AIF) and endonuclease G (EnG) cause DNA fragmentation. Oncogenic signaling pathways including the PI3K, PKC, ERK-MAPK, and PKA can modulate the apoptotic pathway

of apoptosis, oncogenic signals modulate these pathways to promote cancer progression.

1.10.3 Cancer Initiation as Defects in Apoptosis

Pioneering works led by Korsmeyer and colleagues led to the realization in the 1980s that apart from oncogenic stimulation and inactivation of tumor suppressor genes, primary apoptotic suppression can lead to malignant transformation [10–13]. BCL-2 was identified through studies of chromosomal translocation in follicular lymphomas. Molecular studies revealed the presence of t(14;18)(q32;q21) translocation in ~85 % of lymphomas. Subsequent cloning of the chromosome 14 breakpoint revealed a transcription factor named then as BCL-2 [14–16]. Transgenic technology enabled the study of BCL-2 lymphomas in mice. Overexpression of a BCL-2 transgenic construct in mice caused an increase in the number of B lymphocytes in splenic white pulps [11]. These hyperplastic splenic white pulps of BCL-2 overexpressing transgenic mice progressed to monoclonal B-cell lymphomas in ~15 months period [17]. Surprisingly, this finding was not due to differential proliferation rates between transgenic mice and wild-type controls. Instead, it was noted that the BCL-2 overexpressing cells were arrested at G_0/G_1 of the cell cycle [11]. An explanation for the survival advantage conferred by BCL-2 overexpression was later provided when cytokine-dependent cell lines were transfected with BCL-2 overexpressing vectors and observed to survive at G_0 without the needed cytokine, IL-3 [18]. These findings clearly indicate that the ability to avoid apoptosis, and survive without the needed external growth factors, could indeed lead to malignant transformation.

1.11 Multistep Carcinogenesis and Early Detection Cancer Biomarkers

For an overt cancer to be recognized, it must have been through several years of evolutionary progression to overcome all inhibitory mechanisms posed by the metazoan cellular community. This is because the metazoan cellular homeostatic mechanisms are not geared toward “selfish behaviors” of its members. Thus, the cancer cell requires time to accumulate the necessary enabling mutations needed for its independent existence. To achieve this, the cancer cell must undergo multiple steps of genetic or genome-wide alterations, and this is the postulated thesis of the multistep phenomenon of carcinogenesis. While applicable to all cancers, the multistep cancer pathway is well documented for cancers such as those of the colon and rectum, breast, prostate, head and neck, lung, skin, and uterine cervix. In a histopathological context, these stepwise cancer progressive pathways can either follow the hyperplasia-dysplasia (polyps and adenomas)-carcinoma sequence, which is well illustrated by colorectal cancer, or the metaplasia-dysplasia-carcinoma sequence (without evidence of adenomas) as demonstrable in gallbladder and endometrial cancer progression.

Underlying these histopathologic changes are molecular genetic alterations that drive the "proband" cell. These genetic changes that occur with disease progression serve to drive the cell toward overt malignant phenotype. For example, the multistep carcinogenesis of colorectal cancer involves in its simplistic mechanism, initial changes including *APC* mutations, followed progressively by *KRAS*, *DCC*, *SMAD4*, and *TP53* mutations to fully transform the colonocyte into an invasive cancer.

The multistep carcinogenesis model of cancer development offers tremendous molecular explanation for Slaughter's concept of "field cancerization" [19]. The observations of these early scientists were that cancer did not develop in a focal fashion amenable to simple complete surgical resection. Slaughter and his contemporaries were well convinced by their initial observations of oral mucosal lesions that a large expanse of tissue was affected and hence at risk of giving rise not only to one but multiple cancers. The description by D. P. Slaughter and his colleagues indicates global damage to oral epithelium and contiguous upper aerodigestive structures by carcinogens (tobacco and alcohol), leading to the evolution of multiple primary tumors. In the case of colorectal cancer, for example, even if a single initial cell acquired *APC* mutations, it could eventually undergo clonal expansion and spread to occupy vast expanse of the colorectal mucosa. Some cells will acquire other molecular alterations to become cancerous and thus be clinically diagnosed. But the non-overt clones, which are non-histopathologically recognizable, will still be carrying the *APC*, and possibly other genetic alterations, and hence are poised to become malignant sometime in the future. These early molecular and genetic changes provide early detection biomarkers that can be used for risk stratification and early primary prevention such as the application of chemoprevention measures.

1.12 Personalized Cancer Care

In the field of pharmacology, the era of "one agent fits all" is quickly disappearing with the ushering inn of personalized medicine and oncology. With current high-resolution sequencing and other "omics" technologies, individual cancers can be genotyped, such that therapy and management strategies are stratified to patient populations based on the genetic landscape of their tumors. This approach is being made possible with the development of targeted biotherapeutic agents. Also of importance to personalized oncology is the development of biomarkers that inform targeted molecular therapy, often referred to as companion diagnostics.

Brian J. Druker pioneered the concept of targeted therapy in 1996 [20]. He developed the first small-molecule inhibitor, STI571, which is now known as imatinib (Gleevec) for the treatment of CML. The Philadelphia chromosomal translocation (Chr 9 and 22) generates the Bcr-Abl fusion molecule with kinase activity conferred by Abl. Imatinib binds and inhibits this kinase activity such that $>90\,\%$ of patients treated achieve remission.

Targeted molecular therapy also evolved with the discovery of hormone receptor status (ER, PgR), followed by the amplification status of human epidermal growth factor (HER2) in association with its overexpression in some tumors. For example, ER+ tumors are targeted with tamoxifen and the ER antagonist, fulvestrant. HER2, which is amplified in ~20 % of breast cancers, is targeted with the humanized monoclonal antibody, trastuzumab, which binds the extracellular domain of HER tyrosine kinase transmembrane receptor. Another important targeted agent for HER2+ tumors is the humanized monoclonal antibody, pertuzumab, that binds HER2 dimerization domain and thus prevents receptor dimerization, which is not blocked by trastuzumab. The small-molecule tyrosine kinase inhibitor, lapatinib, targets both HER2 and HER1. The synergistic effects of trastuzumab and pertuzumab have been established. Other important targeted therapies are being developed for several cancers including melanoma, NSCLC, metastatic CRC, and CML (Table 1.3). There is a long pipeline of tyrosine kinase inhibitors (TKIs) (~110 targets) under various developmental phases. Current important kinase targets under clinical development include PI3K/mTOR, EGFR, VEGF/VEGFR, HER2, MET, KIT, and PDGF/PDGFR pathways. Efforts by industry are focused on targeting just these few pathways; however, this list is poised to grow with advancing knowledge on cancer molecular networks.

Targeted therapies have proven successful in tumors that harbor unique genetic abnormalities such as the Bcr-Abl fusion in CML and the successful response to imatinib. Another successful targeted therapy, similar to the Bcr-Abl fusion, is the echinoderm microtubule-associated protein-like (EML4) anaplastic lymphoma

Table 1.3 Examples of targeted therapies in oncology. The FDA has approved some agents for the treatment of various cancers

Molecular target	Agents
RAS	Lonafarnib, tipifarnib
RAF	Sorafenib, vemurafenib, dabrafenib
MEK	Sorafenib, selumetinib, crizotinib, cobimetinib, trametinib
mTOR	Sirolimus, everolimus, temsirolimus
AKT	Perifosine, triciribine
VEGF	Axitinib, cabozantinib, lenvatinib, panobinostat, ponatinib, ramucirumab, regorafenib, sorafenib, vandetanib
EGFR (HER1/ERBB1)	Afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, vandetanib, panitumumab
HER2/ERBB2	Trastuzumab, pertuzumab, lapatinib, afatinib
CDK4/6	Palbociclib
cKIT	Sunitinib, imatinib, nilotinib, dasatinib
PKC	Enzastaurin, tamoxifen
Src	Dasatinib
ALK	Alectinib, ceritinib, crizotinib
ROS1	Crizotinib
ABL	Bosutinib, dasatinib, nilotinib

kinase (ALK) fusion found in ~4 % of NSCLC patients. Treatment with crizotinib that targets ALK activity is also associated with remission in ~90 % of patients.

An issue with targeted therapy is the fact that many cancers harbor several mutations (4–10 mutations). Therefore, finding and targeting the most relevant "driver" mutation can be a daunting task. Also, in spite of the numerous efforts being made, resistance emerges following treatment with some targeted therapies, especially when single agents are used. An illustrated example is resistance to TKI treatment of NSCLC. EGFR TKI can effectively treat up to 30 % of NSCLC patients with EGFR mutations, but resistance is a nuance weeks to several months after initiating therapy. Current efforts have revealed several possible mechanisms used by the cancer cell to achieve such resistance, and these include:

- Deployment of downstream signaling pathway members independent of EGFR.
- Activating *PIK3CA* mutations that can sustain tumor growth and survival through PI3K pathway activation.
- EGFR amplification to overcome inhibition.
- EGFR exon 20 p.T790M mutation, which reduces drug-receptor interaction in about half of all NSCLC patients.
- MET amplification leading to TK switching or receptor dimerization.
- Low expression of the inhibitor, IkB, leading to increased NF-kB signaling to obviate TKI-mediated apoptosis.
- The presence of cancer stem cells, which may be resistant to TKIs such that following treatment they emerge as new cancers.
- Tumor heterogeneity is another possibility. It is likely NSCLC may contain unrecognizable SCLC cells. Thus, it is known that SCLC emerges in some NSCLC patients on TKIs. SCLC cells that are not targeted by TKIs may be selected for by this therapy.

Detailed tumor profiling and the use of multiple or combinatorial targets may help overcome the resistance to targeted therapies.

Current personalized oncology efforts target "driver mutations" of key genes in primary tumors as predictive biomarkers of therapy response that inform patient selection for treatment. Not only is this approach scientifically sound, but also the future should embrace the complete adoption of personalized biotherapy instead of nonspecific chemotherapies. In so doing, tumor heterogeneity and evolution should be considered. Hence, "liquid biopsy" offers an attractive approach to patient management. Not only does this overcome tumor heterogeneity, but it also offers noninvasive longitudinal sampling to monitor genetic evolution with disease progression and hence tailor treatment appropriately.

1.12.1 Nomenclature of Targeted Therapies

Targeted therapeutic agents are mostly small molecules or monoclonal antibodies. Their names depend on the state of development. Agents in preclinical trials are designated with one or two proprietary names (e.g., STI-571). Once successful, the agent then receives a generic name with stems and sub-stems that have specific meanings. Finally, the agent is given a brand name used for marketing purposes. The parts of the generic name may describe the nature, source, and targets of the agents. A prefix at the beginning of a name is unique to each agent. An end designation with a "-mab" refers to monoclonals, while small-molecule inhibitors end with "-ib" (for inhibitory). Various technologies are used in generating monoclonal antibodies, and these methods reflect as sub-stems in the name of the agent. As examples, "-mumab" is a sub-stem that indicates a fully humanized monoclonal antibody, "xumab" is a humanized mouse monoclonal antibody, and "-ximab" is a chimeric human-mouse monoclonal antibody. The target of the agent may also be indicated with a stem in the middle of the name. For monoclonals, "-tu" stands for a tumor target, and "-ci" stands for a circulatory system target. The middle stem, "cicl," in small-molecule inhibitors stands for cyclin-dependent kinase inhibitor, and "-tin" indicates a tyrosine kinase inhibitor.

1.12.2 Tests of Interest in Personalized Oncology

Molecular genetic analyses are conducted on patient samples prior to treatment planning. This may involve the use of fresh or fixed tissue, exfoliated cells, or body fluid samples for testing. Here are some biomarker assays of interests in personalized oncology:

- *Pharmacogenetic biomarkers*: These are validated assays that identify or stratify individual's genetic variation that usually control the pharmacokinetics and pharmacodynamics of drugs. These assays often target gene variations such as polymorphisms in drug uptake and metabolizing pathways including enzymes, receptors, and transport proteins. Specifically, these proteins modulate how drugs get absorbed, distributed, and function in the body.
- *Pharmacogenomic biomarkers*: These are equally validated assays with similar intent as pharmacogenetic biomarkers but rely on whole genome variations including haplotype markers, SNP maps, and global gene expressions that identify an individuals' response to a therapeutic agent, both efficacy and toxicity.
- *Companion diagnostic biomarkers*: These assays rely on existing or newly discovered and validated diagnostic genetic, proteomic, or transcriptomic biomarkers that also inform the selection of an efficacious medication for an individual or class of cancers.

- *Pharmacometabolomic biomarkers*: These are novel approaches using metabolic profiles to predict drug actions for individuals. For example, metabolism of the commonly used medication, acetaminophen, can be interfered with by high levels of gut microbiome metabolite, p-cresol. Thus, individuals with high pre-dose urinary p-cresol sulfate had low post-dose urinary ratios of acetaminophen sulfate to acetaminophen glucuronide. This reduced excretion of acetaminophen sulfate is because of competition between p-cresol and acetaminophen for O-sulfonation [21].

1.13 Epigenetic Modifications and Cancer

Conrad H. Waddington (1905–1975) introduced the word epigenetics in the early 1940s as an ontological term or phenomenon in organismal developmental control. The term derives from the Aristotelian word "epigenesis" meaning that an embryo develops through successive differentiation of undifferentiated egg cell, rather than enlarging (akin to inflating a balloon) of a preformed entity. The word was initially used to explain the relationships between genotype and phenotype and later as a global process of gene regulation in regard to cellular differentiation and development. The cardinal feature of the epigenetic phenomenon was the fact that the observed regulatory mechanisms were inexplicable by conventional genetics. Thus, the process is occurring above (*epi*) and beyond genetics. Berger et al. in 2009 defined epigenetics as ". . . a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequences" [22]. Thus, a feature of epigenetics from this definition is that the effects on altered gene expression are independent of changes to the structure of the DNA or RNA sequences. The discipline of epigenetics is thus the scientific study of heritable gene expressional changes that are controlled by phenomenon other than the genomic sequence (*epigenetics*—epi, Greek for above, because it is controlled *above* the genetic structure). Factors regulating these heritable gene expressional changes are not fully understood, but chromatin structural control factors such as acetylation and methylation of histone proteins are implicated, and most of the studied events are cytosine methylation at CpG dinucleotides. The main modes of epigenetic control are:

- DNA methylation at CpG sites.
- Polycomb/trithorax protein complexes.
- Noncoding RNA control of gene expression.
- ATP-dependent chromatin remodeling by posttranslational histone modifications.

The most proven heritable epigenetic mechanisms are DNA methylation and polycomb/trithorax protein complexes. Heritability of other mechanisms such as histone modifications and noncoding RNA are yet to be proven. A newer mechanism of epigenetic control is internal compartmentalization of the nucleus into what

is referred to as "chromosomal territories." This phenomenon suggests that transcription of specific gene loci depends on their nuclear position. Thus, chromosomal architectural organization or changes in the nucleus exerts changes to how genes are expressed.

1.13.1 DNA Methylation

DNA methylation occurs on carbon 5 of cytosine residues, mostly when followed by a guanine base, the so-called CpG methylation (non-CpG methylation has been reported). Cytosine methylation can cause point mutations in the genome, because deamination of 5′-methylcytosine yields thymidine, a natural DNA base that can escape the DNA repair system. To prevent or minimize the occurrence of such mutations in the genome, there is general paucity of CpG dinucleotides in the human genome. Because active gene transcription tends to prevent DNA methylation, unmethylated CpG dinucleotides are more frequent in gene promoter regions. These "CpG clusters" are called "CpG islands," and they constitute about 1–2 % of the entire human genome.

DNA methylation reduces gene expression because of reduced accessibility of regulatory sequences to interaction by transcription factors. Thus, DNA methylation causes histone deacetylation and the organization of the chromatin into a dense compact structure. DNA methylation is required in the normal cell for:

- Gene imprinting.
- X chromosome inactivation.
- Stabilization of ALU and LINE repetitive sequences.

1.13.2 Histone Modification

These are complex modifications leading to changes in chromatin structure and gene expression. There are at least eight recognizable histone modifications: acetylation, methylation, phosphorylation, sumoylation, ubiquitination, ADP-ribosylation, deamination, and proline isomerization. These modifications commonly occur primarily at N-terminal residues of histone tails. However, well-characterized histone modifications are methylation of lysine and arginine and lysine acetylation. The chromatin remodeling complexes (a multi-subunit of up to 12 proteins) have either a chromodomain or a bromodomain. The chromodomains interact with methylated histones, while the bromodomains (acetyl-lysine-binding domains) bind acetylated histones.

Histone acetylation occurs on only lysines and is regulated by histone acetyltransferases (HATs). On the contrary, histone deacetylases (HDACs) can remove the acetyl group from lysine. The acetylation of histones reduces their

basicity through neutralization of positively charged histone tails. This relaxes the interaction of histones with the negatively charged DNA, thus opening up chromatin structure that then enables transcription factor binding.

During histone methylation, up to three methyl groups can be added to lysine and a maximum of two to arginine. Gene expression modulation by histone methylation depends on the type of residue and the number of methyl groups added. For example, histone H3 trimethylation of lysine 4 (H3K4me3) and monomethylation of lysine 9 (H3K9me1) and 27 (H3K27me1) enhance gene expression. However, dimethylation or trimethylation of lysines 9 (H3K9me2 or H3K9me3) and 27 (H3K27me2 or H3K27me3) hinders gene expression.

1.13.3 The Cancer Cell Epigenome

Normal DNA methylation status and patterns are reversed in cancer cells. Selective gene promoter hypermethylation and global demethylation and hypomethylation characterize the cancer cell. Disproportionate heavy cluster of CpG dinucleotides referred to as CpG islands are found in promoter regions of ubiquitously expressed genes such as tumor suppressor genes. In the normal cell, these regions are unmethylated enabling gene expression when given the appropriate signals.

Gene promoter hypermethylation of tumor suppressor genes, developmentally controlled homeobox genes, and caretaker genes are rampant in cancer. For example, promoter methylation and silencing of the tumor suppressor gene and cell cycle regulator *CDKN2A* are common in many malignancies. Noteworthy, some gene promoters are, however, hypomethylated, leading to increased expression rather than silencing as in the normal cell. These genes include those that encode heparinase, mesothelin, claudin4, S100A4, trefoil factor 2, maspin, PLAU/uPA, POMC, and PGP9-5. Gene-specific hypomethylation can initiate cancer (e.g., imprinting of IGF2 and other H19 genes). Additionally, hypomethylation can involve classical oncogenes (thus increasing their expression), or cancer invasion and metastatic genes such as synuclein-γ, to favor the spread of cancer. Gene methylation maps can prove useful as biomarkers for the classification of cancers into various subtypes and are very helpful in several aspects of cancer management including early detection and diagnosis, prognosis, and therapy selection and monitoring.

Approximately 45 % of the human genome is of highly repetitive noncoding DNA. These tandem and intersperse repetitive DNA sequences are derived from transposable elements or transposons. There are two main classes of transposons. First are retrotransposons, which are DNA sequences that migrate and insert into the genome through RNA intermediates. This class includes SINES and LINES. ALU sequences alone make up over 10 % of the genome. The second class consists of DNA transposons that migrate by excision and reinsertion into different parts of the genome. Global hypomethylation of these repetitive sequences is a cardinal feature of several neoplastic growths. Indeed, the cancer cell genome has up to 60 %

less 5mC than the normal cell. The effects of loss of cancer genome methylation include reactivation and transcription of transposable elements. Silenced LINES and ALU repeats can translocate and be inserted into other parts of the genome with possible interference with gene function, and the loss of imprinting of some genes favors tumor formation through predisposition to chromosomal instabilities.

DNA hypomethylation is more widespread than hypermethylation. These hypomethylated loci tend to span megabase regions involving multiple chromosomes, often in CpG-poor regions. The hypomethylated regions are often punctuated by hypermethylated promoter and non-annotated CpG islands. These regional gains and losses in methylation are associated with late replicating, lamina-associated nuclear regions and contain many genes with bivalent, chromatin promoter domains that are highly susceptible to CpG island methylation in cancer.

The epigenetic changes are at the forefront of carcinogenesis, being present before invasive tumors are formed. Additionally, epigenomic alterations promote cancer initiation. For example, promoter hypermethylation of several classical tumor suppressor genes is found in normal-appearing cells in cancer fields as well as in preneoplastic cells. Methylation of *BRCA1*, *APC*, and *VHL* underlies both familial and non-familial cancers of the breast, colon, and kidney, respectively. Methylation of *GSTP1* is common in normal-appearing prostate epithelial cells of cancerous prostate gland, and *CDKN2A*/*P16INK4A* is methylated in several preneoplastic lesions.

The loss of genetic imprinting is one example of how the epigenome is exploited for neoplastic transformation. Histone modifications and DNA methylation of imprinted genes prevent biparental expression of both alleles (only enables heritable expression of paternal allele with suppression of maternal pair). A notable and classic example is the role of loss of imprinted *IGF2* in tumor initiation. The loss of imprinting and hence biparental expression in cells results in the formation of excess or more than normal growth factor. Ample evidence suggests the increased abundance of this growth factor (IGF2) can trigger neoplastic changes or provide permissive states for easy transformation. Some notable features of *IGF2* loss of imprinting are:

- The creation of immortalized cells that are poised for easy transformation.
- Expansion and maintenance of progenitor cell pools easily amenable to cancer initiation.
- The presence of loss of imprinting in circulating leukocytes confers risk for cancer development.
- The demonstration of the concept of field cancerization in colon carcinogenesis.

Novel technologies for measuring the epigenetic changes are very sensitive such that only very little sample amounts are required. Hence, epigenetic biomarkers are very attractive for clinical applications using body fluids. Noteworthy, methylation assays are developed using body fluids such as plasma/serum, urine, sputum, bronchoalveolar lavage fluids, and stool.

1.14 Genetic Alterations in Cancer

The genome is altered in almost all types of cancers, and these aberrations occur at multiple levels. The genomic structure can be abnormal in the cancer cell at the following levels:

- Chromosomal structural variations. Breakages within a single or between multiple chromosomes can lead to structural abnormalities. The consequences of these abnormalities are chromosomal translocations or inversions resulting in gene fusions.
- One or few nucleotide variations. One or few nucleotide insertions or deletions; simultaneous deletion and insertions of one, but mostly few nucleotides (indels); and consecutive nucleotide duplications can occur. The changes may maintain the coding reading frame (in-frame mutations) or change it (frameshift mutations).
- Point mutations or single nucleotide variations are base substitutions at a single nucleotide that may cause insertion of a different amino acid (missense mutations) or create a premature stop codon (nonsense mutations).
- Gene copy number variations including whole gene amplifications or deletions can occur, and there can also be deletions or duplications of large genomic sequences that may involve exon(s) of a gene creating exon copy number variations.

1.14.1 Chromosomal Alterations

The 6.8×10^9 bp human nuclear genomic DNA is compactly organized as chromosomal structures in the nucleus. This large amount of genetic material is scattered on 46 chromosomes (22 pairs or 44 autosomes and one pair or two sex chromosomes). For each chromosome, the DNA strand is wrapped around core histone proteins involving 146 bp of DNA. This initial compact structure is called the nucleosome, which in turn spirals at 30 nm diameters as solenoid fibers. More compacted spiraling, visualized by the light microscope, is referred to as chromatin fibers. Each chromosome basically comprises of two chromatids connected at a constriction called centromere. The centromere is a reference point for denoting the two regions or parts of the chromosome, a short or p arm (p for petit) and a long or q arm (q for the next alphabet after p). Microscopic examination of cancer cell chromosomes has revealed abnormalities in their structure (deletions, gains, and translocations) and numbers (aneuploidy).

1.14.1.1 Chromosomal Instability

Chromosomal instability (CIN) can give rise to the variation in numbers (aneuploidy), which characterizes most human cancers. Several mechanisms allow for defective chromosomal segregation during cell division. These mechanisms include defective telomere metabolism, centrosome amplification, dysfunctional centromeres, and/or defective spindle checkpoint control. Two mechanisms that are thought to be the first step in the development of CIN involve telomere length abnormalities, which result in telomere aggregation, breakage-bridge-fusion cycles, and centromere fragmentations and fusions. The resulting aneuploidy can autocatalytically contribute to greater CIN, leading to genomic instability.

1.14.1.2 Microsatellite Alterations

Microsatellite alterations (MSA), or allelic imbalance (AI), are deviations from the normal one to one ratio of alleles. In general, they include microsatellite instability (MSI) and loss of heterozygosity (LOH). Loss of heterozygosity is common in cancer cells and their precursors, but it has also been demonstrated in normal tissues such as normal mammary tissue adjacent to breast cancer. Microsatellites are satellite DNA interspersed throughout the human genome. Specifically, these consist of two to four nucleotide units that occur in tandem repeats. Miscopying or errors caused by DNA polymerase during replication can either expand or shorten microsatellites leading to MSI, which is a common feature of cancer cells. The accelerated proliferation of these cells requires expedited and efficient polymerase functions, the lack of which increases the likelihood of such errors.

1.14.1.3 Gene Amplification

Duplication of chromosomal regions that contain genes can occur in somatic cells or in the germline, giving rise to molecular evolutionary changes. This process often leads to acquisition of new or more genetic materials. Errors in DNA replication and repair mechanisms commonly give rise to gene duplication in somatic cells, and this can result in oncogene amplification and overexpression in cancer.

1.14.1.4 Telomere Dysfunction

The terminal 15–20 kilobase pair of DNA at each end of chromosomes is comprised of the unique TTAGGG repeat sequences referred to as telomeres. A specific DNA polymerase called telomerase produces these repetitive sequences. Telomeres are important in maintaining chromosomal integrity. They prevent end-to-end

chromosomal joining, nuclease digestion, activation of cell cycle checkpoints, and recombination. The classic experiments by Hayflick and Moorhead [23] demonstrating that fibroblasts grown in culture have a limited replicative potential due to age-related loss of telomeres contributed enormously to our knowledge of how loss of telomerase activity with aging causes cellular senescence. While normal telomerase activity attempts to maintain normal telomeres, elevated telomerase activity in a cell can cause cellular immortalization, cancer initiation, and progression. Not surprising then, many cancer cells overexpress telomerase.

1.14.2 Gene Mutations

Since the publication of the reference human genome sequence in 2002, DNA sequencing costs have plummeted partly due to technological advancements such as modifications to next-generation sequencing. This feat has enabled the rapid sequencing and cataloguing of whole cancer genomes.

A number of international consortia and freely available databases provide enormous information on cancer genomes (currently > 1000 cases of cancers have been sequenced). For example, the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) focus on cataloguing somatic variations in cancer genomes, while the OncoArray Network is making available a genome-wide data on germline variations. The COSMIC database also houses somatic cancer mutations. These databases have been and continue to be analyzed in detail to identify important cancer-causing mutations. The in-depth analyses of the available sequenced data reveal the complexity of the cancer cell genome to include these features:

- Compared to the normal cell, the cancer cell genome is studded with multiple mutations. On average, there are ~3000 mutations uncovered, of which ~300 are located in exons or protein-coding regions.
- Of the mutations in exons, 30–60 are protein altering or non-synonymous mutations. These non-synonymous mutations vary in number between tumors, with tumors of known mutagens tending to harbor more mutations than those without definite mutagens. For example, tobacco smoke-induced lung cancer and UV-mediated melanoma have hundreds of these mutations.
- The mutations fall into two broad categories, namely, "driver" and "passenger or bystander" mutations. Many of the non-synonymous mutations are passenger or bystander mutations that do not contribute to tumor initiation or progression.
- Two independent investigators suggest there are between 120 and 140 "cancer driver" genes in the genome. Vogelstein et al. [24] uncovered 138 "cancer driver" genes from analyses of the COSMIC database, while 127 were identified in the TCGA database by Kandoth et al. [25]. Sixty-six genes were common between the two studies. A "cancer driver" gene by definition must be mutated in more than one type of cancer, suggesting its important role in driving malignant

progression. On average a tumor harbors 2–8 mutations in any of these driver genes.

- The cancer driver genes fall into two functional groups – oncogenes and tumor suppressor genes (TSGs). The types, distribution, and functional consequences of the mutations have enabled this dichotomy to be reached. Oncogene mutations tend to be missense non-synonymous mutations that change amino acid residues leading to enhanced functions of the protein (gain of function mutations). On the contrary, TSG mutations are often found all over the coding regions as missense or nonsense mutations, which lead to loss of protein function (loss of function mutations). Of the 138 possible cancer driver genes identified by Vogelstein et al., 64 and 74 are oncogenes and TSGs, respectively.

All the cancer driver genes are altered in some fashion in different types of cancers. However, when considering genes significantly mutated in >5 % of cancers (an arbitrary cutoff), the top 10 genes in decreasing order of frequency from analyses by Kandoth et al. [25] are *TP53* (42 %), *PIK3CA* (17.8 %), *PTEN* (9.7 %), *APC* (7.3 %), *VHL* (6.9 %), *KRAS* (6.6 %), *MLL3* (5.9 %), *MLL2* (5.4 %), *ARIDIA* (5.4 %), and *PBRMI* (5.4 %). These findings are important in view of our yearning efforts to identify the best targets for therapeutic interventions in cancer.

1.14.2.1 Cancer Genome Consortia

There are a number of multinational organizations that are making it not only possible but also easy to obtain high-quality genomic data by scientists for further research. In regard to cancer, these international consortia are working independently and cooperatively in obtaining and cataloguing huge amounts of cancer genomes. Below are some of these organizations and their primary focuses.

The ICGC is concerned with coordinating a large number of research projects with the common focus on comprehensive deciphering the genomic changes that characterize various malignancies. In addition, the ICGC facilitates communication among its members and provides a forum for research coordination to maximize efficiency among the scientists. The primary goals include cataloguing the various somatic mutations, abnormal gene expression, and epigenetic modifications in tumors from 50 different cancer types and/or subtypes. The data is available to the entire research community as rapidly as possible with the aim of accelerating further research into the causes and control of cancer.

The Cancer Genome Atlas is a collaborative effort of the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), with the aim of generating comprehensive, multidimensional maps of the key genomic changes in major types and subtypes of cancer. Its initial 3-year pilot project in 2006 confirmed that an atlas of genomic changes could be created for specific cancer types. Moreover, it demonstrated that a national network of research and technology teams working on distinct but related projects could pool the results of their efforts together and make them publicly accessible, for further research into

cancer. The consortium finalized a collection of cancer and matched normal tissues from 11,000 patients for the comprehensive characterization of 33 cancer types and subtypes, including ten rare cancers.

The OncoArray Network was formed to develop a new custom genotyping array (the "OncoArray"). The OncoArray Network brings together multiple disease-based consortia and is funded by grants from multinational organizations. The OncoArray is a custom array manufactured by Illumina. It includes approximately 570 K single nucleotide polymorphism (SNP) markers. The backbone consists of approximately 260,000 SNPs that cover most common variants. Additionally, there are markers of interest for each of the five diseases identified through genome-wide association studies (GWAS), fine mapping of known susceptibility regions, sequencing studies, and other approaches. The array also includes loci of interest identified through studies of other cancer types and other interesting loci such as those involved with cancer-related phenotypes, drug metabolism, and radiation response. Additionally, SNPs concerned with quantitative phenotypes such as body mass index, height, and breast density that are common cancer risk factors are included on the array. The OncoArray Network has collected more than 400,000 samples from existing studies and several biobanks, and genotyping began in October 2013.

The COSMIC database stores the vast amounts of somatic cancer genome mutation data being generated. There are two ways data is collected for storage. These are COSMIC expert curation and systematic screen data. It is important to recognize how data is collected and catalogued and as to how mutation frequencies are obtained in order to use the database properly. First, the expert manual curation data is a targeted screen data obtained by COSMIC experts from PubMed searches using relevant key terms. It includes comprehensive literature curation and relevant data point pertaining to the disease and is updated periodically. This data source gives more accurate mutation frequency data as mutation-negative samples are specified. Erroneous or incomplete data is not fully curated. The second data source also provides objective mutation frequency data. Data is from sources involving whole genome molecular profiling of diseases and also imports from TCGA and ICGC or uploads from publications involving large-scale genome data sets.

1.15 The Cancer Cell Transcriptome

The epigenetic and genetic events that characterize a cancer cell reflect on downstream gene expressional changes at the message and protein levels and subsequently produce a metabolite profile of the particular cancer. Specifically, changes in oncogenes and tumor suppressor genes result in increases and decreases in their transcript levels, respectively.

While mRNA alterations are rampant in many tumors, a few have been characterized in circulation. Examples include *tyrosinase* in melanoma; *hTERT, BIRC5*, *KiSS1*, *CCND1*, *BMI1*, and *TYMS* in breast cancer; and *EGFR* and *hTERT* in

lung cancer. However, considerably much has been achieved in oncology through global gene expressional profiling of cancer cells. These gene signatures enable cancer diagnosis, prognosis, treatment decision-making, and subtype classification. For example, the Oncotype DX® multigene breast cancer test offered by Genomic Health provides a score to guide treatment and make recurrence predictions. Oncotype DX® score for DCIS predicts risk of local recurrence, while a score for invasive cancer predicts benefits from chemotherapy and risks for distant recurrence. Similarly, based on the field cancerization concept of carcinogenesis, Veracyte is developing a gene expressional signature (Percepta™ Bronchial Genomic Classifier) of normal airway cells from smokers to predict the risk of lung cancer. Several other gene expression signatures are clinically available for cancer management.

There are a few technical issues associated with expressional profiling of blood and body fluids. Exogenous transcripts in plasma are rapidly degraded by ribonucleases. However, it appears endogenous transcripts are protected from ribonuclease digestion by their association with small protective particles in plasma. For whole blood gene expression analysis, the shear abundance of globin RNA can easily mask RNA in transcriptionally active cells. In view of all these technical difficulties, many gene expression studies in blood have targeted transcript differences between patients and controls. The expressed transcripts in patients have mostly been attributed to CTCs, which often represent advanced metastatic diseases. The CE-marked PAXgene™ Blood RNA Tube standardizes and integrates important steps in whole blood collection, nucleic acid standardization, and RNA purification that enables preservation of intracellular RNA in circulation. This has enabled accurate gene expression profiling studies using blood samples.

In regard to expression profiling of blood for cancer biomarkers, the patented platform technology named the Sentinel Principle® from GeneNews is worthy of note. Developed in the 1990s by Dr. Liew Choong-Chin, it is based on the scientific tenet that circulating blood cells harbor expressional signatures reflective of conditions of the body in physiologic or pathologic states. It assumes that:

- As blood percolates through tissues, there is communication between blood and tissue cells.
- Changes in tissue from pathologic states cause specific changes in blood cell gene expression.
- These disease-associated changes can then be profiled and validated as biomarkers for the specific disease.

This platform is used to develop and commercialize diagnostic biomarkers. Developed and commercialized is the blood-based colon cancer screening test, ColoSentry®. This test is based on expression signature of seven genes with reported sensitivity of 61–78 % and a specificity of 66–77 %.

1.16 The Proteome or Peptidome

The human genome houses 20,000–25,000 protein-coding genes. However, due to alternative splicing, the use of alternative promoters, and posttranslational modifications, the human proteome is estimated at ~1 million. The discipline of proteomics deals with the study of the complete protein repertoire of a cell, tissue, or organ of an organism at a point in time. A challenge in proteomics studies is the fact that the human proteome is complex and demonstrates a large dynamic range, especially in plasma. This poses problems to protein quantification. The technologies available for proteomic studies enable the study of the proteome or peptidome repertoire, posttranslational protein modifications, and protein-protein interactions. Advances in technology now enable global analysis of tissue and body fluid proteome or peptidome.

1.16.1 Serum Peptidome

Analysis of serum proteins for cancer diagnosis and management has been around for several decades. Examples include the use of serum PSA for prostate cancer; CEA for breast, lung, pancreatic, and colorectal cancer; and CA125 for ovarian cancer detection and management. These biomarkers are still used in clinical practice despite concerns about their accuracy.

The serum proteome comprises the myriads of or composite proteins in an individual's serum (Fig. 1.14). This serum proteome is primarily made up of low molecular weight protein biomarkers and their metabolites referred to as the serum peptidome, fragmentome, or degradome. The field of serum proteomics was catapulted by the seminal work of Petricoin and Liotta that SELDI-TOF-MS peaks achieved sensitivity and specificity each of 95 % and a PPV of 94 % for ovarian cancer detection [26]. Since this report, the field has seen several technological challenges and improvements including qualitative and quantitative proteomics.

Mass spectrometry-based technologies have been used extensively for serum proteomic analyses. An advantage of MS, unlike gel-based approaches for protein separation, is that it can resolve peptides less than 10 kDa in size. Because of this advantage, MS has been employed for mining low molecular weight molecules as disease-associated biomarkers. Another advantage of proteomics studies using MS is that, unlike the other methodologies, such as microarrays, prior knowledge of the peptide or protein amino acid sequence is dispensable in analyte detection. As such, biomarker signatures, often of multiple analytes detected, simultaneously constitute defined classifier of disease presence, prognosis, or other clinical endpoints.

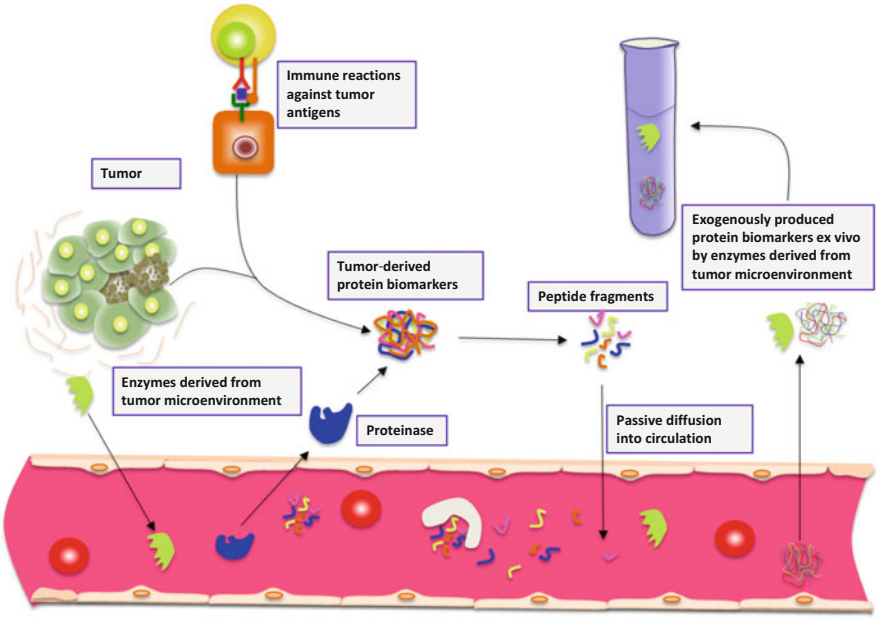

Fig. 1.14 The serum proteome. The serum proteome/peptidome is generated by diverse mechanisms. Cell death can lead to the release of cancer cell-specific proteins into the ECM. Similarly, immunologic reactions to cancer antigens in the tumor microenvironment can lead to the release of tumor-related proteins. These proteins can be acted upon by proteinases to generate a cancer-specific peptide signature that can diffuse into the circulation. Low molecular weight proteins can be protected by sequestration in albumin to evade renal clearance. Alternatively, enzymes derived from tumor microenvironment can act on exogenously produced proteins to generate a peptide signature of the specific cancer

1.16.2 Serum vs. Plasma Proteomics

It is still an unresolved issue as to which should be the preferred blood sample for proteomic studies to enable disease-specific biomarker identification. Researchers have used both types of sample to interrogate the peptidome for obvious reasons such as sample availability or the type of research question being asked. Indeed, some scientists have criticized the circulating peptidome as being uninformative.

The work by Villanueva et al. proposes that most cancer-specific peptidome signatures are produced by enzymatic cleavage at known endoprotease sites in ex vivo blood, as a consequence of the coagulation and complement fixation process [27]. They argue that disease-specific proteinases derived from within the tumor and/or tumor microenvironment perform these cleavages ex vivo, thus giving rise to cancer-associated peptidome profile. The peptidome in this case is a surrogate biomarker of the disease. Blood clotting is central to the ex vivo production of these low molecular weight cancer biomarkers, and hence, blood clotting should not be suppressed with proteinase inhibitors. The human proteome organization

recommends plasma as the preferred fluid for circulating proteomics. However, Kojima et al. achieved resolution between pancreatic cancer patients and controls only with serum and urine samples but not plasma [28]. Thus, the verdict as to which sample to use is still out there and partly depends on the research question.

1.16.3 Issues with Proteomic Biomarker Studies

Apart from the dilemma as to which blood sample to use, there are several issues with current proteomic studies that hamper reproducibility and comparison for authentic biomarker discovery. Some of these drawbacks apply to other biomarker development processes, while others are unique to proteomic approaches. These include:

- Non-standardized modes of sample collection, handling, storage, transport, and processing.
- Interlaboratory and array-to-array variability.
- The use of diverse bioinformatics approaches.
- Some proteomic approaches use only m/z peaks as biomarker signatures without protein identification.
- Cancer heterogeneity and hence lack of specificity of protein biomarkers identified.
- Different study designs and the use of controls.
- Lack of platform validation and standardization (e.g., protein microarrays currently in use vary among different platforms).
- False-positive rate (type I error) due to small sample sizes.

The proteome is, however, very informative for various diseases including cancer. With improving technologies and standardization, body fluid proteins that are specific to many cancer types are being discovered for use in patient care.

1.17 The Cancer Cell Metabolome

Metabolic alterations of the cancer cell are well established since the seminal observations reported by Otto Warburg in 1927 [29]. Glucose addiction by the cancer cell is a generalized metabolic phenotype of a third of all cancers. Glutaminolysis (dependence on glutamine as a metabolic substrate) is another established metabolic alteration of the cancer cell. Additionally, the cancer cell relies on the utility of nonessential amino acids such as arginine, proline, and serine. Other generalized metabolic phenotypes of cancer cells include increased phospholipid levels (i.e., elevated total choline containing compounds and phosphocholine) and increased glycolytic isoenzyme, specifically pyruvate kinase type M2 (M2-PK). Some of these metabolite differences between the cancer and

normal cell have been exploited for cancer diagnosis and treatment. For example, FDG-PET is developed for cancer detection based on the glycolytic phenotype. Thus, patients are administered ^{18}fluorodeoxyglucose (FDG), which is taken up by the glycolytic cancer cell and phosphorylated by hexokinase. Because of the lack of 2′ hydroxyl group necessary for subsequent metabolism, the molecule remains trapped in its phosphorylated state, and the radioactive emissions are measured with positron emission tomography (PET). False-positive rates (due to acute inflammation), cost prohibition, and short half-life hinder its generalized use. However, FDG-PET is clinically useful for the detection of lymphomas and other cancers.

Developing tumor-specific metabolic assays for clinical use has been tricky because many of the changes fluctuate reflecting various physiologic adaptations. However, promising metabolites are probably those that integrate genetics with metabolism, such as metabolites altered as a consequence of mutations in genes encoding enzymes involved in their metabolism. For example, loss of function mutations in fumarate hydratase, and succinate dehydrogenase associated with the accumulation of fumarate and succinate, respectively, in the cancer cell, or increased 2-hydroxyglutarate (2HG) levels as a result of oncogenic mutations in isocitrate dehydrogenase (IDH).

In addition to the single metabolite assays, the "omics" science is applicable to metabolite profiling. Metabolic activities (anabolic and catabolic) leave signatures that can be measured in body fluids and tissues. The terms metabolomics and metabonomics deal with the discipline of studying metabolites. The discipline of metabolomics applies to the measurement of the concentrations or amounts and location of cellular low molecular weight ($<$1 kDa) metabolites. The complete collection of the end products of cellular, tissue, or organismal enzymatic activities (metabolites) defines the metabolome. Metabonomics is a subdiscipline of metabolomics that deals with the measurement of cellular metabolic response to pathophysiologic stimuli (environmental or nutritional stress) or genetic variation. Specifically, metabonomics dwells on the study of the interactions between metabolites over time as a result of these stimuli. The two terminologies are sometimes, however, used interchangeably. The concentrations of these metabolites measured at a given time, which is reflective of the state of the organism at that time, constitute its metabolic "fingerprint." Metabolic "profiling" focuses on quantitative measurement of metabolites associated with a particular metabolic pathway (more limited than fingerprint).

To study body fluid metabolome, the repertoire of metabolites will be a reflection of both endogenous and exogenous sources. Altered cancer cell metabolism will produce metabolites of endogenous changes in metabolism. Additionally, factors such as diet, medication intake, and the microbiome are possible sources of exogenous metabolites. The microbiome of an individual consists of $>$100 trillion microorganisms from 300 to 500 different species and is unique to every individual ("microbiome fingerprint"). This microbiome changes with age, diet, medication intake, health status, and various medical and surgical interventions. Metabolites from the microbiome contribute significantly to the metabolome of an

individual. For example, the measurement of stool microbiome of colorectal cancer patients will constitute metabolites from both gut microbiome and colorectal cancer cells.

A powerful way of harnessing authentic biomarkers of cancer will be the integration of various "omics" data. Because "upstream" changes in genes and proteins reflect on "downstream" changes in cellular metabolism, systems biologic approaches, which deal with the interactions between interconnected networks involved in changes in the genome, transcriptome, proteome, and metabolome, should be the future of biomarker mining.

1.17.1 Body Fluid Metabolomics

Important biomarker targets poised to revolutionize body fluid utility for cancer signatures are the measurements and analysis of metabolites. Similar to other "omics" biomarkers such as gene expression, metabolomic biomarkers are influenced by the physiologic and pathologic conditions of the biologic system under study. The well-accepted methodological approach to metabolomics includes analytical techniques such as MS and NMR spectroscopy, as well as pattern recognition and bioinformatics. Metabolomics is aptly applied in tissues. However, it is even much easier in regard to acquisition and sample preparation, to target analyses of body fluids such as blood, urine, saliva (commonly assayed fluids), and others such as CSF, lung aspirates, digestive fluids, synovial fluids, amniotic fluid, seminal fluid, fecal water, ascetic fluid, cyst fluids, and dialysis fluid. For optimal comparisons, sample collection should take into account physical activity, diet, age, medication use, and other patient and control individual characteristics that can confound metabolite profiles. Body fluid volumes are usually within the range of 0.1–0.5 ml. For NMR spectroscopy, urine requires minimal preparation; however, blood, serum, or plasma requires acid, acetonitrile, or methanol/chloroform extraction. Whereas the entire human metabolome is yet to be deciphered and catalogued, much progress has been made in recent times with the discovery of thousands of metabolites.

1.18 Summary

- The study of cancer at the molecular level adds substantial knowledge and resolution to cancer biology.
- Cancer is a disease caused by exogenous and endogenous mutagens (epimutagens).
- These mutagens (epimutagens) cause alterations in the epigenome and genome that affect gene functions and genomic integrity.

- Altered oncogenes and tumor suppressor gene functions impinge on cellular signaling networks that establish and sustain the hallmarks of cancer to orchestrate malignant transformation and progression.
- Genomic efforts are unraveling the complex heterogeneity of cancer, such that cancer of the same histology can even show dramatic differences at the molecular level.
- Therapies are currently being developed to target specific molecular genetic alterations in cancer, ushering in personalized oncology. This treatment approach enables cancer cells to be targeted based on their specific molecular nature and for therapies to be changed as and when the molecular landscape of the same cancer cell changes.
- The "omics" sciences allow global snapshot of the cancer cell epigenome, genome, transcriptome, proteome, and metabolome. Such profiling studies are providing insights into cancer prognosis, treatment predictions, and also subclassification.

Internet Resources

- ICGC: http://icgc.org/
- TCGA: http://cancergenome.nih.gov/
- OncoArray Network: http://epi.grants.cancer.gov/oncoarray/
- COSMIC: http://cancer.sanger.ac.uk/cosmic
- Cancer Proteome: http://www.proteinatlas.org/humanproteome/cancer

References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.
2. Boveri T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci. 2008;121(Suppl 1):1–84.
3. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. J Natl Cancer Inst. 1960;25:85–109.
4. Cox AD, Der CJ. Ras history: the saga continues. Small GTPases. 2010;1:2–27.
5. Ames BN. Dietary carcinogens and anticarcinogens. Oxygen radicals and degenerative diseases. Science. 1983;221:1256–64.
6. Weinstein B. Relevance of the concept of oncogene addiction to hormonal carcinogenesis and molecular targeting in cancer prevention and therapy. Adv Exp Med Biol. 2008;617:3–13.
7. Semenza GL, Wang GL. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. Mol Cell Biol. 1992;12:5447–54.
8. Kroemer G, Galluzzi L, Vandenabeele P, et al. Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. Cell Death Differ. 2009;16:3–11.
9. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer. 1972;26:239–57.

10. Raffeld M, Wright JJ, Lipford E, et al. Clonal evolution of t(14;18) follicular lymphomas demonstrated by immunoglobulin genes and the 18q21 major breakpoint region. Cancer Res. 1987;47:2537–42.
11. McDonnell TJ, Deane N, Platt FM, et al. bcl-2-immunoglobulin transgenic mice demonstrate extended B cell survival and follicular lymphoproliferation. Cell. 1989;57:79–88.
12. Nunez G, Seto M, Seremetis S, et al. Growth- and tumor-promoting effects of deregulated BCL2 in human B-lymphoblastoid cells. Proc Natl Acad Sci USA. 1989;86:4589–93.
13. Korsmeyer SJ, McDonnell TJ, Nunez G, et al. Bcl-2: B cell life, death and neoplasia. Curr Top Microbiol Immunol. 1990;166:203–7.
14. Bakhshi A, Jensen JP, Goldman P, et al. Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. Cell. 1985;41:899–906.
15. Cleary ML, Sklar J. Nucleotide sequence of a t(14;18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. Proc Natl Acad Sci USA. 1985;82:7439–43.
16. Tsujimoto Y, Gorham J, Cossman J, et al. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. Science. 1985;229:1390–3.
17. McDonnell TJ, Korsmeyer SJ. Progression from lymphoid hyperplasia to high-grade malignant lymphoma in mice transgenic for the t(14; 18). Nature. 1991;349:254–6.
18. Vaux DL, Cory S, Adams JM. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. Nature. 1988;335:440–2.
19. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953;6:963–8.
20. Druker BJ. Imatinib: a Viewpoint by Brian J. Druker. Drugs. 2001;61:1775–6.
21. Clayton TA, Baker D, Lindon JC, et al. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci USA. 2009;106:14728–33.
22. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes Dev. 2009;23:781–3.
23. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. Exp Cell Res. 1961;25:585–621.
24. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. Science. 2013;339:1546–58.
25. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333–9.
26. Petricoin EF, Liotta LA. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotechnol. 2004;15:24–30.
27. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271–84.
28. Kojima K, Asmellash S, Klug CA, et al. Applying proteomic-based biomarker tools for the accurate diagnosis of pancreatic cancer. J Gastrointest Surg. 2008;12:1683–90.
29. Warburg O, Wind F, Negelein E. The metabolism of tumors in the body. J Gen Physiol. 1927;8:519–30.

Chapter 2
Advanced Technologies for Body Fluid Biomarker Analyses

Key Topics

- High-throughput technologies
- Next-generation sequencing
- Single-cell analysis
  - Whole genome amplification
  - Whole transcriptome amplification
  - Whole genome analysis of circulating tumor cells (CTCs)
- Proteomic and metabolomic technologies
- Lab-on-a-chip

Key Points

- The low abundance and dilution effects of body fluids on liquid biopsy biomarkers call for sensitive technologies for their detection and quantification.
- Profiling single cells such as CTCs or circulating cancer stem cells poses a much harder technical challenge. However, the ability to faithfully amplify a single-cell genome, coupled with deep coverage using next-generation sequencing technologies, ushers in a new era of high-resolution cancer genomics.
- Microfluidic and nanofluidic technologies enable the accurate and sensitive detection and quantification of liquid biopsy targets, using picoliter quantities of samples.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_2

2.1 Introduction

The age of personalized cancer management has dawned. Targeted therapies are clinically available to cancer patients, and although there are still challenges to overcome, they contribute to the successes that are being made in cancer treatment. To further advance our efforts, therapies should focus on the driving forces behind the progression of each individual cancer. The heterogeneity and complexity of cancer has been recognized for decades now. Additionally, stromal and "normal" epithelial cells confound tissue analysis. Thus, to improve on research outcomes, efforts had drifted from whole tissue analysis to microdissection, with the aim of isolating individual cancer cells. Yet, many technicians or operators have focused on multiple cancer cell dissection (to increase template amounts), which still faces the issue of cancer cell heterogeneity. The quality of microdissected cells also relies on the expertise of the pathologists or technician. Although of minor concern, tissue thickness may preclude perfect homogeneous sampling of cancer cells.

Cancer biotherapies target important genetic alterations and pathways that drive disease progression. Hence, treatment planning requires prior knowledge of the genotype of specific cancers. This feat has easily been accomplished by examining tumor biopsies. Unfortunately, tumor heterogeneity and sensitivity levels of detection technologies may not permit capturing of the entire or relevant tumor genome. Besides, targeted therapies mainly aim at killing metastatic cells. However, metastatic cancer cells (and their stem cells) undergo genetic evolution with acquisition of novel molecular alterations.

For all the above reasons, "liquid biopsy" offers optimal applications in cancer management. Unfortunately, liquid biopsy faces tough analytical challenges. Biomarkers in circulation and other body fluids are technically more difficult to measure because of dilution effect, degradation, wide dynamic range (especially for proteins), and masking by high-abundant non-tumoral biomolecules. Novel technologies to overcome these hurdles are rapidly being developed and optimized. They include sensitive high-throughput genomic, proteomic, and metabolomic technologies, as well as those for the capture and analysis of shed cancer cellular particles and circulating tumor cells. Methods for single-cell genomics and transcriptomics have also emerged. While there are technical issues to overcome, they offer tremendous impetus into not only enabling the biology of cancer to be well understood but also the adoption of suitable treatments with disease progression and emergence of new mutations.

While deaths from cancer have not declined at the same frequency as those from cardiovascular diseases (CVDs) since the early 1990s (39 % for CVDs compared to 21 % for cancer), considerable progress toward the war on cancer has been made, partly as a consequence of accrued immense molecular genetic data on various types of cancer. However, for several reasons, the mortality reductions achieved are not globally uniformly distributed. Cancer mortality figures show wide geographic disparities, with >70 % of all cancer deaths occurring in the less developed or

resource poor parts of the world. Thus, the disturbing trends in global cancer statistics have been noted:

> By 2030, it is projected that there will be 26 million new cancer cases and 17 million cancer deaths per year. Moreover, the global distribution of cancer and types of cancer that predominate continues to change, especially in economically developing countries. Low- and middle-income countries accounted for about half (51 %) of all cancers worldwide in 1975; this proportion increased to 55 % in 2007 and is projected to reach 61 % by 2050. Cancers of the lung, breast, colon/rectum and prostate are no longer largely confined to Western industrialized countries but are among the most common cancers worldwide. (Thun et al. [1])

Even more worrisome is the emerging evidence that the less developed nations will bare much of the cancer burden. Indeed, the 2008 World Health Organization's *world cancer report* reveals a similar disturbing pattern. While incidence and death rates will show a modest decline in developed nations, the reverse is occurring in the developing world, such that deaths from cancer will lead cumulative deaths from HIV/AIDS, tuberculosis, and malaria. Contributing factors to this global trend include lifestyle changes such as increase dependence on Western diet, tobacco and alcohol use, as well as ineffective treatment of infectious diseases such as *Helicobacter pylori* gastritis. All these factors are also causative agents of precancer fields or "field cancerization." Arguably, lifestyle alterations appear to be the most important factor that could help reverse this global catastrophe. However, employing other strategies such as targeting genetic footprints of carcinogens has potential for early detection and diagnosis, as well as effective cancer management. Thus, to reverse the emerging scenario, cancer, I believe, must be tackled at its "*genesis and not exodus*."

Validated (Oxford Center for Evidence-Based Medicine level 1a) biomarkers in body fluids that can be incorporated in lab-on-a-chip platform for point-of-care deployment will help reduce the current geographic disparities in incidence and mortality of cancer in the world. Several efforts aimed at achieving these goals are far approaching reality, as some industries are making these devices available for the management of diseases including cancer. Platform technologies including microfluidics, nanofluidics, and biosensors usher in the next generation of biomarker detection and measurements at the miniaturized levels. Various technologies for biomarker analyses are explored in this chapter. However, some technologies appropriately appear in other chapters as well.

2.2 High-Throughput Technologies for Biomarker Analyses

The "omics" sciences offer a number of high-throughput technologies for assaying biomarkers in body fluids, cells, and tissues. These technologies range from the detection of "omics" fingerprints of the epigenome to the metabolome, each with its strengths and weaknesses. Also being advanced are biosensors, biochips,

microfluidics, molecular labels, as well as novel materials and multifunctional platforms comprised of super-paramagnetic nanoparticles coupled with quantum dots, dyes, or gold nanoparticles and the proper surface modifications for specific targeting. These technologies are constantly being improved upon for specific detection and measurement of very low-abundant targets.

Undoubtedly, the PCR has revolutionized molecular biology, and its modifications are commonly employed in epigenetic, genetic, and RNA analysis using body fluids. High-throughput technologies have equally been applied for detection of alterations in the epigenome, chromosome, genome, transcriptome, and proteome. The epigenome can be examined by methylation and acetylation microarrays. Gene copy number variations are determined using various methods such as comparative genomic hybridization, SNP arrays, molecular inversion probes, digital karyotyping, representational oligonucleotide microarray (ROMA), and NGS. MassARRAY (Sequenom) GWAS and NGS allow HTP mutation detection. Exon arrays and genome tiling arrays unravel alternative splicing, while gene expression arrays enable HTP expressional analysis. Mass spectrometry enables circulating proteins/peptides and metabolite analysis. Traditionally, immunoassays are used in research and clinical settings for circulating protein analysis. While mainly developed based on the use of tissue samples, many of these devices have also been used to measure targets in body fluids.

This chapter, however, focuses on microfluidics and other miniaturized technologies that enable accurate analyses of even picoliter amounts of samples, as these are more adept to body fluid biomarker analysis. Moreover, numerous biomedical engineering devices are developed for biomarker analyses mostly using tissue samples, but some of these have been successfully applied to body fluid analyses as well.

2.3 Next-Generation Sequencing

Next-generation sequencing (NGS), massively parallel sequencing, or deep sequencing has considerably advanced genomic analysis. The ability to apply rapid massively parallel sequencing of the entire genome or parts of it enables simultaneous detection of multiple alterations in a particular situation. The various modifications to this technology provide partial or complete information of genomic aberrations. The technology is applicable to the entire genome (whole genome sequencing, WGS); only protein-coding genes (whole exome sequencing, WES) or custom panels can be developed for specific purposes. Custom panels tend to be cheaper and faster to run, but are limited in scope because they are focused only on regions of interest, and hence unable to determine all genetic lesions. The amplicon capture custom platforms (e.g., Agilent HaloPlex, RainDance, Illumina TruSeq, and Life Technologies Ion Torrent AmpliSeq) detect 1–100 mutations per sample, while hybridization-capture custom panels (e.g., Illumina TruSeq, NimbleGen SeqCap, and Agilent SureSelect) detect from 50 to 1000s of mutations. In addition

to substitutions, duplications, insertions, deletions, and indels detectable by amplicon capture panels, the hybridization panels can also detect gene or exon copy number variations and predetermined translocations. But these custom panels tend to cover higher sequencing depth than WGS and WES approaches.

Genome-wide coverage at the exome or whole genome level is similar, except for the extent of comprehensiveness. Both can detect single nucleotide variants (e.g., substitutions), duplications, deletions, insertions, indels, as well as gene copy number variations. However, large-scale chromosomal aberrations (e.g., translocations and inversions) are more easily detected using WGS. Currently, the drawbacks on these technologies include high costs, large sample amount requirements, limited throughput, and need for complicated bioinformatics. Additionally, some platforms show imperfections and may introduce artifacts especially in regard to the detection of gene rearrangements. Depth of coverage is important in detecting low-abundant mutations, which is important for liquid biopsy analysis. With decreasing costs, NGS will be a routine clinical laboratory assay.

In principle, NGS technology requires initial sample preparation, generation of a library before loading onto specific platforms for sequencing, and subsequent specialized bioinformatics detection of sequencing data. Amplicon capture panels require multiplex PCR amplification of desired regions of interest, while hybridization capture and WES involve hybridization capture of regions of interest or all exons, respectively.

The DNA for sequencing is used to create a library. This involves:

- DNA fragmentation into small sizes, depending on platform requirements.
- Blunt ends are created on each fragment.
- Adapters are ligated to both 3′ and 5′ ends of the fragments. The ends of the adapters are blunt or have A/T overhangs.
- Adapters are unique to specific platform chemistry.

The libraries can be loaded directly onto some platforms and sequenced or pre-amplified prior to sequencing. Direct sequencing is possible on platforms that are sensitive enough to detect the signal. However, pre-amplification serves to increase the signal-to-noise ratio, enabling detection of incorporated nucleotides on platforms with reduced signal detection sensitivity. Depending on technology, sequencing occurs by synthesis (Illumina, Roche, Ion Torrent, and Proton), ligation (Life Technologies ABI SOLiD), or single-molecule synthesis (Pacific biosciences, Helicos). Detection of incorporated nucleotide signal can also be by fluorescence methods, photon release, or pyrophosphate detection. Each base is sequenced multiple times during the process, and the sum of reads of a particular nucleotide or base constitutes the depth of coverage of that nucleotide at that position.

NGS technology has multiple applications including methylome, genome, and transcriptome analyses, as well as meta-genomics (sequence it all). Meta-genomics applications are mostly focused on infectious diseases.

2.4 Single-Cell Analysis

Traditional cancer biology relies on analyses of tissues or hundreds to thousands of cells where biomarker concentrations are high. While such approaches have generated a wealth of information on the biology of cancer, we are at a time whereby more in-depth knowledge on cancer heterogeneity is needed. This yearning need is pushing technological boundaries into uncharted grounds, thus requiring single-cell analyses. Unless technologies are developed that can perform whole genome sequencing, or expression analysis using picogram amounts of genetic material, the nanogram quantities of required template by current high-throughput technologies such as next-generation sequencing and microarrays necessitate that the ~7 pg diploid single-cell genome or 1 pg mRNA be amplified prior to downstream applications. To achieve this, various methods have been developed for single-cell nucleic acid amplification.

2.4.1 Whole Genome Amplification

Single-cell whole genome amplification (WGA) relies on PCR strategies, multiple displacement amplification, or a combination of both. Each technique has its own advantages and shortcomings, as well as uses in specific downstream situations.

As will be expected, even given the best of enzymes, template, and other reaction conditions, faithful stoichiometric amplification of the genome is bound to have some imperfections. Downstream analyses of such artifacts will introduce errors in data interpretation. Various WGA strategies suffer from nucleotide copy errors, amplification bias due to differences in guanine and cytosine content in the genome, preferential allelic amplifications, polymerase errors in nucleotide reads, and allelic dropouts. The extent of chimeric DNA molecules in the genome also affects WGA.

Multiple displacement amplification (MDA) copies the single-cell genome using isothermal reaction (at 30 °C), a DNA polymerase that has strand displacement activity (e.g., phi29), and random primers (Fig. 2.1). Random priming of denatured genomic DNA is followed by genome copying in the 5′ to 3′ direction at various sites of the genome. When the 3′ end of a newly synthesized DNA strand reaches the 5′ end of a nearby primed sequence, the polymerase displaces the strand, creating a single-stranded DNA. This enables further priming and amplification of the displaced strand. Multiple displacement amplification suffers from preferential amplifications and allelic dropout and hence is difficult to ascertain copy number variations using this method. However, it is useful for SNP and mutation analysis.

The PCR-based methods include primer extension pre-amplification, degenerate oligonucleotide primed, and linker adaptor PCR strategies (Fig. 2.1). Primer extension pre-amplification PCR (PEP-PCR) uses complete degenerate sequence

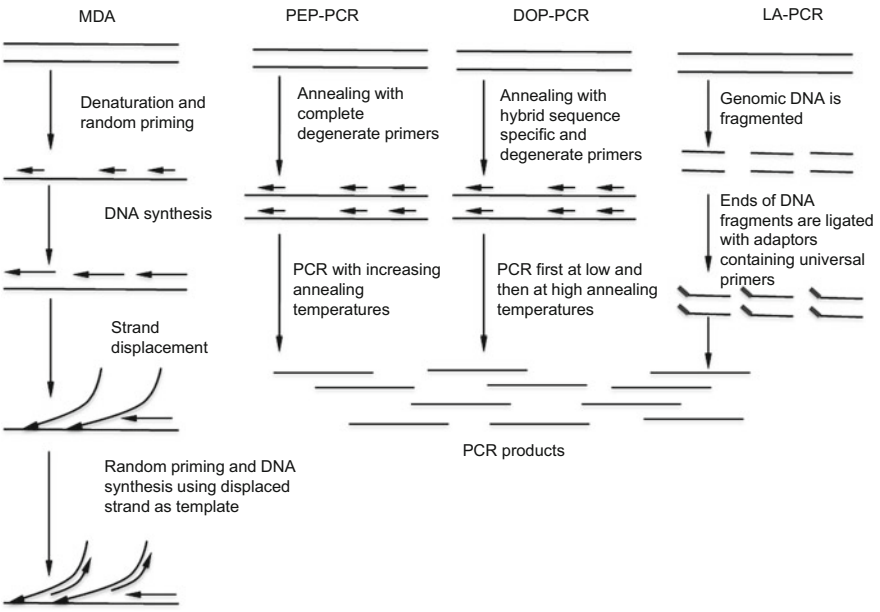

Fig. 2.1 Methods for whole single-cell genome amplification. MDA uses random priming and isothermal (30 °C) amplification to copy the denatured DNA. When the 3′ end of a newly synthesized DNA encounters the 5′ end of an adjacent primed segment, this segment gets displaced to generate a new single-stranded molecule, which is primed and copied as well. The PCR method uses three different approaches. PEP-PCR uses complete degenerate oligos in permissive thermocycling with increasing temperatures. DOP-PCR uses hybrid oligos consisting of both degenerate and defined sequence-specific nucleotides. Cycling begins at low annealing temperature for semi-random priming. This is followed by high annealing temperature for sequence-specific priming and amplification. LA-PCR uses DNA digestion and end ligation of DNA fragments with adaptors containing universal primers, which are used for PCR amplification

oligonucleotides of 15 bases to amplify the entire genome using permissive thermocycling with increasing annealing temperatures. Degenerate oligonucleotide primed PCR (DOP-PCR) uses oligonucleotides with both degenerate and genome-specific sequences. Two amplification steps are involved. An initial thermocycling at low annealing temperature enables semi-random priming and copying. A more stringent high annealing temperature to enable specific priming and genome amplification then follows. Finally, in linker adaptor PCR (LA-PCR), the cell's DNA is first randomly digested and then ligated with adaptors that have universal primer sequences. This is followed by genomic amplification using these primers.

Technologies such as PicoPlex® (Rubicon Genomics, Inc.) and multiple annealing and looping-based amplification cycles (MALBAC) integrate both MDA and PCR techniques (Fig. 2.2). These techniques use displacement pre-amplification to generate DNA fragments for subsequent PCR amplification. Conceivably, these methods more faithfully copy the genome and hence are more accurate for detection of copy number changes, mutations, and SNPs. PicoPlex uses

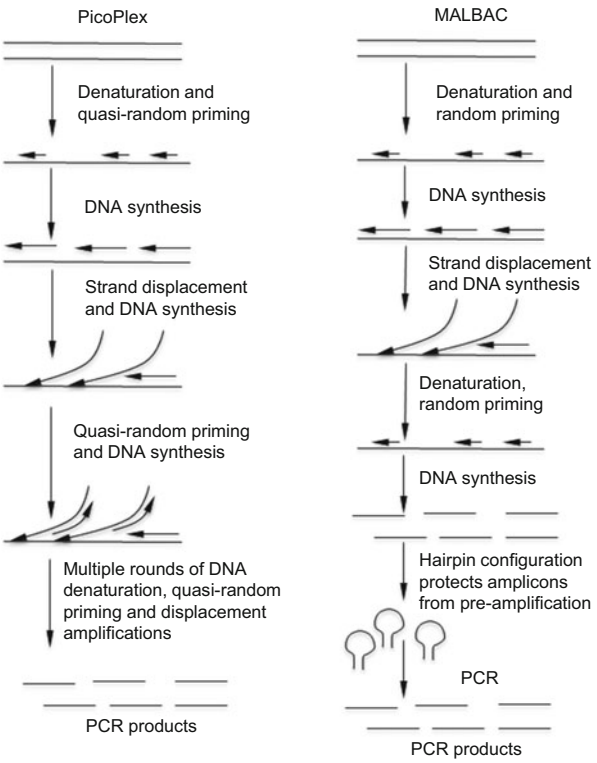

Fig. 2.2 PicoPlex and MALBAC. These approaches to WGA use the combination of displacement pre-amplification and PCR methods. PicoPlex uses hybrid oligos for displacement pre-amplification to generate amplicons for subsequent PCR. MALBAC uses specific oligos for displacement pre-amplification to generate amplicons that form looped structures to prevent copied genomic segments from further pre-amplification. These looped pre-amplified products are then subjected to PCR amplification

several rounds of displacement pre-amplification with hybrid oligonucleotides of degenerate and universal sequences to make amplicons for subsequent PCR. MALBAC also applies MDA method to generate pre-amplification genomic products. However, the primers have specific sequences that create looped pre-amplified products. These looped products serve to prevent further pre-amplifications of these segments and hence provide a safeguard for unfaithful copy number variations.

2.4.2 Whole Transcriptome Amplification

Single-cell transcriptomics adds to the study of cancer cell heterogeneity. Similar to DNA, the single-cell mRNA repertoire is only ~1 pg and hence must be amplified prior to downstream applications. A number of methods have been developed for whole transcriptome amplification (WTA), each with its advantages and shortcomings. The strategies for WTA include:

- STRT-seq: This method involves the use of oligo-dT adaptor primer for initial reverse transcription with incorporation of non-template (novel) nucleotides at the 5′ end. These novel nucleotides enable hybridization of bar-coded template-

switching adaptor primer for reverse transcription. The bar-coded reverse-transcribed products are pooled and subjected to PCR amplification for downstream uses. This technique preferentially amplifies 5′ end sequences (5′ bias).

- CEL-seq: This method also uses bar-coded oligo-dT primers with 5′ adaptor sequence and T7 RNA polymerase priming site. The cDNA generated is used as a template to create cRNA using T7 RNA polymerase. Amplified cRNA is used for downstream applications. This technique preferentially amplifies 3′ end sequences (3′ bias).
- Tang et al. developed a technique that has a better transcript coverage than the previous two methods (Tang/Surani method) [2]. This process involves the initial production of first-strand cDNA molecules, followed by 3′ polyadenylation. These molecules are then primed with adaptor-conjugated oligo-dT used to generate double-stranded DNA for downstream applications. This is also 3′ biased but covers much more transcript sequences than CEL-seq.
- Other methods are developed in an attempt to achieve whole transcript coverage. The SMARTer method is a novelty that uses template switching to create a whole library of full-length transcripts containing adaptor sequences at both ends. These adaptor sequences are targeted for PCR amplification of the transcriptome.

The imperfections of WTA include:

- Lack of whole transcript coverage by some methods.
- Amplification may be biased toward selecting for high copy transcripts.
- Low copy mRNAs ($\leq$10 copies/cell) are unfaithfully amplified, and yet 85 % of transcripts in a cell are between 1 and 100 copies.
- Inherent in all genome amplification strategies are base copy errors.

2.4.3 Whole Genome Analysis of Circulating Tumor Cells

Whole genome amplification techniques have been successfully applied to the study of CTC genomes. Although at its infancy, there are compelling data that reveal the potential of future clinical applications of the technology. Ramskold et al. developed a robust RNA sequencing protocol (Smart-Seq) with improved read coverage across transcripts [3]. This technique, which is applicable to single-cell analysis, was used to reveal distinct gene expression patterns in CTCs. The genetic interrelationships between CTC, primary tumor, and metastatic deposits from colorectal cancer patients indicated shared genomic gains and losses of CTCs with primary and metastatic tissues. Mutations in *APC*, *KRAS*, and *PIK3CA* were shared, but CTCs harbored additional "private" mutations. To ascertain that these were authentic mutations and not introduced by WGA errors, deep sequencing of primary tumors revealed the presence of these mutations at subclonal levels. This study further revealed the importance of CTC genome investigation, because in one patient, 9 out of 10 CTCs had amplifications of *CDK8* that were not present in the

primary tumor, suggesting this patient could have benefited from CDK inhibitor therapy [4]. In lung cancer, sequencing of whole genome amplified DNA showed that copy number changes of individual CTCs from different patients having the same tumor subtype were identical compared to patients with different lung cancer subtype [5].

The ability to analyze CTCs and circulating cancer stem cells hold great potential for the study of cancer metastasis and discovery of important therapeutic agents for drug development.

2.5 Proteomic Technologies for Body Fluid Analysis

Protein biomarkers of cancer are abundant in the circulation and other body fluids. It is estimated there are >10,000 different proteins in circulation. Based on their sources and mechanisms of release, these protein signatures differ between samples from cancer patients and healthy individuals. Thus, a number of technologies have been deployed to discover, validate, and translate body fluid protein biomarker signatures. However, the serum proteome is derived from two sources: (1) high-abundant endogenous proteins such as albumin and immunoglobulin that constitute ~90 % of serum proteins and (2) cell- and tissue-derived low-abundant and low molecular weight proteins. Thus, the serum proteome is complex, such that the high-abundant proteins tend to mask useful disease-associated low-abundant proteins. Various techniques have therefore been developed to deplete samples of the high-abundant body fluid proteins such as albumin and immunoglobulin, prior to downstream analysis. A number of pre-fractionation methods are used to deplete the high-abundant proteins, and these include:

- A simple heparin chromatography and protein G treatment of the sample can be used to deplete high-abundant proteins.
- Affinity chromatography with immobilized protein A or protein G is used to remove immunoglobulin.
- Peptide affinity chromatography, dye-ligand chromatography, isoelectric trapping, and immunoaffinity columns are employed to remove albumin.
- Immunodepletion and reverse-phase separation of the plasma on mRP-C18 columns are also employed.
- The commercially available method (multiple affinity removal system (MARS) Agilent) mixes six highly specific polyclonal antibodies to target the removal of the top six high-abundant proteins.
- Also, the human-14 multiple affinity removal column (Agilent) enables depletion of the top 14 abundant proteins in body fluids.
- Core shell bait-loaded nanoparticles enable selective entrapment and protection of low-abundant proteins. This process involves (1) the use of a sieve to eliminate high molecular weight proteins, and thus prevent them from entering the nanoparticle, (2) small molecules of low-abundant proteins are then captured

with high-affinity baits, and, finally, (3) the shell of the nanoparticles is coated with vinylsulfonic acid (VSA), which repels albumin, even those of small fragment sizes.

Various novel technologies and labeling techniques, each with its unique applications, strengths, and weaknesses, are used to study the proteome (Table 2.1). These include simple 2D polyacrylamide gel electrophoresis (2D-PAGE), difference gel electrophoresis (DIGE), large-scale Western blot, isotope-coded affinity tag (ICAT), isobaric tagging for relative and absolute quantification (iTRAQ), SILAC, shotgun proteomics, multiple reaction monitoring (MRM), protein microarrays, multidimensional protein identification technology (MudPIT), and mass spectrometry.

2.6 Gel Based (2D and DIGE)

The gel-based technology enables separation, quantification, and the study of posttranslational protein modifications. In two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), proteins are separated on polyacrylamide gels in two dimensions. The first dimension is based on charge, and the second on mass. Protein quantification can be achieved by labeling with cyanide dyes prior to separation or with other agents such as SYPRO Ruby, Coomassie, or silver stains after separation. Additionally, in-gel protein digestion can be performed followed by downstream identification using MS. In addition to quantification, this method enables protein molecular weight and isoelectric point (pI) determination as well as the study of posttranslational modifications.

In 2D difference gel electrophoresis (2D-DIGE), proteins are labeled with spectrally distinct cyanide dyes (Cy2, Cy3, Cy5) to lysine residues in proteins prior to 2D separation. For the DIGE, up to three samples can be labeled with the three dyes, mixed, and loaded together on the same gel. This permits quantitative comparisons to be made between the samples. Softwares, such as DeCyder, are available for image analysis. This method enables visualization of up to 1000 proteins, many of which can be identified.

2.7 Non-gel-based Proteomic Techniques

A number of non-gel-based approaches are developed that offer sensitive and high-throughput proteomic analysis. These tools, coupled with novel labeling techniques, are fostering the discovery of protein biomarkers of cancer. These biomarkers are often characterized as mass-to-charge (m/z) ratio, peak intensities, or actual protein/peptide identification.

Table 2.1 Some proteomic methods and technologies

Technology	Uses	Features
2D-PAGE	Separate proteins Quantitative proteomics Information on posttranslational modification	Not sensitive for detection of low-abundant proteins, proteins with extreme pH (<3 and >10) or weights (<10 kDa and >150 kDa), and hydrophobic proteins Poor separation of basic proteins Not reproducible
2D-DIGE	Same as 2D-PAGE	Highly sensitive Easy protocol Reproducible No bias from experimental variation Needs special visualization equipment Expensive fluorophores Proteins without lysine cannot be labeled
MS based	Protein identification Protein characterization	Used for both qualitative and quantitative analysis Can be used to study post-transformation protein modification Sensitive, specific, and high-throughput technology Identification of novel proteins Hydrophobic and basic proteins are not easily identifiable by ESI and MALDI technologies
SILAC	Isotope labeling of cells Differential protein expression	Ease and high degree of labeling Cannot be applied to tissue samples Limited to cells in culture
ICAT	Chemical isotope labeling of proteins Quantitative proteomics	Sensitive for quantification of low-abundant proteins Poor detection of proteins with less cysteine and acidic proteins Biased toward proteins with high cysteine content
iTRAQ	Isobaric protein tagging	High-throughput relative protein quantification Peptide fractionation before MS Multiplexing ability
MudPIT	Protein-protein interaction	Highly sensitive Enables identification of large protein complexes Achieves high protein separation Nonquantitative Nondiscriminatory between protein isoforms Generates huge data set needing detailed and complex bioinformatics for protein identification
Protein arrays	Diagnostics and discovery research involving known proteins	High-throughput sensitive technology Low amounts of samples required – good for body fluid proteomics Limited by poor expression of proteins Restricted to available antibodies

2.7.1 Mass Spectrometry

Mass spectrometry (MS) enables protein identification and possible protein sequence deciphering. Labeled proteins can directly be subjected to MS analysis. Gel-based approaches require spot excision, trypsinization, and possible liquid chromatography (LC) before MS analysis. In LC-MS/MS, proteases are used to specifically digest mixtures of proteins into peptides, which are then separated by one or two dimensions using LC coupled with automated tandem MS.

The MS technology requires protein ion source followed by separation techniques. Electrospray ionization (ESI) and matrix-assisted laser desorption ionization generate the required protein ions. Mass analysis is then accomplished using other methods such as time-of-flight (TOF), quadrupole, Fourier transform ion cyclotron (FTIC) resonance or ion trap. Different mass analysis instruments such as quadrupole-ion trap, quadrupole-TOF, or triple quadrupole used in tandem MS (MS/MS) increase protein resolution and possible deciphering of protein sequences.

Selected reaction monitoring (SRM) is a powerful tandem proteomic technology that increases the sensitivity and specificity of protein/peptide detection in multiple samples. In its simplest form, SRM first selects an ion of a particular mass, which is followed by a second selection of an ion product from the initially selected ion. This approach is very useful especially when selected set of proteins (e.g., candidate biomarkers) ought to be measured in multiple samples consistently and reproducibly and in a quantitatively precise manner. SRM has been useful in body fluid protein biomarker studies. There are variations to how ions are selected, and this has given rise to different versions of SRM, which are multiple reaction monitoring (MRM), consecutive reaction monitoring (CRM), and parallel reaction monitoring (PRM). Multiple reaction monitoring, for instance, involves the selection of multiple reaction product ions from one or more initial precursor ions. Thus, an initial mass analyzer transmits only the protein of interest to a second mass analyzer, where collision and fragmentation leads to the selection and transmission of a specific diagnostic biomarker to a third mass analyzer for eventual detection.

2.7.1.1 Matrix-Assisted Laser Desorption/Ionization (MALDI)-MS

This simple approach enables protein mass measurement. In this technique, the protein sample is mixed with the matrix in solution and co-crystallized on a metal surface. A laser ion source is used to excite the matrix and analyte ions, which are then released into the gaseous phase. MALDI generates peptide masses (mass fingerprint), and relative protein quantification can be achieved through measurement of mass spectral peak intensities or use of peptide and spectral counts.

2.7.1.2 Electrospray Ionization (ESI)

In ESI, ionization of analyte occurs in solution (not on a surface) and converted into the gaseous phase by a fine spray from a high-voltage needle. This process generates multiple consecutive ions as a consequence of multiple charging of the analyte.

2.7.1.3 Surface-Enhanced Laser Desorption/Ionization (SELDI)-MS

This MS method has been used extensively in body fluid proteomic studies due to the following advantages:

- Minimal requirement for sample purification before MS analysis.
- High sensitivity to low molecular weight (<15 kDa) proteins, which are important body fluid biomarkers.

Introduced in the early 1990s by Hutchens and Yip, and commercialized in 1997 by Ciphergen Biosystems, the SELDI technique is a modified version of MALDI [6]. The SELDI-TOF MS (ProteinChip system) is a high-throughput technique for identifying protein signatures. A complex protein mixture is spotted on a surface either through biochemical (e.g., high-affinity protein-protein interaction) or chemical (e.g., charge or hydrophobicity) modifications. Some proteins bind to the surface while others are lost during the washing process. The matrix is then applied to crystallize with the sample peptides on the surface, following which a protein chip reader generates mass profiles of each ion. SELDI-TOF MS can produce specific disease biomarkers or biomarker panels for validation studies.

2.7.1.4 Multidimensional Protein Identification Technology (MudPIT)

This is a "shotgun" non-gel-based approach to proteomic studies. Here, sequence-specific protein digestion with trypsin and endoproteinase Lys-C is initially performed. The generated peptides are then separated by strong cation exchange (SCX) and reverse-phase HPLC. The peptides are then subjected to MS analysis. Database search using MS data then enables protein identification.

2.7.2 Labeling Techniques

2.7.2.1 Isotope-Coded Affinity Tag (ICAT)

This is a gel-free MS-based technique that enables the comparison of up to two samples. It involves chemical isotope labeling of proteins for quantitative MS analysis. Proteins are labeled with either isotope-coded light (e.g., for cancer) or

heavy (e.g., for controls) ICAT reagents on thiol groups in cysteine residues. The ICAT reagent has three distinct regions with defined functions. It has the thiol-reactive group to label cysteines, an isotope linker region that has either heavy or light tag, and a biotin moiety used for affinity purification.

2.7.2.2 Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)

This method is not capable of measuring secreted protein signatures in body fluids, because it involves in vivo protein labeling. But it can be used to discover differentially expressed proteins between cancer and normal cells (e.g., in secretome experiments), which can then be applied to body fluids using other suitable proteomic methods. SILAC enables quantitative MS-based protein analysis of two to three different cell types. The SILAC method involves an in vivo metabolic incorporation of stable isotope-containing amino acids that are added to the cell culture medium. Three stable isotopes, ^{13}C, ^{15}N, and deuterium, enable the labeling of up to three samples. After incorporation, proteins are digested and the isotope-labeled peptides are determined and quantified using MS as differing in mass, and relative abundance is determined by peak intensities or areas.

2.7.2.3 Isobaric Tag for Relative and Absolute Quantitation (iTRAQ®)

iTRAQ® uses multiplex stable isotope reagents that enable labeling of all peptides in four or eight different samples. This amide-specific labeling (labels amine residues on proteins) reagent allows unbiased simultaneous identification and quantification of proteins without interfering with posttranslational modifications. The iTRAQ tags have identical mass (isobaric) but differ chemically. The N-terminal amine and lysine side chain amines are labeled following protein digestion. This labeling method enables the production of differential protein fragment peaks on tandem MS.

2.8 Protein Microarrays

Protein microarrays have commonly been used in body fluid proteomic studies for biomarker discovery, validation, and clinical applications, simply because of their sensitivity and reliability to detect low-abundant proteins. Unlike gel-based and MS-based proteomics that can detect and quantify both known and unknown (novel) proteins and peptides, protein microarrays detect only known proteins selected for by the array manufacturer. Proteins expressed by comprehensive expressed clones can be isolated, purified, and used for microarray fabrication.

Consistent with other microarrays, protein microarrays are composed of microspots that contain captured immobilized protein molecules in rows and columns on a solid surface. These surfaces are spotted with clinical samples such as body fluids, cell lysates, or cells. The spots containing the experimental samples are read using labeled (e.g., chemiluminescence, fluorescence, or radioactive) and label-free (e.g., backscattering interferometry, Brewster angle straddle interferometry, UV fluorometry, or SELDI-TOF MS) technologies. The immobilized proteins can either be antibodies (referred to as forward-phase protein microarrays) or antigens (known as reverse-phase protein microarrays). Thus, forward-phase protein microarrays will capture antigens, while reverse-phase protein microarrays will target antibodies in the test clinical samples.

Protein microarrays can be used for protein identification and quantification for biomarker discovery and utility. Such platforms are referred to as analytical protein detection microarrays. These are commonly antibody microarrays. Protein microarrays are also increasingly being used for biochemical activity or functional studies, including but not limited to protein-protein, protein-nucleic acid, and protein-small molecule interactions, posttranslational modifications, and kinase activity. These platforms are referred to as functional protein microarrays, and they contain functional whole protein molecules or domains. Protein microarrays offer sensitive and high-throughput analyses of proteins and peptides in various body fluids and have been extensively used in these applications.

2.9 Metabolomic Technologies

In addition to some of the proteomic technologies used in metabolomic studies, there are unique technologies (such as nuclear magnetic resonance (NMR) spectroscopy) applied to the study of metabolites.

Metabolomic studies require sample preparation (not required for NMR), such as separation of some analyte from others in order to simplify complex mixtures prior to detection. Separation is achieved by gas chromatography coupled with MS, capillary electrophoresis, or HPLC. Ionization is required to generate the ions needed for detection. Commonly applied are chemical, electrospray, and electron ionization methods. There are various detection methods; however, MS and NMR spectroscopy are commonly used for metabolite detection.

2.9.1 NMR Spectroscopy

This is an analytical chemistry method of determining the structure of organic compounds by exploiting the magnetic properties of their atomic nuclei. The physical and chemical properties of atoms in molecules, as they spin, can change the intramolecular magnetic field around the atoms, which can alter the resonance

frequency and hence the electronic structure of the molecule. Unlike MS, NMR spectroscopy does not require analyte separation, and hence samples can be recovered for reanalyses. The simplicity of sample preparation coupled with the high analytical reproducibility, rapidity of performance (few minutes per sample), and low per-sample costs are some of the strengths of NMR spectroscopy detection method. However, in comparison to MS-based detection method, NMR spectroscopy is less sensitive. Additionally, NMR requires large amounts of samples, thus making it more of a discovery than an application tool.

While any spin-active nuclei can be detected by NMR spectroscopy, those commonly targeted are ^{1}H and ^{13}C, because of their abundance in the body and unique features. ^{1}H is very abundant in the body (99.98 %) and hence offers increased sensitivity, but it has a much smaller chemical shift range compared to ^{13}C. Thus, ^{1}H tends to have more peak overlaps (low resolution), and hence an analysis of output data is more difficult. ^{13}C, on the other hand, has a lower sensitivity (less abundant) but has a high chemical shift range that is 20 times higher than ^{1}H and hence produces much better resolved spectral peaks (greater spectral resolution than ^{1}H).

2.9.2 Commercial Plasma Metabolomic Kits

The AbsoluteIDQ™ from Biocrates Life Sciences is a kit for quantitative targeted plasma metabolomics. This kit enables the absolute quantification of ~160 metabolites in >4 compound classes in just a few minutes per sample. It can be used with standard triple quadrupole MS, but has been optimized for Applied Biosystems API 4000™ and 4000 Q TRAP® LC-MS/MS system. The kit includes the MetIQTM software package that enables complete analytical process from project setup to calculation of metabolite concentration. It supports in-house proprietary data storage and can be used for a number of applications including drug efficacy, toxicity, and biomarker research.

2.10 Lab-on-a-Chip

Developed in the mid 1970s by a team at Stanford University, lab-on-a-chip (LOC) became popular in the early 1990s, fueled mainly by the development of biomedical engineering technologies such as microfluidics, photolithography, and sensor arrays. Lab-on-a-chip is an integration of several conventional laboratory procedures on a miniaturized or microchip device that provides controlled settings for detection and measurement of biomolecules (Fig. 2.3). Micro total analysis system (μTAS) is a version of LOC, but is used more for integration of miniaturized analytical chemistry assays. Microelectrochemical system (MEMS) is another version of this technology but uses mechanical and electrical elements on the

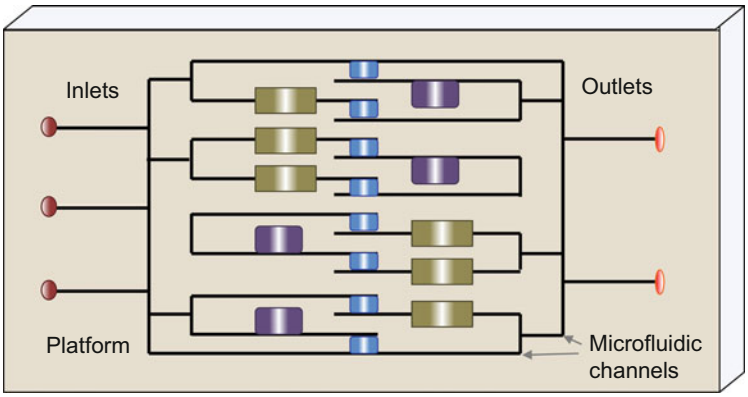

Fig. 2.3 The fundamental design of lab-on-a-chip platform. There are various LOC designs for different applications. In its basic form, a LOC must have inlets for sample and reagent intake and outlets for post-processing waste discharge. The bulk of the device is customized for specific intended uses in regard to sample handling, processing, and detection technologies

chip. There are diverse variations of these technologies. However, the development of these devices requires cutting-edge medical engineering. They are manufactured at the microscale or nanoscale level to perform several functions including sample intake, preprocessing, analysis, detection, or quantification. They are very sensitive and hence adept at handling very small volumes of samples including body fluids. In fact, samples can even be at the picoliter level to be accurately analyzed. An important advantage is their ability to measure numerous and different biomolecules in blood and other body fluids. The manufacturing of these devices includes microfabrication using glasses, metals, silicon, plastics, and ceramics. Microfluidic engineering has played an important role in advancing the LOC technology, because of the following features:

- Permits sample flow through the device by various pumping mechanisms.
- Permits sample mixing by micro-vortices.
- Enables sample flow through capillary channels that can be coated with specific antibodies or molecules targeting specific biomarkers.
- Has flow control valves to modulate sample flow.
- Has heating and cooling systems for various reaction conditions.
- Reactions occur in micro-bioreactors.

Indeed, virtually all sample processing and handling conducted in the conventional laboratory are recapped in this miniaturized device. Advances in microfluidics/nanofluidics, photolithography, computer chip designs, nanotechnology, and bioinformatics over the past few decades have driven LOC features to even much smaller miniaturized scale. These include nanoscale level bioreactors, nanopores for electrophoresis and sequencing, nanoscale cell arrays for cell culture, membrane ion channel studies, protein studies, and toxicity testing among many others. These cell arrays (for cell culture) incorporate microfluidic systems that

function just like capillaries in the body in bringing nutrient supply and removal of waste products.

Several LOC prototypes are being developed for various applications such as routine blood and HIV testing. Indeed, an international team of researchers used latest cutting-edge advances in plasmonics, nanofabrication, microfluidics, and surface chemistry to develop a LOC device to detect very low concentrations of cancer protein biomarkers in blood [7]. This platform is reliable, sensitive, small, and potentially low cost and yet relies on just a drop of blood. Its potential applications are enormous. Additionally, Chuang et al. [8] developed a LOC device using multiplex immunosensors for detection of galectin 1 and lactate dehydrogenase B, which are associated with different grades of bladder cancer. This device could effectively identify different grades of bladder cancer cells. In general, the LOC concept has several advantages, including:

- Fast sample analysis that can occur within minutes.
- Sensitive quantification of targets.
- Automation.
- Small amounts of sample requirements – small droplet of body fluids suffices.
- Low-cost devices.
- Portable and easy to use at the POC.
- No need for specialized technicians or costly laboratory equipment.
- Easy availability in low resource countries and communities.

2.11 Summary

- High-throughput technologies have been developed and advanced for cancer biomarker discovery, validation and product development.
- Although next-generation sequencing is costly at the moment, it enables a depth of genome interrogation that has never been possible before. Depending on the genomic level and depth of coverage, low-level genomic alterations can be uncovered by this technology.
- Because of tumor heterogeneity and evolution with acquisition of novel mutations, single-cell genomics provides great insight into tumor biology, which is important to cancer management.
- With current levels of analytical detection technologies, single-cell genomics will be fruitless without methods for whole genome and transcriptome amplifications.
- At the cancer research front are also edgy proteomic and metabolomic technologies for biomarker discovery.
- The ability to miniaturize the entire conventional laboratory processes on a chip (lab-on-a-chip, LOC) is advancing the utility of cancer biomarkers. Developed and deployed at the point of care, LOC technologies should help reduce the cancer burden in resource-poor communities.

References

1. Thun MJ, DeLancey JO, Center MM, et al. The global burden of cancer: priorities for prevention. Carcinogenesis. 2010;31:100–10.
2. Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6:377–82.
3. Ramskold D, Luo S, Wang YC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30:777–82.
4. Heitzer E, Auer M, Gasch C, et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res. 2013;73:2965–75.
5. Ni X, Zhuo M, Su Z, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci USA. 2013;110:21083–8.
6. Hutchens TW, Nelson RW, Yip TT. Evaluation of peptide/metal ion interactions by UV laser desorption time-of-flight mass spectrometry. J Mol Recognit. 1991;4:151–3.
7. Acimovic SS, Ortega MA, Sanz V, et al. LSPR chip for parallel, rapid, and sensitive detection of cancer markers in serum. Nano Lett. 2014;14:2636–41.
8. Chuang CH, Wu TF, Chen CH, et al. Lab on a chip for multiplexed immunoassays to detect bladder cancer using multifunctional dielectrophoretic manipulations. Lab Chip. 2015;15:3056–64.

Chapter 3
Liquid Biopsy and Cancer Biomarkers

Key Topics

- The liquid biopsy concept
- Cancer biomarkers
- Classification of cancer biomarkers
- Biomarker developmental processes and challenges
- Body fluids, types, and handling
- Clinical application of circulating biomarkers
- Point-of-care cancer management

Key Points

- Concerned with harnessing cancer biomarkers in body fluids, the liquid biopsy concept is a paradigm shift in cancer research and product development. It has important applications in clinical oncology.
- In spite of the technical challenges in assaying cancer biomarkers in body fluids, circulating biomarkers offer several advantages over conventional tissue biopsy or aspiration cytology. Liquid biopsy is a noninvasive procedure adept to multiple longitudinal sampling even in frail patients. It overcomes tumor heterogeneity and enables real-time monitoring of tumor genomic evolution, especially with treatment.
- Advanced miniaturized technologies enable the development of liquid biopsy biomarkers on lab-on-a chip devices that can be deployed at the point-of-care to help bridge the unacceptable disparities in cancer statistics between the resource-poor and resource-rich parts of the world.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_3

3.1 Introduction

Established is the fact that cancer is a complex and heterogeneous disease. Although a particular cancer may be of clonal origin, different selection pressures (akin to Darwinian evolution) drive some clones to acquire additional, possibly private mutations, as well as other molecular alterations. Of particular interest in cancer biology is the presence of a population of rare cancer cells, and probably the "Achilles heel" that needs to be thoroughly understood in order to achieve effective treatment and cure of cancer. Not only are these rare cancer stem cells difficult to sample, but they also tend to be resistant to therapeutic interventions including radiotherapy.

The heterogeneous nature of cancer implies that the current practice of using tissue biopsy for treatment decision-making or disease monitoring has obvious shortcomings. Moreover, in frail patients, of which many cancer patients are, it may not be possible to perform invasive procedures to obtain such biopsy samples. Cancer evolves over time, whether treated or not, which means that for efficient treatment decision-making, the molecular signature of the tumor should be serially assessed at specific points in time. Apart from not being representative, it may be impractical to subject some patients to serial biopsy procedures.

With refinements in technology, it is now well established that the various heterogeneous cells within a tumor constantly release into body fluids, myriads of biomolecules (DNA, RNA, proteins, peptides, and various metabolites) that are free, protein or lipid bound, or in extracellular vesicles. Biomarkers in body fluids (circulating or localized tissue-derived) are of tremendous potential in translational oncology. Thus, not only will the "liquid biopsy" concept overcome tumor heterogeneity in biomarker analyses, but it also resolves many of the other issues associated with obtaining tissue biopsy. Additionally, during the metastatic cascade, circulating tumor cells (CTCs) and circulating cancer stem cells (CCSCs) undergo genetic evolution, acquiring novel mutations that are absent in the primary tumor. Therefore serially sampling body fluids noninvasively should enable adjusting treatments to target these cells, which are responsible for metastatic deposit formation associated with >90 % of all cancer mortalities.

Biomarkers, whether assayed in the primary tumor or in body fluids, are potentially powerful tools in clinical oncology. They are useful in several ways to help curtail the burden of cancer. They may be used for population screening to guide primary prevention (before the onset of disease) or secondary prevention (when disease is detected early and can be cured by surgery or other preventive measures). Biomarkers can be specific indicators of the presence of disease (aid in diagnostics) and can guide treatment decision-making simultaneously, in the realm of companion diagnostics. Although there are ~200 types of cancer, there are several subtypes, each with its unique biologic behavior and response to treatment. Biomarkers are emerging as accurate tools for classification of various cancers into subtypes, which is important in disease management. The course or behavior of a particular cancer can be predicted using biomarkers (prognostication), and they can provide

directions or insights into patient selection for therapeutic interventions. During treatment, biomarkers modulate with disease remission or progression and hence can be assayed to monitor treatment effectiveness, resistance, or even toxicity. To this end, they are used in the scientific disciplines of pharmacogenetics, pharmacogenomics, and pharmacometabolomics. Early disease relapse also modulates with biomarker levels, and hence they can serve as targets to monitor for early detection of recurrence, when salvage interventions may improve survival.

Notwithstanding the importance of cancer biomarkers, the benefits cannot be reaped without advances in technological developments. Thus, over the past few decades, sensitive analytical technologies have been developed to accurately measure biomarkers, including even the low abundant targets in body fluids. An additional armamentarium in optimal biomarker translation is the development of miniaturized devices that can be deployed at the point-of-care. With these sensitive technologies and devices, it is now even possible to perform analyses of the entire biomolecular composition of single cells, e.g., whole genome sequencing and transcriptomics. This feat is important because it enables the Darwinian evolution of cancer cells to be studied accurately and in more detail. Additionally, it has paved the way for whole genomes and transcriptomes of CTCs and CCSCs to be examined for novel alterations that can be used to guide patient management.

3.2 Liquid Biopsy

This book concerns "liquid biopsy," a phrase originally coined to describe CTC capture as an alternative sample to the invasive traditional tissue biopsy often used for biomarker analyses. While cancer management largely depends on biopsy sampling of primary and/or metastatic tumor tissues for pathologic and/or molecular analyses, liquid biopsy enables minimal invasive approaches to sampling these cancer cells and their molecular imprints (Fig. 3.1). Moreover, liquid biopsy is a representative sample of the entire tumor, which overcomes tumor heterogeneity that is seldom captured by tissue biopsy. While surgeons may have easy access to primary and/or metastatic cancer tissue and hence may potentially achieve complete curable resection, CTC capture is not at the moment a procedure for eradication of cancer cells. Yet CTCs are responsible for distant metastasis that account for the majority of cancer-related deaths. Thus, chemotherapy, immunotherapy, and targeted therapies are aimed at killing these metastasizing cells. However, decisions as to what treatment modality or strategy to use is often made based on molecular and genetic information obtained from the primary or metastatic tumor tissue or aspirates. Well established though is the concept of tumor evolution, whereby CTCs acquire novel genetic changes that enable metastasis, resistance to treatment, and avoidance of immune surveillance. Liquid biopsy thus holds tremendous potential for clinical oncology and hence has attracted the much-needed attention.

Although initially coined to designate CTC capture, liquid biopsy encompasses all cancer biomarkers in circulation and in other body fluids. Body fluid analysis is

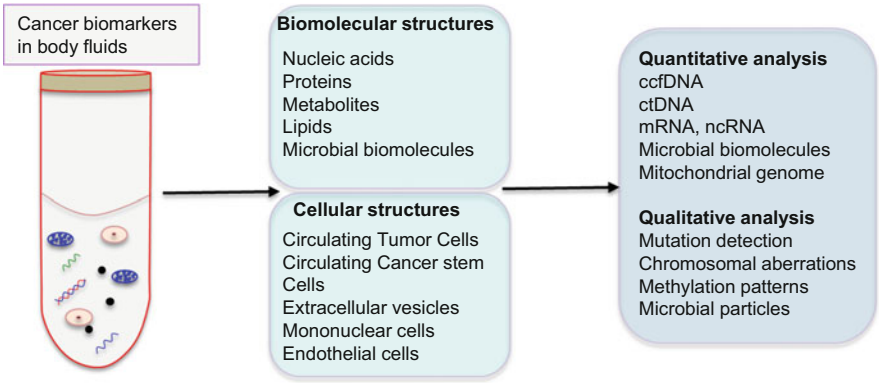

Fig. 3.1 The liquid biopsy concept of biomarker analysis. A body fluid is a rich source of diverse biomarkers that can be detected and measured qualitatively and quantitatively. *CcfDNA* circulating cell-free DNA, *ctDNA* circulating tumor DNA, *ncRNA* noncoding RNA

simple and yet enriched with various cancer biomarkers. The longitudinal analysis of the epigenome, genome, transcriptome, proteome, metabolome and glycome in cell-free extracts, CTCs, or cancer-derived extracellular vesicles offers immense advantage to the discipline of cancer research as well as valid biomarkers (Fig. 3.2).

The presence of "cell-free" cancer biomarkers in circulation and other body fluids is very well established. These cancer-derived biomarkers (e.g., circulating tumor DNA and microRNA) carry signatures reminiscent of the primary and metastatic tumor and hence can be assayed as part of the liquid biopsy phenomenon for patient management.

Also extensively investigated and established is the diagnostic importance of cancer-derived extracellular microvesicles or microparticles. Not only are their numbers elevated in circulation of cancer patients, but also some of these microvesicles (e.g., exosomes) are selectively packaged with specific cancer signatures. Thus, the analyses of cancer-derived extracellular microvesicles in circulation and other body fluids provide tremendous opportunity for the study of cancer biology and the discovery, validation, and eventual clinical utility of cancer biomarkers in the diagnostics, management, and therapeutic targeting of cancer.

Circulating tumor cell enumeration has established clinical relevance in many solid tumors. Additionally, CTC analyses provide further insights into the genomic imprint of cancer cells, and some information into the mechanisms of the metastatic cascade and cancer cell dormancy. Thus, emerging technologies for CTC capture and characterization, including the application of deep sequencing technology on single-cell genomes, should enrich our knowledge of the cancer cell molecular landscape.

In summary, there are ample biomarkers of diseases including cancer in the circulation and other body fluids amenable to discovery, validation, and standardization for clinical translation. All these biomarkers can be subjected to various "omics" analyses. The issue of slow progress in biomarker translation is very well

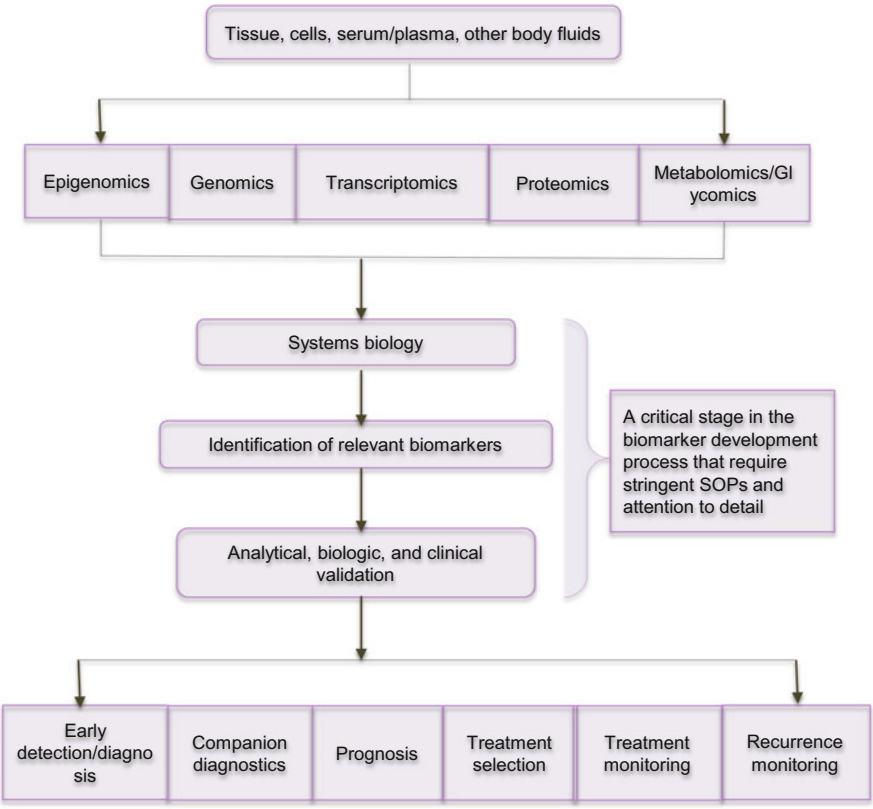

Fig. 3.2 The optimal cancer biomarker development paradigm. The same biologic sample being subjected to all aspects of analysis should enable logical data integration

recognized. Of woeful note is the dissociation between the number of discovered cancer biomarkers and those validated for translation. The paucity of validated biomarkers in the clinic is partly due to poor study designs, the absence of standardization, and hence the irreproducibility of many discovery data. On the bright side, many industry, academia, and regulatory bodies are making it a reality to translate cancer biomarkers into routine clinical practice. The future illuminates.

3.3 Biomarkers

A number of different definitions have been given to the word "biomarker" or biological marker, depending on the intended purpose. First, in connection with chemical safety, the International Programme on Chemical Safety (led and coordinated by the WHO, the UN, and the International Labor Organization) defined a biomarker as "any substance, structure, or process that can be measured in the body

or its products and influence or predict the incidence or outcome of disease." An extended version includes the effects of treatments, interventions, and unintended environmental exposures. Second, concerning environmental risk assessment, the WHO defines a biomarker as "almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological. The measured response may be functional and physiological, biochemical at the cellular level, or a molecular interaction." Third, the NIH Biomarkers Definitions Working Group defines a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." From all these definitions, the following features characterize biomarkers:

- A biological system is involved.
- A characteristic or entity indicative of physical, chemical, and biologic process.
- Objectivity in measurement and evaluation of the characteristic(s) or entity.
- A response by the body to an agent.
- The agent can be a pharmacologic or any chemical intentionally or unintentionally introduced into the body.

The use of the word "characteristic or entity" is all encompassing, suggesting any change in anatomy, physiology, biochemistry, pathology, or pharmacology of an individual constitutes a biomarker. Thus, biomarkers can be changes in physical signs such as blood pressure, pulse rate, body temperature, etc. or changes in traditional biochemical measurements such as cholesterol and blood glucose and radiographic changes. The term biomarker used in this book focuses on cellular, genetic, and molecular changes in cancer. In the field of cancer research and oncology, the NIH Biomarkers Definitions Working Group definition should be adopted as the standard to avoid confusion in use of terminology.

3.4 Cancer Biomarkers

Extending from the definition of a biomarker, cancer biomarkers should indicate cancer presence and modulate with cancer biology and various therapeutic interventions. These biomolecular compositions could be single or few alterations such as specific epigenetic changes, cancer-causing mutations that alter genes, gene products (e.g., proteins and their modified forms), or molecular compositions of cancer-causing organisms. They could also be composite signatures obtained from genome-wide ("omics") interrogations such as the use of microarrays and other technologies for SNP profiling, comparative genomic hybridization, gene expression, and proteomic and metabolomic profiling. Cancer biomarkers can be used for familial risk assessment, screening of target populations for primary (before disease occurrence) or secondary (early cancer detection) prevention, differential diagnosis, prognosis, selection of patients for specific or targeted therapy, monitoring for

therapy response, and the early detection of cancer recurrence. Thus, cancer biomarkers can be measurable changes in any of these cellular structures and biomolecules:

- Cellular morphologic variations (the anatomy of the cancer cell) or changes in cell number.
- Cellular fragments in circulation or body fluids (e.g., exosomes, microvesicles, microparticles, or apoptotic bodies).
- Nonstructural changes in DNA, such as histone and CpG alterations.
- Structural changes in DNA such as mutations, gene rearrangements, and chromosomal alterations.
- Changes in transcripts of structural or coding genes (expressional changes).
- Changes in the expression of noncoding transcripts (short and long noncoding RNAs).
- Differential protein expression and circulating proteins and peptides.
- Changes in metabolites or metabolomic fingerprints and profiles.
- Changes in lipid and glycan profiles.
- Presence and/or changes in the biomolecular compositions of cancer-causing organisms (e.g., Epstein-Barr virus).

3.4.1 Features of an Ideal Cancer Biomarker

There are various clinical utilities of biomarkers. However, the ideal biomarker should accurately identify or predict what it is intended for, with negligible false proclamations. Thus, to optimize the chances of developing a clinically useful biomarker, the "intended clinical utility" should be determined a priori to biomarker exploration and development. This initial focus informs the study design in regards to patient and control population characteristics, types of samples required, and importantly the statistical performance needed to make this a valid biomarker, whereby the benefits far outweigh the harms. As an example, for a biomarker that when positive triggers an invasive procedure for definite diagnosis, in addition to the optimal screening accuracies based on disease incidence, prevalence, and other measures, it must have a very high negative predictive value, such that almost all patients who truly are devoid of the disease test negative and be spared the potential harm of further investigation. Ideally therefore, a valid cancer screening biomarker should have the following minimum characteristics:

- Be an early detection biomarker that is present at the very early stages of disease, to enable early detection of potentially aggressive cancer when cure is optimal.
- Have the ability to accurately (high sensitivity and specificity) differentiate people with cancer from those without.
- Be easily measurable in noninvasive manner using body fluids.
- Be easy to perform, especially at the point-of-care.

- Be cost-effective, enabling all targeted population to be able to afford it, irrespective of their economic status.
- Have companion diagnostic capability, enabling a diagnostic that can be linked to therapy decision-making.
- Have positive measurable clinical impact on morbidity and mortality.

Because biomarkers are used for various clinical purposes, under certain circumstances, other measures of accuracy (e.g., high NPV or PPV) may be more important than just sensitivity and specificity.

3.4.2 Classification of Cancer Biomarkers

Consistent with the broad definition of biomarkers, various diverse schemes have been used to classify them. For instance, the FDA's suggested classification focused on their use in drug development. Hence, the designated four classes include exploration, demonstration, characterization, and surrogacy. Only surrogate biomarkers are considered valid for use in drug approval processes. Frank and Hargreaves, however, divide all biomarkers into three main categories (types 0, 1, and 2) [1]. Type 0 biomarkers are associated with the natural history of the disease and also modulate with the clinical course of the disease. These include diagnostic and predictive biomarkers. Type 1 biomarkers are more in the realm of pharmaceutical development. These biomarkers determine the biologic effects of therapeutic interventions including response, nonresponse, and toxicity. Finally, type 2 biomarkers are surrogate measures of the disease. Usually they modulate with disease course during interventions and are therefore used as substitutes for the clinical outcome of the disease, especially during clinical trials.

Cancer biomarkers can be classified based on (1) molecular or cellular alterations in cancer cells or (2), as mostly referred to in the literature, based on their clinical and research applications.

3.4.2.1 Molecular Classification of Cancer Biomarkers

Considering cancer biomarkers from a molecular perspective, they can be:

- Epigenetic alterations, specifically aberrant gene methylation patterns, histone modifications, and miRNA expressional changes.
- Changes in nuclear DNA; this encompasses chromosomal abnormalities, SNPs, haplotypes, unusual DNA deletions and rearrangements, specific mutations, DNA content alterations, and ALU/SINE and LINE sequence variations.
- Mitochondrial alterations at the structure of the organelle, genome, transcriptome, proteome, and metabolome.
- Nuclear RNA alterations; both coding and noncoding gene expressional changes.

- Proteins and their modifications; these include the cancer cell peptidome, proteome, and posttranslational alterations.
- Metabolites altered as a consequence of cancer development.
- Abnormal glycosylation of macromolecules; proteins, glycans, and lipids.
- Cellular elements; these include cancer-derived extracellular vesicles and apoptotic bodies.
- Cancer cells, cancer stem cells, and other cells such as immune and endothelial cells assayed in the field of cytomics.

Traditionally, the various different molecular targets have been explored singly, or at the omics level, targeting a specific molecular class for biomarker exploration. Thus, single biomolecular alterations or set of genomic, transcriptomic, proteomic, or metabolomic signatures assayed vertically have served as good biomarkers for cancer. The horizontal integration of such data from a single cell will be very informative. Indeed, the ability to study the interactions of genes, proteins, and metabolites in a particular cell (the interactome) should expand our knowledge on the interconnected communication networks that produce a particular cancer cell phenotype. Moreover, the study of the composite phenotypic traits of a cell (the phenome), especially when correlated with genomic, proteomic, or metabolomic profiles, should offer further insights into the discovery of actionable cancer biomarkers.

3.4.2.2 Classification Based on Clinical and Research Utility

Cancer biomarkers are adopted for use in different clinical situations, each of which requires specific biomarker characteristics and performances. On a classification scheme, these cancer biomarkers can have the following uses:

- *Risk assessment*: Biomarkers that change early in cancer development, especially those associated with field cancerization, can serve as alarm sensors of impending cancer development, for instance, *KRAS* mutations in colorectal cancer or *BRCA* mutations in women and breast and/or ovarian cancer. Sensitive technologies that enable detection of these alterations in body fluids should enable early detection of cancer risk, especially when used to screen high-risk individuals. Such findings will enable enhanced surveillance or institution of preventive measures such as lifestyle changes, epigenetic reversal medications, or surgery.
- *Early detection and diagnosis*: Screening or early detection biomarkers provide evidence of the likelihood of disease in asymptomatic population at risk for a particular disease. Positive outcomes usually trigger further noninvasive diagnostic investigations. Diagnostic cancer biomarkers are those that complement other clinical findings and investigational procedures in the establishment of disease presence in patients who are symptomatic. The early detection biomarkers are offered in screening programs, and those who test positive are then referred for diagnostic workout that usually includes invasive procedures.

- *Companion diagnostics:* In advanced-stage disease, patients are often provided with systemic therapies. Companion diagnostic biomarkers are useful in this setting. A companion diagnostic biomarker should provide both diagnostic and treatment predictive information on safety and effectiveness for the specific cancer and patient. In addition to its use to determine responders from adverse responders, the FDA includes the use of companion diagnostics in a longitudinal manner for detection of treatment changes and hence for modifications to be made accordingly.
- *Prognosis*: True prognostic biomarkers are those that are able to predict at the time of diagnosis before any treatment intervention as to which cancers are most likely to be indolent and which ones will be aggressive, metastasize, and be lethal. This obviously has tremendous advantages of enabling patient triaging for aggressive management of the bad tumors, and sparing those with indolent tumors the toxicities and harms associated with therapy and interventional procedures. Patients with indolent disease can be placed on active surveillance protocols.
- *Therapy prediction and selection*: Commonly referred to as "predictive biomarkers," these biomarkers predict the likelihood of a cancer patient responding to a particular therapeutic agent. Additionally, they also detect those who will not respond and even those in whom the agent may be toxic.
- *Therapy response or toxicity monitoring*: These are biomarkers that alter with various therapeutic interventions. These biomarkers, especially when assayed longitudinally in body fluids, inform oncologists about which patients are responding and which patients are not or getting worse with treatment. Importantly, even early responders may become nonresponders due to tumor evolution, enabling changes in treatment plans to be made at that point in time.
- *Recurrence monitoring:* The need to detect cancer recurrence early is important to optimal management. Biomarkers are emerging as being able to detect cancer recurrence earlier and with better performances than traditional imaging or even by the Response Evaluation Criteria in Solid Tumors (RECIST). While new biomarkers may be better at detecting only recurrences, in many instances, the diagnostic biomarkers of a particular cancer could serve as early indication of relapse, as their levels change. Such biomarker levels usually reset to within normal range with treatment, indicative of response. The return to disease state (could be elevated circulating biomarkers) is usually indicative of disease relapse or nonresponse.
- *Pharmacokinetics and pharmacodynamics of medications*: These are pharmacogenetic or pharmacogenomic biomarkers that identify an individual's genetic variation that controls the pharmacokinetics and pharmacodynamics of drugs. These assays often target gene variations including polymorphisms in drug uptake and metabolizing pathways including enzymes, receptors, and transport proteins. Specifically, these proteins control how drugs get absorbed and distributed and function in the body. They could also be whole genome variations including haplotype markers, SNP maps, and global gene expressions

that identify an individuals' response to a therapeutic agent (efficacy and toxicity).

- *Surrogate endpoints:* Often used in drug development processes, these biomarkers modulate with specific therapeutic agent, indicating whether the treatment is effective or not. Validated surrogate endpoint biomarkers are very cost-effective compared to the traditional means of measuring drug effectiveness, such as imaging or obtaining tumor tissue biopsy. Additionally, drugs that fail to alter surrogate endpoint biomarkers (i.e., ineffective medications) can be eliminated from the drug development process to save time and money.
- *Drug target development:* These biomarkers serve useful purposes in drug development as they inform the development of a specific agent. These biotherapeutic agents are developed to target and alter specific molecular or signaling pathway to effectively treat the disease.

3.4.3 Cancer Biomarker Development

There are specific developmental phases that a valid biomarker must successfully go through to be accepted for product development. Specifically, these include phase I (discovery), phase II (validation), phase III (qualification), and phase IV (standardization). All these phases require a priori planning, including statistical requirements and the development and utility of standard operating procedures.

- *Phase I or discovery phase*: This involves comparison of diseased and healthy or benign disease tissues of the same organ. The use of tissue at this stage is justified because this will contain enough biomolecules that will enable the ideal multidimensional interrogation of the epigenome, genome, transcriptome, proteome, and metabolome/glycome of the same tissue (Fig. 3.2). The wealth of data accrued at this stage can be integrated into a systems' biologic approach or other bioinformatics algorithms to enable discovery of authentic biomarkers. Biomarkers that elevate with disease and are secreted or otherwise released into the circulation and other body fluids will be very ideal for validation, because they will enable development of cheap and noninvasive clinical assays for translation.
- *Phase II or validation phase*: This phase involves analytical and biologic validation. Analytical validation involves assay performance at measuring the biomarkers of interest accurately and reproducibly. Also assay precision and range of detection are determined. Biologic validation relates the biomarker to biologic mechanisms of the cancer, and not just a mere association without causation. Knowing the biologic importance of the biomarker can also inform targeted drug development.
- *Phase III or qualification phase*: This is clinical validation, whereby the clinical utility of the biomarker in terms of disease detection, assessment, or therapy

prediction is accurately determined. This requires various clinical trials focused on the intended use of the biomarker.

- *Phase IV or standardization phase*: Following acceptance of the biomarkers into clinical practice, there is a need for assay standardization across laboratories. The use of, for instance, private bioinformatics algorithm will thwart interlaboratory comparison. Thus, the College of American Pathologists has recommended guidelines for the use of non-FDA-approved laboratory-developed test. In clinical translation, it is even more commendable to use a single platform for a specific assay.

The Early Detection Research Network (EDRN) has specific guidelines for the development of biomarkers for early cancer detection [2]. Five phases of biomarker development are identified and are summarized herein:

- *Phase I* (*Preclinical exploratory*): This is akin to the discovery phase where promising biomarkers are determined and prioritize in the case of large-scale omics data.
- *Phase II* (*Clinical assay development and validation*): Here a clinical validated assay is developed that detects the intended cancer. Emphasis is placed on the use of noninvasive samples for assay development and to relate this data to the phase I data. Biomarker performance has to be related to other confounding variables such as age, sex, lifestyle, etc. Also biomarker utility in determining disease stage, grade, and outcome is established.
- *Phase III* (*Retrospective longitudinal study*): Primarily, this stage is to establish lead-time performance of the biomarker in cancer detection. Thus, samples collected from patients before the diagnosis of cancer (duration before cancer diagnosis is important) are compared to those who did not develop cancer. This enables establishment of how early the biomarker detects cancer and also allows determination of screening interval and positive outcomes to be implemented in a phase IV prospective study.
- *Phase IV* (*Prospective screening study*): This phase establishes the clinical performance of the biomarker such as the true positive rate. How many test-positive individuals get diagnosed with cancer? It is also important to determine the false positive rate, which is how many people positive for the test actually do not have cancer. The stage, grade, and other pathological findings of early-diagnosed tumors should also be determined and possibly relate this to prognosis as well.
- *Phase V* (*Cancer control studies*): This phase is to determine the usefulness of the screening test in reducing cancer-related mortalities in the intended population. A biomarker that detects indolent cancers, which will otherwise be harmless, and yet having these patients subjected to treatment, will not be a useful screening biomarker.

3.4.4 Challenges to Cancer Biomarker Development

The biomarker development space is fraught with numerous challenges that explain the paucity of translated biomarkers, despite the plethora of reported biomarkers in the literature. There are technical, biological, analytical, economic, societal, and regulatory issues that need to be ironed out to create a smooth path to biomarker development. This needs the coordinate cooperation of industry, academia, and regulatory bodies such as the FDA. A synopsis of some of these challenges is provided herein:

- A major issue with biomarker development is the lack of reproducibility of discovered biomarkers. Variability in assay conditions, which could arise from types of specimens used, the manner of specimen handling, and the nature of the specific biomarker(s) being targeted could all affect assay performance.
- The inadequacy of available quality biospecimens needed for authentic biomarker discoveries and validation often leads in some cases to the use of convenient samples, which often yield data that fail subsequent phases of development. Indeed, in many academic settings, biomarkers are developed without a prior identification of their intended use. The utility is then defined after the fact, making it difficult to ascertain whether the experimental design and types of specimen used are indeed optimal for the uncovered biomarkers. For example, available samples may be heterogeneous, or enriched for late stage and grade cancers. Biomarkers discovered from such specimens may then be subsequently subjected to validation studies intended for cancer screening, which may therefore fail at this stage.
- Cancer heterogeneity poses a challenge to uncovering biomarkers. Thus, single biomarker assays are often inferior to panel biomarkers for disease detection.
- While of tremendous importance in oncology, the low signal-to-noise ratio often hinders development of accurate biomarkers in body fluids. However, this hurdle is being overcome with the development of sensitive analytical technologies.
- The absence of regulation and standardization creates many suboptimal discovery assays that lack reproducibility.
- There is also a lack of technology standardization and hence the often-observed interlaboratory variability of the same assay. It is necessary to have a common technology for specified biomarker use in the clinics.
- While it may look simplistic by putting the onus on the FDA to define a path for biomarker development, all stakeholders need to cooperate in this effort. Thus, the desired clarity and standardization of the regulatory path for biomarker development, often requested by stakeholders, needs expert discussions.
- Lack of progress in assay development may also stem from stakeholders including academia, industry, government agencies, regulatory agencies, and patient advocacy groups, each of which may have its own interest. Shareholders of a company for instant can thwart the critical path needed for a valid biomarker development because of the need to quickly go to market in order to bring in returns.

- Although critical, industry often finds it unattractive to invest a lot in biomarker discovery and validation. Moreover, at a large-scale clinical validation stage, it may be difficult for industries to secure funds, because of the need to show improved patient outcomes in regard to morbidity and mortality, which are often lacking at this stage.
- Although discovery biomarkers may be authentic, once they are in the public domain without intellectual property protection, they become unattractive to investors.
- Biomarker development for rare cancers or diseases are even more challenging in a number of ways. First, apart from philanthropists, they are often not attractive to investors. Second, it is often challenging to establish statistical significance at initial development and during clinical trials because of the low population of participants.
- Screening tests need to overcome the issues of lead-time bias, overdiagnosis, and overtreatment.
- Screening tests also require optimal accuracy in terms of sensitivity and specificity that are easily obtainable at the discovery phase, often due to type I error, but fail at a large-scale validation phase.
- Screening tests may face challenges in design in regards to optimal time to begin screening, frequency or screening intervals, population to screen, and test performance across different ethnic groups.

These and many other issues, especially concerning tumor biology and regulatory requirements, account for the dismal biomarker product development. Chapter 10 details the issues and efforts by the US Food and Drug Administration (FDA) to help accelerate biomarker and drug development.

3.5 Evidence-Based Medicine

Evidence-based approaches are paramount to improvements in current and future clinical practice. The integration of clinical expertise, patient values and inputs, and best research evidence constitute the evidence-based clinical practice paradigm. An important component of this healthcare delivery system is evidence-based medicine, which is the use of evidence on well-designed and executed randomized studies for decision-making. In this context, research validity is critical to optimized decision-making in patient care. To help ascertain such evidence, the Oxford center for evidenced-based medicine (OCEBM) has provided levels of evidence on reported research, which are applicable to cancer biomarker research work as well. The five levels of evidence from the OCEBM include:

- The highest evidence to be considered for the purpose of clinical decision-making is level 1, which is divided into 3 parts, 1c (good), 1b (better), and 1a (best). Studies with very high sensitivity such that a negative test result rules out the diagnosis (absolute SnNout) and very high specificity such that a positive test

result rules in the diagnosis (absolute SpPin) are evidentiary at level 1c. Validation cohort studies with reference standards that are independent of the test and applied blindly or objectively to all patients (good reference standards) or the use of algorithm or scoring system that leads to a diagnostic category (clinical decision rule) tested within one clinical center forms level 1b evidence. The highest evidence, 1a, is based on systematic review of homogeneous level 1 studies or clinical decision rule with 1b studies from different clinical centers.

- Level 2 is divided into 2b and 2a. Exploratory cohort studies with good reference standards, or clinical decision rule after derivation, or validation only on split samples (same samples split into derivation and validation cohorts) are level 2b evidence. Systematic reviews of homogeneous level 2 diagnostic studies elevate the evidence to 2a.
- Level 3 also comprises of 3b and 3a. Nonconsecutive studies or those without consistently applied reference standards are level 3b evidence. Level 3a evidence includes systematic reviews of homogeneous studies of level 3b or better studies.
- Level 4 are case-control studies with poor or non-independent reference standards.
- At the extreme end is level 5 evidence. These may be considered exploratory studies and consist of bench research and expert opinions based on physiology, without obvious critical appraisal of the literature.

3.6 Body Fluids

The field of biomarkers in body fluid is faced with different terminologies in the literature. Terms often used are body fluids, bodily fluids, biological fluids, or simply biofluids. Irrespective of the terminology, these are liquids produced and secreted or excreted by body cells. They arise as blood filtration, cellular secretion, or excretion. These fluids are produced for normal body functions (e.g., blood and cerebrospinal fluids, CSF), as waste products (e.g., urine), or in pathological conditions (e.g., ascitic or excess nipple aspirate fluids, NAF). Body fluids include saliva, blood, sweat, urine, lymph, and tears; bronchoalveolar (BAF), follicular, amniotic, synovial, peritoneal, ascitic, prostatic, and seminal fluids; as well as vulvar, vaginal, and cervical discharge, among several others. Fluids commonly assayed for systemic conditions include blood, urine, and saliva. Fluids and exfoliated cells for localized diseases include NAF, CSF, stool, gastric juice, BAF, synovial fluids, pancreatic juices, semen, prostatic massage fluid, expressed prostatic secretions, post digital rectal examination urine, menstrual flow, cervical, vaginal and vulvar washes or secretions, skin swabs, etc. Given their origin from cells of the body, these fluids and exfoliated cells serve as important sources of genetic and nongenetic biomarkers of disease processes. These biomarkers are present in secreted or free form or are in exfoliated cells or other membrane-

enclosed vesicles. Such clinical samples have several advantages in oncologic applications including:

- Noninvasive or minimally invasive means of acquisition.
- Rich source of cancer biomarkers.
- Easy storage and transportation.
- Cost-effective utility; in general they do not require specialist for collection.
- Multiple samples can be collected for target identification and validation and for disease and therapy monitoring.
- Compared to biopsy samples, biomarkers in body fluid overcome issues of tumor heterogeneity in the case of multifocal tumors. Target or biomarkers in body fluids reflect entire organ/tissue and thus overcome the heterogeneity of cancer biomarker detection in tissues from biopsy or surgery (such tissues require microdissection of multiple areas for analysis).
- Because of the noninvasive nature and availability, body fluids can be sampled multiple times for detection, confirmation of diagnosis, and monitoring of disease progression and therapy response, even in frail patients.

The main concerns of body fluid analyses are low copy number of targets, biomarker degradation (e.g., by nucleases), and possible lack of marker specificity such as those in circulation. In proteomics, the large dynamic range of proteins is a major technical concern.

Body water is also considered as body fluid in physiological terms. This type of body fluid is contained in three separate but interrelated compartments – intracellular, interstitial, and vascular compartments. This physiologic fluid is tightly controlled by hormonal actions. On average, 50–60 % of total body weight is water. It is into this body water that biomarkers are released.

3.6.1 Historical Perspective on Body Fluids and Cancer

The history of medicine and etiologic mechanism of disease has centered immensely on body fluids. Ancient Chinese doctors, for example, used ants to examine urine for detection of diabetes. Similarly, many ancient cultures examined the color, taste, and smell of urine for diagnosis of various diseases. Consistently, while cancer as a disease had been known for several decades, its etiology had to wait for several more decades. The name cancer was given by Hippocrates (ca. 460BC–ca. 370BC) to describe its external growth with projecting veins reminiscence of a crab or crayfish. Thus, he used the term *carcinos* (Greek for crab) to describe cancer. Because human dissection was unacceptable at the time, Hippocrates postulated that the four humors or body fluids (blood, phlegm, yellow bile, and black bile) controlled health. He indicated that a healthy body needed balance of these four humors and that any lack of balance caused disease. He associated increased levels of black bile with the etiology of cancer. His approach to cancer treatment therefore centered on the “humor theory” being able to restore

balance to the four humors. Bleeding, diet, and use of laxatives were therefore employed to achieve this balance. This humor theory, also held by the Greek anatomist and philosopher, Aelius Galen (129AD–c.200 or c.216), persisted for over 1300 years before autopsy studies enabled scientific enquiry into the causes of diseases. Another misconception at the time was that cancer originated from lymph. Stahl and Hoffman held the view that cancer was made of fermenting and degenerating lymph that varied in acidity and alkalinity. The Scottish surgeon, John Hunter, also agreed that tumors grew from lymph removed from blood.

Johannes Muller, a German pathologist, in 1838 made the observation that cancer developed from cells and not lymph, but that they were from budding elements (blastema) found between normal tissues. It was Rudolph Virchow (1821–1902), Muller's student, who demonstrated that all cells came from other cells, ushering in the era of the cell theory. Rudolph, however, held the view that cancer was caused by chronic irritation, but spread like fluid or liquid. While he did not associate the movement of cancer cells in body fluid, this may have helped the correct hypothesis of Karl Thiersch, who in the 1860s demonstrated that malignant cells spread as cancer "cells" and not as fluids or just liquids.

Another misconception of cancer came from two Dutch doctors who studied familial breast cancer patients independently. Zacutus Lusitani (1575–1642) and Nicholaes Tulp (1593–1674) popularized their findings in 1649 and 1652, respectively, that cancer was a contagious disease, leading to the implementation of different healthcare strategies for cancer patients and hospitals.

While we now know a lot about cancer, these early observations are informative, lending support for what we can further learn from the role of body fluids in cancer.

3.6.2 Issues with Cancer Screening

Cardinal requisites for an optimal screening test include high sensitivity to avoid false negatives, high specificity to avoid false positives, ease of assay performance, cost-effectiveness, accessibility to intended population at risk, ease of performance at the point-of-care or at the community level, comfortable and painless to perform, and importantly the benefits of the test must outweigh its harms.

Screening can help with early cancer detection, but it should be noted that the kinetics of cancer development are different and, as such, screening cannot detect all cancers early. Also depending on study design, screening test can be misleading and useless. Unless a carefully planned randomized controlled study is performed (to provide level 1 evidence), a test can have issues with selection bias, lead-time bias, and length-time bias, as well as overdiagnosis that could result in overtreatment.

3.6.2.1 Selection Bias

This is caused by lack of proper randomization. Under coverage, whereby some members in the intended population are not properly represented, and the use of convenient samples are some issues in biomarker discovery. Also the cohort of participants may be skewed due to the nature of the screening test. For example, a test that is unattractive due to high cost, pain, or discomfort will more likely be taken only by those who view themselves to be at high risk for the condition. By sampling such cohort for a study will not reveal the true performance of the test for the entire population it is intended for.

3.6.2.2 Lead-Time Bias

This is when a screening test detects cancer early and yet actually has no effect on the natural course of the disease. With such a test, those diagnosed late will still die at the same time as those diagnosed early. It then becomes questionable whether early detection changes the time course of the disease. Or does it just inform the patient of having the disease and yet has no impact on the prognosis. If so, then early detection of cancer only harms, as it only prolongs the time of psychological stress. Cancer prognostic studies looking at just years of survival are thus inadequate. They need to know the differences in survival between those with cancers diagnosed early from those without early detection by the screening method. Does early diagnosis improve survival time over late diagnosis?

3.6.2.3 Length-Time Bias

Tumors are heterogeneous with regard to their rates of growth. Some grow very rapidly and are clinically detectable in just a few months. Other tumors grow slowly with several years of indolence before becoming clinically evident to be diagnosed. In view of these differences in growth rates, if the same number of both fast growing and slow growing tumors were to occur in a particular year, the screening test will pick up more slow growing than fast growing tumors, because the majority of the rapidly growing tumors will be clinically overt and be diagnosed. Thus, by detecting many of such indolent tumors that are also more likely to have better prognosis may give the false impression that the screening test saves lives, but in actual fact it does not. It really has no impact on survival.

3.6.2.4 Overdiagnosis and Overtreatment

Early detection of cancer can involve precancerous lesions and indolent (virtually nonprogressive) cancers as well as aggressive cancers. In most cases, such

precancerous lesions and indolent cancers will not cause harm. These patients can lead a normal life expectancy without harm from these lesions. Detecting such lesions (overdiagnosis) and treating them (overtreatment) only increases discomfort, causes possible harm, and increases healthcare costs without any benefit to the patient or society. Similarly, evaluating the performance of a test or biomarker that detects such lesions early will lead to the perceived view that the test enables a high cure rate. For instance, if a test detects prostatic intraepithelial neoplasia (PIN) at a high rate of 100 % (and yet only 25 % of PIN actually becomes prostate cancer), then the test will falsely seem to cure 75 % of cancers, while in fact the actual rate of cancer is 25 %.

3.6.3 Mechanism of Cancer Biomarker Release into Body Fluids

Several mechanisms explain how cancer biomarkers may enter the circulation and other body fluids (Fig. 3.3). Circulating tumor cells and dead cancer cells can release large amounts of biomolecules for genomic, transcriptomic, proteomic, and metabolomic analyses. Moreover, cells can actively secrete nucleic acids, proteins, and metabolites into body fluids. Some mechanisms of biomolecular release into body fluids include the following:

- Death of cancer cells in tissues or in circulation (apoptosis or necrosis).
- Breakdown of micrometastatic cancer cells.
- Active or spontaneous release. Some proteins and metabolites are released by this mechanism.
- Tumor-derived microvesicles (e.g., exosomes) can selectively package cancer molecules in their cargo.
- Trauma or surgical procedures including diagnostic or therapeutic maneuvers (e.g., endoscopies and biopsies) may release biomolecules into circulation. But, their levels should return to normal shortly thereafter.
- Tumor angiogenesis associated with porous vasculature should enhance micromolecular and macromolecular entrance into the circulation.

Formation of serum is associated with release of nucleic acids, and they can emanate from circulating tumor cells or white blood cells; thus serum nucleic acid levels are much higher than plasma. In addition to the above mechanisms, the high cellular turnover of epithelial cells implies the constant exfoliation of cells into lumens of various organ systems. Thus, organ or tissue-specific body fluids such as saliva, NAFs, stool, or urine, should be enriched with biomarkers from this source. A particular body fluid for cancer biomarker discovery, especially by proteomic efforts, is tumor interstitial fluid. This protein-rich media is less variable, diluted, or complex as other body fluids such as blood.

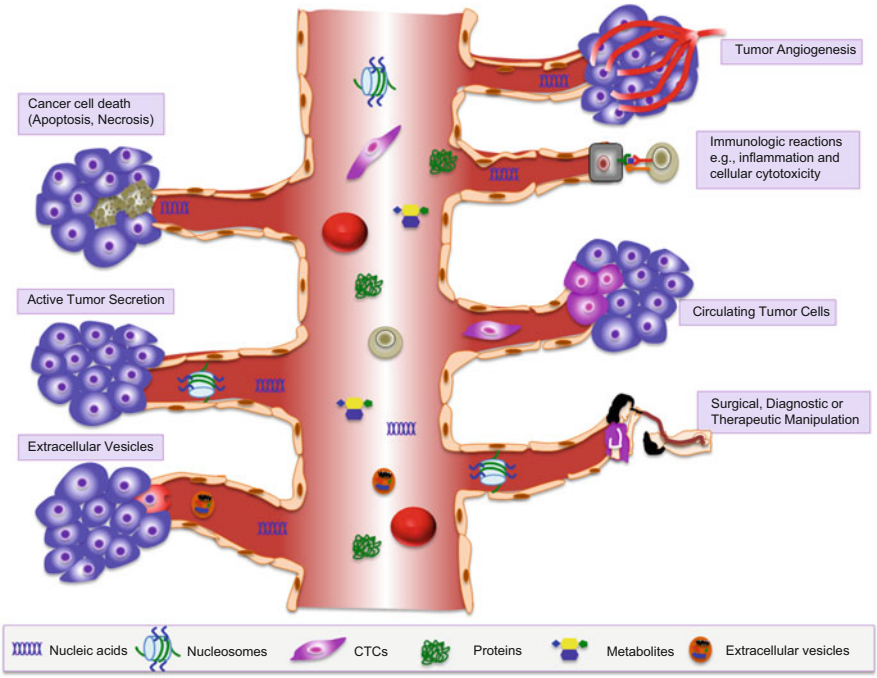

Fig. 3.3 Sources of circulating cancer biomarkers. Biomarkers from cancer cells can enter the circulation and other body fluids by various mechanisms

3.6.4 Body Fluid Handling to Preserve Biomarker Integrity

One important hindrance to biomarker development is sample collection and handling. Pre-analytical variables account for over 60 % of erroneous biomarker discovery and validation and hence contribute to diagnostic inaccuracies. Important issues to be considered for accurate biomarker analyses using body fluid include sample acquisition, preservation, storage, transportation, and processing. Across the world, there are several millions of biospecimens stored in biobanks, and millions more are being added each year. These biorepositories are "treasure chests or goldmines" for biomarker developments, and their value cannot be overemphasized. But to be of real clinical usefulness, these banked samples need standardized collection, preservation, transportation, processing, and storage procedures in order to maintain biomarker integrity.

It may even be important to collect, process, and store some biospecimens for specific biomarker exploration. For instance, the stability of several protein biomarkers including C3a, TNF, fibrinogen, IL-1α, IL1-β, and IFN-γ requires different storage conditions. Samples meant for LDH assays are not to be frozen, and MMP9 degrades at −80 °C storage. The proteome or peptidome is significantly altered by

freeze-thaw cycles. Specifically, MMPs, ADAM, TIMP, transthyretin (TTR), and glycoproteins are sensitive to freezing and thawing of samples.

To minimize errors in biomarker development in the ideal world, patients and controls should be matched for age, gender, physiologic states such as circadian rhythm, and habits such as smoking, alcohol use, exercise, and drug use among other pertinent variables in connection with the research question. However, achieving this is almost a utopian ideology. Hence, strict precautions based on standard operating procedures must be developed and adhered to, such that any sample-to-sample variations may be minimized. At the minimum:

- Samples should be collected at about the same time of day, and with consideration of other possible confounding variables such as fasting or postprandial state.
- Blood can be collected by venipuncture or phlebotomy, but the same procedure must be used to collect all samples.
- For plasma, blood should be collected in an anticoagulant and immediately placed on ice (2–4 °C) prior to subsequent processing or transportation.
- To obtain serum, blood should be allowed to clot for ~30 min at room temperature, following which the tube should be placed on ice before further processing.
- Samples should be transported on dry ice to the analytical laboratory.
- At the processing laboratory, samples should be processed immediately. Centrifugation should be done at low temperatures (2–4 °C). To avoid biomarker degradation due to freeze-thaw cycles, aliquots should be made before freezing at −80 °C.

These minimum recommendations should help reduce some of the pre-analytical variables, especially in proteomic studies.

3.6.5 Precautionary Measures for Handling Body Fluids

Body fluids and other clinical samples are not without danger to those who obtain, assemble, transport, store, process, and analyze them. Several microorganisms, including bacteria, viruses, fungi, and parasites, both pathogenic and nonpathogenic, are found in various body fluids and can be transmitted by direct contact. Most microbes in body fluids are not of a major concern, because they are of low infectivity and/or virulence and are easy to eradicate by the immune system. However, three viruses, human immunodeficiency virus, hepatitis B virus, and hepatitis C virus, are very high on the radar of healthcare providers. While these three infectious agents might not be that ubiquitous in body fluids, all those who work with them, from collection, packaging, transportation, reception, storage, processing, and waste disposal, must take adequate precautions to prevent being infected.

Indeed clinical laboratories handle body fluids on routine basis and there are detailed precautionary measures in place to prevent technicians from being infected. Training on laboratory safety, periodic refresher courses, continuing education on new techniques and products, and review of safety manuals are more valuable for infection control than anything else that can be provided in textbooks. Thus, summarized here are basic safety guidelines. Because the source of body fluid cannot be authenticated with regards to it being infested with pathogens, all body fluids must be considered as potential sources of pathogenic microbes. In view of this, the Centers for Disease Control and Prevention advocates the application of "universal blood and body fluid precaution" to all patient samples. This involves handling and treating all body fluid samples carefully regarding self-protection. Additionally, employers are mandated to ensure acquired knowledge, expertise, skill, and provision of adequate and reliable protective equipment for workers. Basic universal precautions, including the following synopses are mandatory for handling body fluids in the laboratory:

- General barrier protective measures to enhance the natural barrier protection by the skin must be taken. Note that not only mucous membranes are much more susceptible to infections. Wear gloves, marks, protective eyewear or face shields, and gowns or aprons each time body fluids are handled. Before leaving the laboratory, all protective clothing should be removed and hands thoroughly washed with soap.
- Wash hands before and after wearing gloves. Wash any skin surface immediately should they come into contact with body fluids.
- Prevent any injury and direct exposure of your circulation to body fluids. Needles, scalpels, and sharp instruments should rarely be used, and if used, take utmost care in handling them. Avoid recapping needles, and do not attempt to break or bend any needle. Dispose these sharp objects immediately in a puncture-resistant container.
- Body fluid specimens should be collected by trained personnel and secured in containers with tight lids to prevent leakage or spillage. Care must be exercised to prevent contamination of external surfaces of the container or any ancillary accompanying documents or materials.
- All laboratory processes, such as mixing, blending, and sonication that can generate aerosol or droplets, must be performed in class I or II biologic safety cabinets.
- Use germicidal compounds to decontaminate surfaces after work or when there is a spill, as well as laboratory materials before they are disposed or reprocessed for use, and equipment periodically or after a spill.
- Employers must provide orientation and continuing education and training to staff and make available the required equipment and supplies for infection prevention control. Additionally, employers must ensure adherence to recommended precautionary procedures at all times.

3.7 The Peripheral Circulation as a Repository of Cancer Biomarkers

Blood is one of the most explored body fluids for cancer biomarkers, probably because of its accepted use in clinical practice. Also because blood percolates almost all cells in the body, all body tissue biomarkers can be represented in this body fluid (Fig. 3.4). Blood is simply a fluid connective tissue composed of cells and extracellular matrix, the plasma. The cellular components are erythrocytes (RBCs), leukocytes (WBCs), and platelets (megakaryocytes). In the adult, the cellular or formed elements are produced in the bone marrow. Primarily, blood functions to distribute nutrients and oxygen to cells and remove waste materials and carbon dioxide from these cells. Other important functions include homeostasis (buffering and thermoregulation) and transport of hormonal and humoral factors. Plasma is over 90 % water, with the rest being dissolved solutes, mostly proteins (albumin, globins, and fibrinogens). Also are electrolytes, nutrients, hormones, enzymes, nitrogenous substances, and dissolved gases.

Plasma can be obtained by manual or automated methods. Centrifuged blood gives three distinct layers, an uppermost liquid plasma, a middle thin layer composed of WBCs and platelets is referred to as buffy coat, and a larger bottom layer of packed erythrocytes. Plasma devoid of fibrinogen is referred to as serum. Because both plasma and serum are devoid of cells, they serve as useful samples for assaying cell-free biomarkers such as nucleic acids and other clinical biochemical markers. They, however, contain extracellular vesicles.

To obtain plasma, blood is collected in an anticoagulant such as ethylenediaminetetraacetic acid (EDTA) and preferably processed within an hour of collection. Peptides and proteins are stabilized at low temperatures, but storage at

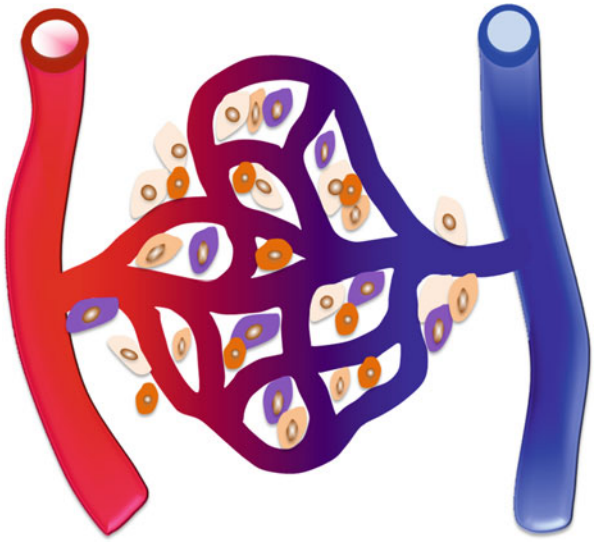

Fig. 3.4 The vasculature. Circulating blood percolates almost all cells of the body. At the capillary level where metabolic exchanges occur, the porous vessels facilitate entry of biomolecules into the circulation. Additionally, biomarkers can enter into the lymphatic circulation (not shown) and be emptied into the blood circulation

4 °C is not advisable because of possible platelet activation. But incorporating a centrifugation step can deplete platelets. Preferably, make aliquots and freeze plasma at −80 °C for future usage. Similarly, to obtain serum, collected blood is allowed to coagulate at room temperature for 30–60 min, the optimum time for clotting. Transport the clotted blood on ice or spin immediately. Do not allow clotted blood to sit on ice for >3 h before centrifugation. Similarly, make aliquots and store at −80 °C for future applications.

The dilemma as to use plasma or serum for circulating cancer biomarker analyses is a challenging one that depends on multiple factors. For instance, several biobanks have archived sera with annotated patient information. These serve as easy sample access for analyses. Similarly should the research question require the release of biomolecules from formed blood elements, serum may be the preferred sample. However, the drawbacks on the use of serum samples are well established. The clotting process leads to the release of biomolecules from other cells into blood fluid, which may confound the study of solid tumor-specific low abundance biomarkers because such molecules from normal cells can overshadow them. Plasma is thus preferred in those circumstances. Plasma also contains less peptides of degradation and the use of platelet-depleted plasma is necessary to avoid platelet peptide interferences. Plasma appears to give less variability in proteomic analyses. But it has been demonstrated that some cancer-specific peptidome biomarkers are generated by cancer-derived proteases ex vivo during the clotting process [3].

Other sources of biomolecular targets from tissues and cells brought into the blood circulation are the lymphatic circulation, the other component of the peripheral circulation. The lymphatic circulation is responsible for removing extracellular fluid and returning it back into the blood circulation. Hence, biomarkers released by cells, including cancer cells, percolate through the lymph nodes and end up in the blood circulation.

3.7.1 Clinical Utility of Biomarkers in Circulation

Biomarkers in circulation are of importance in management of every disease as they easily fit into current clinical practice. Because blood and its associated lymphatic system percolates almost all tissues and cells of the body, secreted molecules or otherwise released biomolecules from cells in “preconditioned cancer fields” or from occult and overt cancer cells are represented in the circulation. Sensitive technologies that are able to siphon these imprints are of tremendous value in cancer early detection and management.

The role of circulating cancer biomarkers in oncology is not new, but current technological advancement on measuring and monitoring biomolecular targets have expanded the utility to encompass the entire spectrum from risk assessment to detection of early recurrence. The diverse utility of circulating cancer biomarkers includes the following:

- Biomarkers in circulation can be sensitive enough for assessing risks of individuals (select population at risk). Because of convenience and minimal discomfort, body fluids offer an acceptable and hence compliant mode of early cancer detection through screening.
- Histologic subtyping and classification of tumors, especially those not easily accessible for biopsy, can be improved by assaying biomarkers in circulation. Such tumor biologic characteristics in circulation help planned management including therapy decision-making.
- The need to select patients for specific treatments based on tumor behavior (indolent vs. aggressive) should rely on both clinical parameters and circulating tumor markers. These prognostic, predictive, or companion diagnostic biomarkers in blood of cancer patients form part of the adage of personalized medicine.
- Because of the availability of several alternative therapies for particular cancers, especially those targeting specific cancer pathways, the "one-size-fit-all" concept is ancient. Therapy targets assayed in circulation can inform initial therapy selection. During treatment for any cancer, circulating biomarkers are important in monitoring response, nonresponse, and toxicity, because serial sampling is convenient. Also because these biomarkers appear early in blood, they offer obvious advantages over imaging and other late measures of therapy response monitoring.
- Circulating biomarkers can effectively detect the recurrence of cancer earlier than conventional means, and this helps early interventional salvage procedures.
- Circulating biomarkers can help identify the primary tissues of origin, in cases of some tumors of unknown primaries.

The enormous translational potential of circulating biomarkers in the management spectrum of cancer patients is realized in several clinics.

3.8 Body Fluid Cancer Biomarkers and Point-of-Care Patient Management

Point-of-care (POC) testing and diagnosis is the ability to test samples or detect changes in individual health in the community, field, client's home, doctor's office, ambulance, or at any health facility. This paradigm shift in healthcare should revolutionize future care, because POC testing enables rapid and timely intervention. The success and impact of POC testing relies on the development of efficient, sensitive, and cost-effective portable devices for sample testing and/or molecular imaging. Devices that can test noninvasive samples such as saliva, urine, and other body fluids including blood (akin to blood glucose testing) are even more readily acceptable. The widespread development of these tests should enable preventive and personalized care, which undoubtedly will reduce the global health disparities

between the developed, developing, and underdeveloped worlds, as well as between remote and urbanized communities.

In an era where cellular telecommunication networks are available in almost all remote parts of the world, POC testing can easily be incorporated into existing telemedicine. Thus, information generated at remote points-of-care can be channeled to the nearest specialist or center where intervention can take place in a timely fashion.

OPKO Diagnostics Inc. and Siemens Healthcare Diagnostics are currently developing POC devices and platform technologies for clinical applications. For instance, OPKO Diagnostics has developed a POC diagnostic device for analysis of protein targets in blood in just about 10 min. It basically performs the typical laboratory blood immunoassay tests quickly on a miniaturized device. In oncology, the pioneering OPKO POC device is a PSA panel, consisting of total PSA, free PSA, testosterone, and novel biomarker targets for prostate cancer management.

3.9 Traditional Circulating Tumor Markers in Clinical Practice

The use of serum biomarkers in clinical practice dates several decades back. Currently, there is a whole armamentarium of these in clinical use. They consist of secreted proteins such as PSA and cell surface-expressed molecules such as HER2/neu proteins. The clinical evidence and hence utility of these biomarkers are detailed in their respective chapters in "Cancer Biomarkers in Body Fluids: Biomarkers in Circulation".

Oncofetal antigens and other serum tumor markers have been in clinical use for several decades. However, their impact on cancer mortality reduction has been at least not more than modest. These tumor markers are so entrenched in clinical practice to the extent that, in spite of their very limited reliability and accuracy, they remain the standard for cancer detection, monitoring of treatment response, and recurrence. The need for emerging sensitive and specific biomarkers is making a change to this landscape. In general, these established tumor markers are nonspecific because they are also expressed in several physiologic and other noncancerous pathologic states and are moreover secreted by different types of cancer. Thus, they are mostly used to monitor surgical and other treatment efficacies and disease recurrences after diagnosis of specific tumors. Table 3.1 illustrates the utility and limitations of some of the commonly used serum biomarkers in the clinic.

Table 3.1 Some common traditional serum biomarkers in clinical practice

Biomarker	Clinical utility	Limitations
AFP	Used to monitor germ cell tumors and HCC. Levels >500 ng/ml is associated with these tumors	Lack of specificity as a screening test, because AFP levels are elevated in gastric, pancreatic, colonic, and bronchogenic carcinomas
B2M	Commonly used to monitor treatment in multiple myeloma. Normal levels are below 2.5 mg/l	Lack of specificity. Levels are also elevated in chronic lymphocytic leukemia and lymphoma
Calcitonin	Useful for diagnosis of medullary thyroid cancer (MTC). Normal levels are <2 pg/ml. Levels >100 pg/ml are associated with MTC	Levels are also elevated in leukemia and lung cancer patients
CA-15.3	Used to monitor breast cancer treatment response and disease recurrence. Normal levels are <30 μg/ml	Elevated levels are seen in liver, pancreatic, gastric, colorectal, lung, ovarian, uterine, and cervical cancers. Also levels are high in inflammatory liver and benign breast diseases
CA-19.9	Used in conjunction with CEA to monitor pancreatic cancer. Normal levels are below 37 U/ml	Levels are elevated in gastric, colorectal, and biliary tract cancers. Pancreatitis, thyroid diseases, and inflammatory bowl diseases also cause increased serum levels
CA-125	Used to monitor treatments of gynecological cancers (ovarian, tubal, and uterine). Normal levels are below 35 μg/ml	Levels are also elevated in lung, breast, and pancreatic cancers and also during pregnancy, menstruation, liver inflammation, pericarditis, endometriosis, and even up to 2 % of healthy individuals
CEA	Used in monitoring medullary thyroid cancer. Also useful in combination with CYFRA21-1 for differential diagnosis of lung cancer. Normal levels are <2.5 ng/ml (5 ng/ml for smokers)	Levels are also elevated in gastric, colon, pancreatic, liver, lung, breast, and ovarian cancers. Pancreatitis, chronic lung disease, cirrhosis, and inflammatory bowel disease also cause levels to rise
CYFRA21	CYFRA21-1 is used for the detection of NSCLC, especially those of SCC histology	Elevated in gastrointestinal, gynecologic, and urologic tumors.
NSE	Biomarker of neuroendocrine tumors similar to chromogranin A. Used in the management of SCLC and neuroblastoma. Levels >100 ng/ml are suggestive of SCLC. Used in therapy response and recurrence monitoring	Levels are elevated in lymphoma, seminomas, and hepatomas and breast, gastric, colorectal, and pancreatic cancers
ProGRP	Levels above 200 pg/ml are suggestive of lung cancer, and even more specific levels >300 pg/ml indicate SCLC	Levels can also climb as high as 200 pg/ml in renal failure
PSA	Prostate cancer screening, with elevated levels >4 ng/ml	Not specific to prostate cancer. BPH, prostatic massage/manipulation increases levels. Screening leads to overdiagnosis and overtreatment
SCCA	Have merit in the differential diagnosis of lung cancer, especially when used in conjunction with CEA and CYFRA21-1	Levels are also elevated in cancers of the head and neck, lung, esophagus, and cervix

3.10 Summary

- Traditional biopsy involves invasive process of tissue acquisition for cytologic, histopathologic, and molecular genetic analyses.
- The "liquid biopsy" concept is a paradigm shift in obtaining and analyzing cancer cells. Liquid biopsy is a noninvasive or minimally invasive mode of sample acquisition that has numerous other advantages over traditional biopsy.
- Cancer biomarkers are objectively measurable entities indicative of cellular transformation, predictive of the natural course and treatment response of cancer, modulate with treatment interventions, and provide indications of cancer recurrence or relapse. While clinically useful in patient management, almost half of cancer biomarkers are coupled to drug development.
- Cancer biomarker development requires careful study design and execution to uncover and validate authentic biomarkers that are safe and effective for their intended applications.
- There are various challenges that biomarker development faces, making only a few being able to meet regulatory standards. Clinically useful cancer biomarkers should meet the highest level of evidence (the Oxford Center for Evidence-Based Medicine levels 1a).
- Biomarkers are released into body fluids through several mechanisms including cell death, secretion, angiogenesis, and trauma.
- All body fluids are potentially hazardous. Hence, precautionary measures must be applied at all times when they are handled.
- Traditional biomarkers, though with their limited accuracies, are still a useful part of cancer management, especially following cancer diagnosis.

References

1. Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. Nat Rev Drug Discov. 2003;2:566–80.
2. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93:1054–61.
3. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271–84.

Chapter 4
Methylated DNA as Cancer Biomarkers in Circulation

Key Topics

- DNA methylation in cancer
- Methylation as field cancerization biomarkers
- Clinical application of circulating methylated tumor DNA (metDNA)
  - Diagnosis
  - Prognosis
  - Treatment monitoring
- Analysis of metDNA in circulating tumor cells (CTCs)
- MetDNA in other body fluids
- Methods for analysis of metDNA
- Commercial products based on DNA methylation

Key Points

- Cancer cells are characterized by aberrant epigenetic modifications, such as DNA methylation. Promoter CpG island hypermethylation inactivates several tumor suppressor genes. Oncogenes (e.g., *RAS*) can be activated via promoter hypomethylation. Normal hypermethylated repetitive sequences such as LINE are hypomethylated in the cancer cell.
- The methylated genes and patterns in cancer cells are stable and can be detected and quantified in ctDNA and CTCs, as well as in other body fluids such as saliva, sputum, and urine.
- The clinical value of circulating cancer epigenetic signatures is being realized with the development of commercial products for clinical applications. The prototype is the FDA-approved methylated *SEPT9* blood test for colon cancer.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_4

4.1 Introduction

Cancer is a disease of epigenetic deregulation induced by environmental cues (epimutagens) or the natural aging process. The epigenetic aberrations in the cancer cell fall into two main categories. First, epigenetic modifications such as the coordinated activities of CpG island methylation of TSG (e.g., *CDKN2A*) promoters by DNMTs, and subsequent binding by methyl-CpG-binding proteins and recruitment of histone-modifying complexes, lead to the creation of inactive chromatin sites and gene silencing. Second, the global cancer genome is hypomethylated. This hypomethylome has a number of consequences: (1) the involvement of repetitive sequences can cause genomic instability, (2) the involvement of imprinted genes (e.g., *IGF2*) can lead to its overexpression, and (3) the involvement of proto-oncogene promoter (e.g., *RAS*) can lead to its activation and overexpression.

DNA methylation, especially tumor suppressor gene promoter hypermethylation and silencing, has been extensively studied in cancer biology. Methylation and aberrant expression are important early events in cellular transformation. While all the factors controlling DNA methylation are yet to be understood, environmental cues and the aging process account for some of the changes. In cancer, these external exposures include known carcinogens such as tobacco smoke, alcohol, chemicals, toxins, and cancer-causing microbes. Similarly, promoter hypermethylation of progeroid genes including *LAMIN A/C* and Werner syndrome gene (*WRN*) is associated with some malignancies. Hypermethylation and inactivation of *WRN* are associated with cancers of the gastrointestinal tract, skin, and sarcoma, while *LAMIN A/C* is silenced in leukemias and lymphomas.

DNA methylation analysis has been extensively conducted using cancer tissue samples. As a diagnostic biomarker, gene promoter methylation demonstrates very high specificity. The sensitivity, however, has always been modest or low, especially for single markers, but tends to increase toward perfection when the right panel of genes is selected as diagnostic biomarker. This finding suggests almost all cancer cells harbor some gene promoter hypermethylation and that multiple genes are silenced by methylation in cancer progression, possibly in an initial polyclonal fashion.

The methylated genes in cancer tissue samples tend to show matched presence in the peripheral circulation and other body fluids. Thus, DNA methylation in blood, saliva, bronchial washes, sputum, and urine has all been extensively evaluated for cancer detection and management. The clinical relevance is being advanced by a number of companies including Epigenomics AG, Orion Genomics, and MDxHealth, with the development of diagnostic and personalized oncology products. Prototypes of these efforts are the FDA-approved methylated *SEPT9* blood test for colon cancer and the CE-marked methylated SHOX2 in bronchial washes as a reflex test for lung cancer detection in patients with inconclusive cytology. This chapter focuses on the relevance of DNA methylation as noninvasive biomarkers

and has intentionally excluded other epigenetic modifications such as ncRNA (e.g., miRNA) and histone modifications in cancer.

4.2 DNA Methylation and Cancer

The cancer genome shows altered DNA methylation patterns that favor malignancy. DNA methylation is an enzymatic reaction that adds methyl groups to the 5′ position of cytosine pyrimidine ring at CpG dinucleotides. The DNA methyltransferase (DNMT) family of enzymes (DNMT1, DNMT3A, and DNMT3B) catalyzes the reaction, using S-adenosylmethionine as cofactor. Transcriptional silencing by CpG island methylation involves histone modification and chromatin-modifying factors. The creation of a transcriptionally inactive chromatin domain requires information flow from 5mC to the histone-modifying factors by methyl-CpG-binding proteins (MBDs). These MBDs are reversibly associated with hypermethylated CpG islands in gene promoters and can be removed by demethylating agents to allow gene expression. Not surprising therefore, cancers demonstrate aberrant expression of a number of these factors involved in DNA methylation. DNMTs, for example, are overexpressed, partly through gene duplication, in a number of cancers including breast, pancreatic, and colorectal cancers, and DNMT3B is mutated in ~20 % of acute myeloid leukemia (AML).

Complex diverse processes control DNA hypomethylation. The TET (ten-eleven translocation) family of proteins is involved in this process. This protein family possesses a DNA demethylase activity that can convert methylcytosines back to cytosine. The mechanism involves oxidation of methylcytosine to 5-hydroxylmethylcytosine (5hmC), leading to the reduced DNMT1 activity. Similarly, *TET* is downregulated in a number of cancers including breast and liver cancers, and *TET2* is mutated in AML.

The cancer cell also demonstrates unique DNA methylation patterns. First, ~50 % of all gene promoters harbor clusters of CpG dinucleotides referred to as CpG islands. Hypermethylation at these promoters often silence tumor suppressor genes (TSGs) in cancer. While this is important in cancer biology, not all methylation of gene promoters play a role in tumor initiation and progression. The "passenger genes with promoter hypermethylation" are referred to as "methylated in tumor" or MINT genes. On the other hand, in some types of cancers such as in colorectal cancer, promoter hypermethylation may be more important than mutations in simultaneously silencing a number of TSGs, to drive carcinogenesis. This phenomenon is called the CpG island methylator phenotype (CIMP). Second, the cancer cell genome demonstrates reduced methylation at CpG clusters outside of promoter regions, as well as CpG dinucleotides in repetitive sequences. Hypomethylation in coding genes can cause oncogene activation and may also be associated with mitotic recombination and genomic instability. Third, DNA methylation is an early event in cancer progression that demonstrates the concept of

molecular field cancerization, which is important in cancer risk assessment and profiling of tumor margins.

In a number of cancers, methylation and inactivation of genes involved in repair of DNA damage constitute the major molecular alterations. Two illustrative examples are provided herein. First, the DNA mismatch repair system involves several genes including *MSH2*, *MSH3*, *MSH6*, *MLH1*, *MLH3*, *PMS1*, and *PMS2*. Loss of function in these genes can cause various levels of elevated mutation rates in the cell, which may result in microsatellite instability (MSI). Two types of microsatellite alterations are recognized, low (MSI-L) and high (MSI-H), which are defined using five microsatellite markers (BAT25, BAT26, D2S123, D5S346, and D17S250). Microsatellite stable (MSS) tumors do not harbor any alterations at these loci. However, instability at two or more of these loci constitutes MSI-H and at a single locus is MSI-L. Of importance, in patients with MSI-H colorectal cancer, promoter hypermethylation and loss of *MLH1* accounts for 80–90 % of all cases. Second, O^6-methylguanine-DNA methyltransferase (*MGMT*) is another DNA repair gene that is methylated and silenced in many cancers. MGMT repairs O^6-methylguanine back to normal guanine, otherwise it is read by the DNA polymerase as an adenine nucleotide. Thus, loss of MGMT function can lead to G:C to A:T transition mutations during DNA replication. Indeed, MGMT is mutated at various frequencies in a number of cancers including head and neck cancer (61 %), colorectal cancer (46 %), esophageal cancer (44 %), glioblastoma (40 %), gastric cancer (37 %), diffuse B-cell lymphoma (36 %), and NSCLC (21 %). A number of genes demonstrate frequent promoter hypermethylation in different cancers (Table 4.1).

An advantage of knowledge on DNA methylation is the fact that this is a reversible process that can be targeted for cancer chemoprevention and treatment. There are FDA-cleared agents, including azacitidine and decitabine, that are useful for treatment of myelodysplastic syndrome and acute myeloid leukemia. These agents also have utility in treatment of some solid tumors and are undergoing clinical trials.

4.3 DNA Methylation and Molecular Field Cancerization

As an early event in malignant cellular transformation, DNA methylation demonstrates molecular field cancerization. This epigenetic phenomenon likely occurs in some cancers and serves as the triggering event that alters gene expression and signaling networks as well as genomic instability to drive cellular transformation. DNA hypermethylation has been demonstrated in several precursor or precancerous lesions, and in histologically normal cells adjacent to surgically resected tumors, suggestive of the multistep model of carcinogenesis. A few illustrative examples are provided herein.

In HNSCC, promoter hypermethylation of *CCNA1*, *CDKN2A*, and *DCC* in histologically normal surgical margins could predict all patients who developed recurrent disease [1]. Similarly, promoter hypermethylation of *DAPK* in normal

Table 4.1 Some genes with promoter hypermethylation in several cancers

Gene	Function	Cancers with methylation
CDKN2A/p16	Cell cycle	Breast, lung, head and neck, colon, kidney, brain, lymphoma
CDKN2B/p15	Cell cycle	Lymphoma, leukemia
CDKN2A/p14	P53 pathway	Stomach, esophagus, colon, kidney, bladder
MGMT	DNA repair	Stomach, esophagus, pancreas, colon, glioma
RARβ2	Vitamin response	Stomach, liver, colon, neuroblastoma
RASSF1A	Apoptosis, cell cycle, DNA repair	Breast, lung, kidney, bladder, prostate, ovary, glioma, neuroblastoma
DAPK	Apoptosis	Lung, head and neck, ovary, leukemia, lymphoma
MLH1	DNA repair	Stomach, liver, colon, ovary
APC	WNT pathway	Stomach, esophagus, pancreas, liver, colon
CRBP1	Vitamin response	Stomach, liver, colon, bladder, lymphoma
TIMP3	Metastasis, angiogenesis	Breast, lung, liver, kidney, glioma
CDH1	Cell adhesion	Breast, esophagus, leukemia
BRCA1	DNA repair	Breast, lung, liver, colon
WRN	DNA repair	Breast, head and neck, stomach, colon, sarcoma
GSTP1	Detoxification	Breast, lung, liver, colon, prostate
RUNX3	Apoptosis	Lung, head and neck, stomach, liver
PRDX1	Antioxidant	Skin
P73	P53 pathway	Kidney, leukemia, lymphoma
IGFBP3	Cell proliferation, growth	Skin, lung, colon, leukemia
EXT1	Metastasis	Skin, stomach, sarcoma, neuroblastoma
EMP3	Metastasis	Neuroblastoma, glioma
VIM	Metastasis	Esophagus, colon, cervix
TWIST	Metastasis	Breast
LAMIN A/C	Nuclear lamina, transcriptional control	Breast, stomach, colon
SFN	Cell cycle, apoptosis	Breast, lung

surgical margins was prognostic in oral squamous cell carcinoma patients as it was significantly associated with shorter overall survival [2, 3]. Indeed, in saliva samples from high-risk individuals for head and neck cancer with premalignant lesions, *EDNRB*, *HOXA9*, and *DCC* methylations were demonstrated. DNMT is overexpressed in precancerous lesions of the pancreas [4], and hypermethylation of *MGMT* and *MLH1* is demonstrated in pancreatic intraductal papillary mucinous neoplasm [5]. Additionally, early-stage pancreatic precursor lesions harbor DNA methylations that increase in frequency with disease progression [6]. Progressive increases in methylation frequencies have also been demonstrated in normal to precursor lesions and to overt gastric cancer [7]. Prostatic intraepithelial neoplastic lesions demonstrate *GSTP1* promoter hypermethylation [8].

The gene promoter hypermethylations demonstrated in "cancer fields" have clinical implications. First, with proper research design, cancer-associated methylated genes can be identified and validated. These biomarkers can then be used to develop noninvasive products for screening, at least for those at elevated risk, for early detection of cancer and possible chemoprevention. Second, DNA methylation analysis of tumor margin tissues should enable prudent selection of those who may need adjuvant therapy to help curtail local recurrences.

4.4 Circulating MetDNA as Diagnostic Biomarkers

Methylated DNA in circulation of cancer patients have been explored as diagnostic biomarkers. Since the first demonstration of methylated genes in serum samples from lung [9] and liver [10] cancer patients, a number of genes with promoter methylation have been associated with other cancers as well (Table 4.2). Circulating methylated tumor DNA (metDNA) as diagnostic biomarkers have a few features in common worthy of note:

- There is a strong general correlation between gene promoter hypermethylation in primary tumors and those in body fluids.
- Promoter hypermethylation is a frequent and early event in cancer development.
- Almost all cancers harbor aberrant promoter methylation. Indeed, the use of panel markers in tissue samples (to overcome tumor heterogeneity) can detect all types of cancer.

Table 4.2 Methylated genes in circulation as diagnostic cancer biomarkers

Cancer	Methylated genes
Melanoma	*FASSF1A*, *RARβ2*, *ER-α*, *MGMT*
Breast cancer	*RASSF1A*, *DAPK*, *APC*, *RARβ2*, *CDKN2A*
Head and neck cancer	*MGMT*, *DAPK*, *CDKN2A*, *GSTP1*
Lung cancer	*DAPK*, *CDKN2A*, *GSTP1*, *MGMT*, *RASSF1A*, *RARβ2*, *TMS1*, *APC*, *CDH13*, *SHOX2*, *SFN*
Esophageal cancer	*APC*, *CDKN2A*, *CDKN2* (*p14*/*ARF*)
Gastric cancer	*CDKN2A*, *CDKN2B*, *CDH1*, *DAPK*, *GSTP1*
Pancreatic cancer	*CDKN2A*, *PENK*
Hepatocellular carcinoma	*CDKN2A*, *CDKN2B*
Colorectal cancer	*CDKN2A*, *SEPT9*, *DAPK*, *APC*, *HLTF*, *HMLH1*, *THBD*, *CAHM*
Renal cell carcinoma	*APC*, *CDKN2A*, *CDKN2 (p14/ARF)*, *CDH1*, *GSTP1*, *MGMT*, *RASSFA1*, *TIMP3*, *RARβ2*
Prostate cancer	*GSTP1*
Ovarian cancer	*BRCA1*, *RASSF1A*, *APC*, *CDKN2A*, *DAPK*, *CDKN2* (*p14*/*ARF*)

- The diagnostic sensitivity of methylation biomarkers is generally low, especially when single genes are considered, but the specificity is mostly high.
- Selecting the appropriate genes as panel methylation biomarker significantly enhances the sensitivity of cancer detection.
- The methyl group is covalently attached to the cytosine pyrimidine ring, such that methylation is very stable and robust and can withstand some of the rough handling during sample processing.

A few examples are provided to illustrate the diagnostic potential of metDNA in circulation. The prototype of diagnostic metDNA in circulation is *SEPT9* methylation, which has been developed and approved by the FDA (referred to as Epi proColon®) for colorectal cancer detection. As a diagnostic biomarker of colorectal cancer, the sensitivity and specificity of circulating methylated *SEPT9* (*meSEPT9*) are 72–90 % and 88–90 %, respectively [11]. Circulating metDNA of another single gene, *THBD*, achieved comparable performance, with a sensitivity of 71 % and specificity of 80 % in a pilot study of colorectal cancer patients [12]. In lung cancer patients, circulating me*SHOX2* achieved a diagnostic sensitivity of 60 % and specificity of 90 % [13]. The promise of panel biomarkers has also been informative. The diagnostic sensitivity of circulating methylated *RASSF1A* and *RARβ2* was 95 %, with 100 % specificity for breast cancer. In esophageal cancer, a sensitivity of 82.2 % and specificity of 100 % were achieved for a panel of circulating methylated *RASSF1A*, *CDH1*, *DAPK*, and *CDKN2A* [14]. A panel of four genes, *GSTP1*, *APC*, *RASSF1A*, and *SFRP1*, achieved a sensitivity, specificity, positive predictive value, and negative predictive value of 92.7 %, 81.9 %, 90.5 %, and 87.2 %, respectively, in discriminating between hepatocellular carcinoma patients and healthy controls and corresponding performances of 84.7 %, 81.1 %, 89.7 %, and 73.2 % when people with benign liver conditions were used as controls [15]. Salvianti et al. showed a significant increase in the frequency and levels of *meRASSF1A* in melanoma patients than controlled healthy people [16]. However, there was no correlation between *meRASSF1A* and CTCs, and the two complemented each other in the detection of metastatic melanoma. Methylated *RASSF1A*, *CACLA*, and *EP300* in circulation had a sensitivity of 90 % and specificity of 87 % for discriminating between ovarian cancer patients and healthy controls, and as a panel, methylated *RASSF1A* and *PGR* performed at a sensitivity and specificity of 80 % and 73.3 %, respectively, for differentiating between women with ovarian cancer and benign ovarian disease [17].

4.5 Circulating MetDNA as Prognostic Biomarkers

Several methylated genes in tumor tissue samples have been used to predict aggressive disease with increased likelihood of recurrence and have also been linked to various outcome measures of disease survival. Some of these prognosis-associated methylated genes have been assayed in circulating DNA and demonstrated to have prognostic potential (Table 4.3).

Table 4.3 Methylated genes in circulation as prognostic cancer biomarkers

Cancer	Methylated genes
Melanoma	*RASSF1A*, *ESR1*, *RARβ2*, *AIM1*, *LINE1*
Breast cancer	*GSTP1*, *ESR1*, *RARβ2*, *SOX17*, *SFN*, *PITX2*, *PR*, *BRCA1*
Lung cancer	*SFN*, *BRMS1*
Esophageal cancer	*APC*, *DAPK*, *MSH2*
Gastric cancer	*RASSF1A*, *APC1*, *SOX17*, *TIMP3*, *MINT2*, *XAF1*
Colorectal cancer	*DAPK*, *CDKN2A*, *MGMT*, *RUNX3*, *HLTF*, *HPP1*, *CDH1*
Bladder cancer	*APC1*
Prostate cancer	*RASSF1A*, *GSTP1*, *APC*, *MDR1*, *PTGS2*, *SRD5A2*, *CYP11A1*
Neuroblastoma	*RASSF1A*, *DCR2*

In breast cancer, *meSOX17* in blood was associated with tumor-node-metastasis (TNM) stage, and in multivariate analysis was an independent prognostic factor [18]. Similarly, me*RARβ2* in plasma or serum samples predicts worse outcomes in women with breast cancer [19, 20]. In gastric cancer, *meSOX17* in serum samples correlated with tumor differentiation status and overall patient survival [21]. As well, methylated *TIMP3* and *XAF1* in circulation were independent predictors of poor disease-free survival in gastric cancer patients [22, 23].

Tumor burden or TNM stage and degree of differentiation and aggression may all reflect on the amounts of tumor DNA released into the circulation. Advanced-stage large tumors release more DNA into the circulation and are obviously associated with worse prognosis. The facile detection of high levels of circulating biomarkers, even with less analytically sensitive technologies, implies the detection frequencies in advanced-stage tumors should be high. It is thus unclear whether the simple detection of methylated DNA in circulation and their association with disease outcomes reflect tumor burden. In other words, patients with low-volume tumors will have low levels of circulating metDNA (less amounts), which will be infrequently detected (low frequency of detection in this population). It should however also be noted that, apart from "methylated in tumor" (MINT) genes, promoter hypermethylation of TSGs, hypomethylation of oncogenes, and other CpG clusters even in repetitive elements have biologic consequences on tumor initiation and progression. Therefore, aberrant methylation in tumors could have functional prognostic predictions of tumor behavior.

4.6 Circulating MetDNA as Treatment Response Biomarkers

Tumor DNA is very dynamic in circulation. The natural course of ctDNA appears to be short. Diehl et al. demonstrated a half-life of ~2 h for ctDNA [24]. Hence, with tumor removal or shrinkage from various treatment interventions, and hence reduction in the source of ctDNA, the levels should decline with time. This thesis forms

Table 4.4 Methylated genes in circulation as treatment response and predictive cancer biomarkers

Cancer	Methylated genes
Melanoma	*RASSF1A*
Breast cancer	*RARβ2*, *BRCA1*, *MSH2*, *ESR1β*, *RASSF1A*, *SFN*
Lung cancer	*RASSF1A*, *CHFR*, *STRATIFIN*, *SOX2*, *APC*, *RARβ2*
Esophageal cancer	*APC*
Gastric cancer	*RUNX3*
Hepatocellular carcinoma	*CDKN2A*
Prostate cancer	*GSTP1*
Neuroblastoma	*DCR2*

the basis for the profiling of ctDNA changes in postoperative monitoring for the completeness of surgery or for the efficacy of chemotherapy treatment. Given the specificity of DNA methylation biomarkers in cancer, and the ease of performing methylation assays, the potential utility of metDNA in disease monitoring has been demonstrated as well (Table 4.4).

In breast, esophageal, and gastric cancer patients, postoperative decline in metDNA in circulation has been associated with complete tumor removal. In breast cancer patients, the ratio of *meCDKN2A* to total *CDKN2A* in circulation (referred to as the methylation index) was 35 % in preoperative compared to 3.5 % in postoperative plasma samples [25]. Postoperative and tamoxifen treatment decreased methylated gene detection in plasma samples from breast cancer patients. *MSH2*, *RARβ2*, and *ESR1β* showed significantly more methylation in preoperative than postoperative plasma from ER^+ breast cancer patients [26]. Methylated *APC* is a predictor of worse outcome in esophageal cancer. Thus, *APC* and *DAPK* methylation have been associated with shorter overall survival, and detection of *meAPC* in serum samples 10 days after surgery was indicative of the presence of residual disease [27]. Methylated *RUNX3* is associated with advanced-stage gastric cancer. Median postoperative *RUNX3* methylation index was 12-fold lower than preoperative levels in gastric cancer patients. Serum *meRUNX3* was a better predictor of gastric cancer recurrence than CEA [28]. Tumor dynamics following chemotherapy also reflect early on circulating biomarkers. Consistently, circulating metDNA has proven useful in monitoring tumor burden in cancer patients on treatment (Table 4.4). Additionally, circulating metDNA may have predictive utility. Loss of serum *meRASFF1A* a year after tamoxifen treatment was associated with good prognosis. In this cohort of women with resectable breast cancer, detection of *meRASSF1A* one year following tamoxifen treatment was an independent predictor of poor recurrence-free survival (RR, 5.1) and overall survival (RR, 6.9) [29]. A decrease in methylation levels of a panel of genes (*AKR1B1*, *ARHGEF7*, *GPX7*, *COL6A2*, *HOXB4*, *TM6SF1*, *TMEFF2*, *RASGRF2*, *HIST1H3*, and *RASSF1A*) in breast cancer patients after 1–2 cycles of docetaxel and imatinib or capecitabine treatment was associated with stable disease or partial response [30]. Patients with progressive disease had no reductions in methylated ctDNA. On follow-up, a rise in

methylated serum DNA was associated with disease progression, and this was evident prior to clinical indices of disease progression. In metastatic breast cancer, a continuous decline in serum methylated *stratifin* (*SFN*) was associated with response to chemotherapy [31]. Additionally, in breast cancer patients on chemotherapy, reductions in total serum methylated *BRCA1*, *MDR1*, *MGMT*, *GSTP1*, and *SFN* were predictive of reductions in tumor volume in responders [32]. Importantly, circulating *meBRCA1* frequency demonstrated significant correlation with chemotherapy.

High levels of methylated *SFN* in NSCLC patients predicted sensitivity to cisplatin plus gemcitabine chemotherapy [33]. In NSCLC patients on neoadjuvant chemotherapy, the methylation indices of serum *RASSF1A* and *RARβ2* decreased after chemotherapy, with further decreases after surgery. Detectable serum methylations of *RASSF1A* and *RARβ2* occurred in patients with disease recurrence 9 months following treatment [34]. Moreover, decreased plasma *meSHOX2* was an early and accurate predictor of response to chemotherapy in lung cancer patients. Indeed, high baseline plasma *meSHOX2* achieved an area under the receiver operating characteristic curve (AUROC) of 0.939 after five treatment cycles in differentiating between responders and nonresponders [35].

In prostate cancer patients on docetaxel or mitoxantrone treatment who were followed for a median of 15 months, a rise in *meGSTP1* in plasma following first dose of treatment was predictive of subsequent PSA relapse [36]. Not only was circulating *meGSTP1* a biomarker of early resistance to therapy, but also it was a better prognostic predictor.

4.7 DNA Methylation and CTCs

The epithelial-mesenchymal transition (EMT) process involves aberrant gene expressions. Profiling studies identifies genes aberrantly methylated in association with EMT [37]. Thus, several genes silenced or activated during EMT have promoter hypermethylation or hypomethylation, respectively. These genes include *CDH1*, *CDH2*, *VIM*, and *TWIST*. A class of genes including *Epb4113* is dispensable in tumor development and growth, but required inactivated at the primary tumor site to facilitate metastasis. These genes are called “metastasis suppressor genes” [38]. Interestingly, these “metastasis suppressor genes” are often silenced via epigenetic mechanisms. It will be informative to know the cells in which these genes are silenced and their role in the metastatic cascade. The question as to whether these are CTCs and CCSCs, and also whether these phenotypes are retained in circulation, has been addressed.

A number of studies have examined gene methylation status in CTCs. The paucity of reported studies addressing this important question may be due to a number of reasons. Current technologies are not sensitive enough to enable direct single-cell genome analysis. This limitation necessitates whole genome amplification (WGA) before downstream methylation detection. However, DNA methylation patterns are not faithfully preserved during the WGA process. This caveat can

also be overcome by initial bisulfite modification before amplification. But, bisulfite conversion may introduce DNA fragmentations precluding the capture of the entire methylome of the cell. Notwithstanding these issues, a genome-wide methylation content of single cells has been successfully carried out in embryonic cells using single-cell locus-specific sequencing and reduced representation bisulfite sequencing [39, 40]. Thus, these approaches may likely be applied to CTC/CCSC methylomics, which will enhance our knowledge on the metastatic cascade and potentially offer epigenetic therapeutic targeting.

A few curious pilot studies have however questioned the clinical relevance and biology of DNA methylation in CTCs. One mechanism regulating EMT of invasive cancer cells was unveiled using murine models of hepatocellular carcinoma [41]. Following establishment of tumors in mice, CTCs were isolated and demonstrated to have increased metastatic potential and mesenchymal phenotypes, evidenced by decreased E-cadherin and increased vimentin and fibronectin expression. The EMT phenotype was partly induced by increased HGF and cMET expression, consequent to hypomethylation of their promoters in CTCs. Thus, aberrant gene methylations in association with their expressional changes may control the EMT process, and CTCs appear to retain these phenotypes.

In human studies, Chimonidou et al. first provided evidence of gene promoter methylations in putative CTCs [42–44]. This group enriched for CTCs using anti-EpCAM-coated beads and centrifugation, and methylation analysis was performed by qualitative methylation-specific PCR (MSP). Promoter methylation of three tumor suppressor genes, *cystatin M6* (*CST6*), *breast cancer metastasis suppressor 1* (*BRMS1*), and *SRY* (*sex-determining region Y*)*-box 17* (*SOX17*), was targeted in CTCs enriched from women with breast cancer. In one study, methylation frequency was higher in CTCs enriched from patients than healthy controls. Moreover, the methylation frequencies increased from women with localized resectable cancer to those with metastatic breast cancer [42]. In a follow-up study, a significant correlation in *SOX17* promoter methylation between CTCs and ccfDNA was demonstrated [43]. In another study, *BRMS1* methylation only weakly correlated with loss of protein expression in CTCs [44]. Note that all these studies were conducted on crudely enriched CTCs without purification and a nonquantitative detection technique was used. These limitations should be considered in data interpretation. Nonetheless, these are interesting proof-of-concept studies. Friedlander et al. used a more reliable CTC enrichment strategy to reveal methylation patterns in CTCs [45]. Viable CTCs were enriched from men with metastatic castration-resistant prostate cancer (mCRPC) with the Vita-Assay™ (Vitatex Inc.). An array-based methylation platform with 27,000 CpG sites was used to analyze methylation in the isolated CTCs. Of interest, the mean methylation values of the enriched CTCs were identical to that from primary mCRPC tissue samples, and these values were higher than those from benign prostate tissue. Indeed, the vast majority (86 %) of the 1361 reported hypermethylated loci in mCRPC were represented in the enriched CTCs. Not surprisingly, pathway analysis revealed the association of the genes mostly methylated to VEGF, angiogenesis, and apoptosis signaling pathways.

4.8 MetDNA in Other Body Fluids

Apart from blood, metDNA has been analyzed in other proximal body fluids such as saliva, sputum, bronchial washes, cerebrospinal fluids, and urine. For certain types of cancer, these tissue-associated body fluids offer a better advantage in regard to being enriched with tumor DNA and hence enable sensitive detection of even early-stage disease. Representative examples are provided here for saliva, sputum, and urine.

Saliva, oral rinse, and oral mucosal brush samples are collected noninvasively for analysis of metDNA. Examples of some methylated genes used for diagnosis and monitoring for head and neck squamous cell carcinoma (HNSCC) relapse are provided. Promoter hypermethylation of *MGMT*, *DAPK*, and *CDKN2A* was detected in 56 % of tumor samples and 55 % of corresponding matched saliva [46]. The work by Righini et al. uncovered hypermethylation of *CDKN2A*, *MGMT*, *DAPK*, *CDH1*, *TIMP3*, and *RASSF1A* promoters in 82 % and 78 % of tumor and saliva samples, respectively [47]. Promoter hypermethylation of *KIF1A* and *EDNRB* was associated with HNSCC, and as a potential early detection biomarker, *EDNRB* methylation was present in saliva from patients with premalignant lesions [48]. Moreover, in oral mucosal brushes, *CCNA1*, *DCC*, and *TIMP3* were methylated at frequencies of 60.4 %, 54.2 %, and 35.4 %, respectively [49]. LINE-1 hypomethylation patterns in salivary rises and white blood cells were sensitive and specific for oral squamous cell carcinoma [50, 51]. The potential of using methylated genes in saliva as recurrent monitoring biomarkers has also been addressed. In multivariate analysis, posttreatment promoter methylation of *TIMP3* was an independent predictor of poor local recurrence-free survival of HNSCC patients [52]. Additionally, promoter methylation of *DAPK*, *DCC*, *MINT31*, *CDKN2A*, *TIMP3*, *MGMT*, and *CCNA1* in saliva was associated with local recurrences and survival [53, 54].

Although not representative of peripheral adenocarcinomas, sputum is a noninvasive surrogate sample for lung cancer risk biomarker assay development. Because sputum is easily obtained from smokers who are at elevated risk for lung cancer, metDNA as lung cancer biomarkers have been examined in this media. Hypermethylation and silencing of *CDNK2A* promoter is common to many cancers. In lung cancer, *CDNK2A* promoter was hypermethylated in 84 % of primary tumors and in as many as 76 % of matched sputum samples [55]. Other lung cancer-associated genes with promoter hypermethylation detectable in sputum include *MGMT*, *RASSF1A*, *HOXA9*, *DAPK*, and *MAGEA3*.

Urine has also served as a useful noninvasive surrogate sample for analysis of urogenital cancer biomarkers. A number of promising methylated genes in urine when used as panels have shown impressive accuracies for bladder cancer detection. As examples, urinary methylation of *VIM*, *GDF15*, and *TMEFF2* achieved a sensitivity and specificity of 94 % and 100 % respectively for bladder cancer detection [56]. Methylated *RARβ2* and *APC* in urine when used as a panel were superior diagnostic biomarkers of bladder cancer compared to cytology alone. The

use of urinary cytology with the other two methylated genes as panel biomarker achieved a sensitivity of 80–94 % for detecting different grades of bladder cancer [57]. Methylation of *POU4F2*, *2NF154*, *HOXA9*, and *EOMES* was identified from global methylation scanning of bladder tumor samples. When assayed in urine as a panel, the sensitivity and specificity were 84 % and 96 % respectively for bladder cancer [58]. Some of these methylated genes are early events in bladder cancer and hold potential for noninvasive risk assessment especially in high-risk individuals such as smokers. Urinary sediments have also been surrogate samples for prostate cancer detection. Methylation of *RASSF1* and *RARβ* in urine achieved a good sensitivity for early-stage prostate cancer detection [59]. Some outstanding issues with urinary biomarker analysis are in the standardization of urine collection and processing for biomarker analysis.

4.9 Methods for MetDNA Analysis

There are several methods and techniques available for DNA methylation analysis. These methods have a few noted features including:

- They may be quantitative, semiquantitative, or qualitative. In clinical applications, quantitative measurements are preferred.
- They are applied on either bisulfite-modified or bisulfite-unmodified genomic DNA. The former is currently the method of choice for methylation analysis.
- The regions and extent of the genome targeted vary by use of various technologies (Fig. 4.1). They may be gene or locus specific as in promoter CpG island methylation, involve large-scale genome-wide approach, or whole global methylome interrogation for 5mC content.
- The choice of technology usually depends on the research question, genome coverage needed, cost, throughput, and maybe just what technology is available at a given medical research center.

Bisulfite treatment of DNA causes the conversion of unmethylated cytosines to uracil. Methylated cytosines are refractory to this modification. Therefore, this process allows the use of several technologies to differentiate between methylated (unconverted cytosines) and unmethylated (cytosines converted to uracil) genomic regions.

There have been a number of technologies for DNA methylation analysis including high-performance liquid electrophoresis (HPLE) or high-performance liquid chromatography (HPLC), methylation-specific restriction and PCR, restriction landmark genome scanning (RLGS), amplification of intermethylated sites (AIMS), and methylation arbitrarily primed PCR (MS-AP-PCR). Methylation-sensitive high-resolution melting (MS-HRM) analysis is based on sequence differences between methylated and unmethylated sequences after bisulfite conversion. Melting dissociation curves of test samples are compared with those from where standard PCR products from same loci where methylation is expected. This method

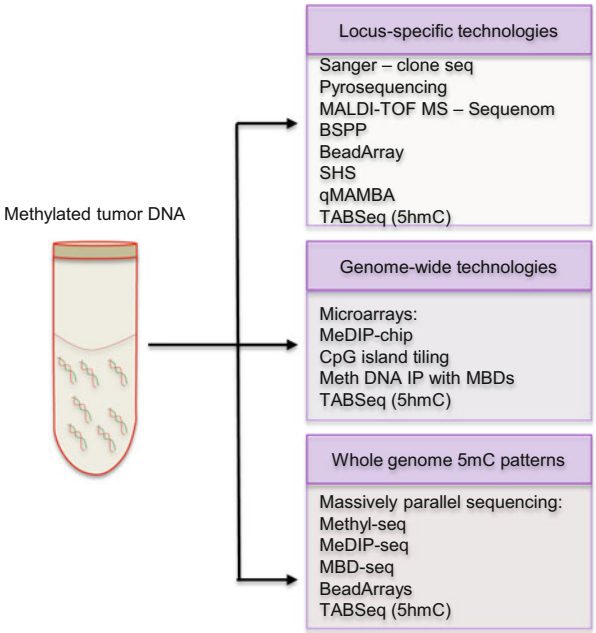

Fig. 4.1 The various methylation detection methods can be used to target different extents of genomic coverage

is one of the sensitive approaches to methylation detection. Other methods and technologies for methylation analysis include the following:

- *Methylation-specific PCR (MSP).* Specific primers can be used in conventional PCR to target bisulfite-converted and bisulfite-unconverted sequences so as to discriminate between unmethylated and methylated DNA (Fig. 4.2). This approach, however, is qualitative and therefore not suitable for clinical applications. A modified version, quantitative MSP (qMSP), also known as MethyLight, has numerous advantages over the gel-based traditional qualitative MSP. First, the quantitation of methylated DNA enables its utility for genes that are methylated at low levels or frequencies in normal or benign cells, but at much higher levels in malignant cells of the same tissue or organ. In this case, a cutoff value can be established for making discriminatory calls. Second, the quantitative value of metDNA can have an important clinical meaning. Values may be determined above which a prognostic prediction can be made. Moreover, dynamic changes in the levels of metDNA may be used to monitor for treatment response, disease relapse, or minimal residual disease. Third, qMSP offers some standardization by providing a universal control gene (*ACTB*) as well as reagents. This should facilitate some level of interlaboratory data comparison, although not everyone uses the same reagents or TaqMan assay. Fourth, labor, cost, and time associated with post-PCR gel analysis are eliminated. Finally, and most important for body fluid analysis, qMSP is analytically very sensitive. Its sensitivity of 1 in 10,000 (equivalent to six methylated alleles or 20 pg of metDNA) is far superior to that of 1 in 1000 for conventional MSP.

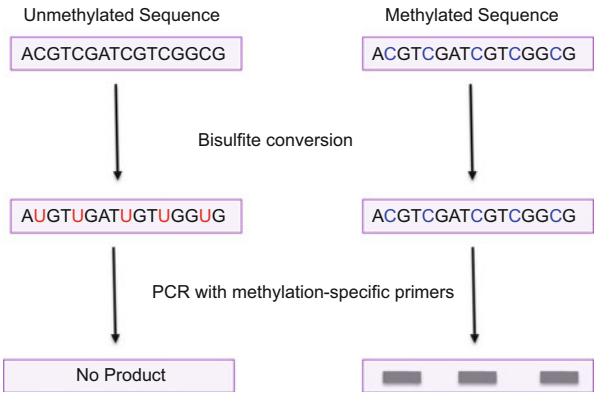

Fig. 4.2 Methylation-specific PCR (MSP). Depicted is qualitative MSP, but MethyLight generates quantitative amplification data in real time

- *Combined bisulfite conversion restriction analysis* (COBRA) uses enzymatic digestion during bisulfite conversion to discriminate between 5mC and unmethylated cytosines. This technology is based on the loss or retention of the restriction enzyme site after bisulfite conversion, which depends on the state of methylation of the involved cytosine at the restriction site.
- *Sequencing technologies*. Different sequencing technologies can be applied to bisulfite-converted DNA. This approach usually provides information on methylation levels or patterns (hypomethylation or hypermethylation) across the whole genome or targeted large genomic regions. It is essentially a screening process used to identify targeted areas of interest where locus-specific analysis (e.g., by qMSP) can be applied. Clone sequencing, pyrosequencing, and next-generation sequencing can all be applied to bisulfite-converted DNA.
  - PCR amplification of bisulfite-converted DNA followed by cloning and sequencing using Sanger chemistry (Clone Seq) can be used for methylation analysis.
  - Pyrosequencing is high-throughput real-time sequencing based on DNA polymerase activity and luminometric detection of released pyrophosphate at each step of nucleotide incorporation. Pyrosequencing is a very quantitative sequencing approach, but is optimal for small PCR products of 50–100 bp, which may miss some large CpG genomic regions. Additionally, pyrosequencing has low analytical sensitivity of ~5 %, which is unsuitable for body fluid analysis.
  - Next-generation sequencing. Massively parallel sequencing enables the generation of methylome maps at single nucleotide level. Bisulfite-converted DNA can be subjected to whole genome shotgun sequencing (MethylC-Seq) to provide a map of genome methylation pattern. Other sequencing approaches include reduced representation bisulfite sequencing (RRBS), single-cell locus-specific bisulfite sequencing (SLBS), solution hybrid selection (SHS), use of bisulfite padlock probes (BSPP), and BeadArrays.

  - Apart from requiring small amounts of DNA input, methods are developed to capture and enrich for methylated regions for sequencing. Antibodies that target 5mC can be used for immunoprecipitation of methylated DNA (MeDIP) for sequencing (MeDIP-Seq). Similarly, immobilized recombinant CpG-binding proteins enable the capture of methylated genomic regions for sequencing (MBD-Seq).

- *MALDI-TOF mass spectrometry (Sequenom).* Global methylation patterns can be achieved by mass spectrometric estimation of 5mC. Enzymatic hydrolysis of DNA can be subjected to HPLE or HPLC coupled with MS to separate individual methylated and unmethylated nucleotides. MS technology is a high-throughput, low-cost, very rapid, semiquantitative screening of genome methylation content.
- *Microarrays.* Arrays enable high-throughput analysis of metDNA patterns in the genome. Several iterations are available for whole genome tilling to focus targeting of gene promoters or CpG islands in the genome. The approaches used include:
  - Commercial platforms such as the Illumina GoldenGate® BeadArray™ and Infinium® Human Methylation 450.
  - Restriction digestion and hybridization of DNA enable analysis of differential methylation patterns.
  - Affinity enrichment of DNA such as 5mC-based antibody immunoprecipitation followed by hybridization (MeDIP-chip).
- Other technologies amenable to body fluid analysis include Tet-assisted bisulfite sequencing (TABSeq) and quantitative methylation analysis of minute DNA amounts after whole bisulfitome amplification (qMAMBA) [60]. TET activity and hence the degree of genome hypomethylation are measurable via TABSeq. This strategy enables single nucleotide resolution that can provide genome-wide or targeted loci maps of 5hmC. The qMAMBA approach involves an initial genome-wide amplification of bisulfite-converted DNA. This increases the template amounts required for quantitative methylation analysis by pyrosequencing. An enormous amount of methylation information is thus garnered form just little initial starting template. The qMAMBA method is compatible with other technologies such as massively parallel sequencing, microarrays, and Illumina Infinium® methylation platform.

4.10 Challenges with MetDNA Analysis in Body Fluids

DNA methylation analysis in body fluids for cancer management holds tremendous potential partly due to its high specific association with cancer. However, standardization of various aspects of the analytical process is required for the development

of robust noninvasive methylation biomarkers. There are a number of issues that hamper interlaboratory data comparison, and these include:

- Issues with specimen type used for circulating metDNA analysis. Plasma is preferable; however, serum has been used in a number of studies for various reasons. Moreover, the inherent issues of sample collection, processing, storage, and stability are pertinent variables as well.
- The efficacy or completeness of bisulfite conversion needs careful consideration. Incomplete bisulfite modification of unmethylated cytosines can occur. Note that such unconverted cytosines to uracil will be the same as 5mCs. Therefore, false-positive results may occur if PCR parameters are not stringent enough.
- Copy number or amounts of metDNA in the body fluid are important for detection. The circulating levels of tumor DNA, for instance, may depend on tumor burden, degree of invasiveness, differentiation, and tumor cell death, all of which will differ from patient to patient and among different samples from different laboratories. Compounding this variable is the analytical sensitivity of chosen technology and amount of input DNA.
- When examining promoter hypermethylation, the choice of target gene, primers, and genomic region targeted by the primers are all variables. For instance, the same gene may be assayed in the same tumor type by different laboratories using primers targeting different regions of the same promoter. Analytical sensitivity issues are bound to arise.
- PCR parameters such as annealing temperatures used for DNA amplifications may differ.
- Choice of technology or technology platform with their different analytical sensitivities will affect data output.

4.11 Commercial Products Based on MetDNA

The potential clinical utility of epigenetic biomarkers is reflected on the number of companies applying various technologies in the development of commercial clinical products. These companies include Epigenomics AG, Orion Genomics, and MDxHealth.

Epigenomics AG is focused on developing noninvasive products based on DNA methylation for use in cancer diagnosis, prognosis, and monitoring as well as personalized medicine. The company uses its proprietary DNA methylation biomarkers for product development. Its marketed products include Epi proColon® and Epi proLung®. Epi proColon® is a blood test based on methylation in *SEPT9*. It is intended as a screening test for early detection of colorectal cancer. Epi proLung® BL Reflex Assay is intended to help confirm the diagnosis of lung cancer in cases of inconclusive cytology. It is based on *meSHOX2* in bronchoalveolar lavage fluid.

The molecular diagnostic unit of Orion Genomics develops cancer risk assessment and predictive products based on epigenetic modifications. It has two proprietary platforms, MethylScope® and MethylScreen™, used for epigenetic biomarker development. MethylScope® is a microarray platform used for genome-wide methylation pattern detection, while MethylScreen™ is a quantitative PCR approach used for targeted detection and quantification of gene methylation. The company's lead product is a colon cancer risk test. This test is used to assess an individual's risk for developing colon cancer, such that those at elevated risk may be selected for screening. It is a blood test based on aberrant methylation and loss of imprinting of the proto-oncogene, *IGF2*, which is associated with colorectal cancer development. Their pipeline products include breast, bladder, lung, and other cancers.

MDxHealth provides epigenetic information for personalized cancer diagnosis and treatment. The company focuses on developing products for urologic cancers, with its marketed lead product being ConfirmMDx. It uses its patented methylation-specific PCR technology and proprietary portfolio of genes for product development. ConfirmMDx targets biopsies from men at elevated risk for prostate cancer, but who return a negative outcome. The benign-looking biopsies are examined by ConfirmMDx to help differentiate between those who are really true negative and those with missed tumors as a result of biopsy sampling error. The company is also developing products for bladder, kidney, and other urologic cancers. It has licensed its two non-urologic products, Cologuard® for colon cancer to Exact Sciences and PredictMDx® for glioblastoma to LabCorp.

4.12 Summary

- The cancer cell is studded with epigenomic alterations at all levels. DNA methylation and associated histone modifications, chromatin structural changes, and ncRNA expressional changes affect genes involved in cancer progression.
- Cancer cells often have promoter hypermethylation and inactivation of TSGs. Additionally, the cancer cell genome exhibits global hypomethylation, which may activate oncogenes or alter the integrity of repetitive elements leading to genomic instability.
- DNA methylation is an early event in carcinogenesis and demonstrates multistep carcinogenesis and field cancerization.
- DNA methylation changes in cancer are also present in body fluids of cancer patients. These methylated tumor DNA can be detected and quantified as biomarkers for cancer diagnosis and management.
- Genes involved in metastatic cascade often demonstrate alterations in DNA methylation. The role of these changes in CTC production and phenotypes is unclear. Data suggest the methylation patterns are conserved; however, this is an active area of research.
- Clinical products based on DNA methylation of body fluid targets are being developed and commercialized.

References

1. Tan HK, Saulnier P, Auperin A, et al. Quantitative methylation analyses of resection margins predict local recurrences and disease-specific deaths in patients with head and neck squamous cell carcinomas. Br J Cancer. 2008;99:357–63.
2. Supic G, Kozomara R, Jovic N, et al. Prognostic significance of tumor-related genes hypermethylation detected in cancer-free surgical margins of oral squamous cell carcinomas. Oral Oncol. 2011;47:702–8.
3. Schussel J, Zhou XC, Zhang Z, et al. EDNRB and DCC salivary rinse hypermethylation has a similar performance as expert clinical examination in discrimination of oral cancer/dysplasia versus benign lesions. Clin Cancer Res. 2013;19:3268–75.
4. Peng DF, Kanai Y, Sawada M, et al. Increased DNA methyltransferase 1 (DNMT1) protein expression in precancerous conditions and ductal carcinomas of the pancreas. Cancer Sci. 2005;96:403–8.
5. House MG, Guo M, Iacobuzio-Donahue C, Herman JG. Molecular progression of promoter methylation in intraductal papillary mucinous neoplasms (IPMN) of the pancreas. Carcinogenesis. 2003;24:193–8.
6. Sato N, Fukushima N, Hruban RH, Goggins M. CpG island methylation profile of pancreatic intraepithelial neoplasia. Mod Pathol. 2008;21:238–44.
7. Lee JH, Park SJ, Abraham SC, et al. Frequent CpG island methylation in precursor lesions and early gastric adenocarcinomas. Oncogene. 2004;23:4646–54.
8. Brooks JD, Weinstein M, Lin X, et al. CG island methylation changes near the GSTP1 gene in prostatic intraepithelial neoplasia. Cancer Epidemiol Biomarkers Prev. 1998;7:531–6.
9. Esteller M, Sanchez-Cespedes M, Rosell R, et al. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Res. 1999;59:67–70.
10. Wong IH, Lo YM, Zhang J, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res. 1999;59:71–3.
11. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 2014;63:317–25.
12. Lange CP, Campan M, Hinoue T, et al. Genome-scale discovery of DNA-methylation biomarkers for blood-based detection of colorectal cancer. PLoS One. 2012;7:e50266.
13. Kneip C, Schmidt B, Seegebarth A, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J Thorac Oncol. 2011;6:1632–8.
14. Li B, Wang B, Niu LJ, et al. Hypermethylation of multiple tumor-related genes associated with DNMT3b up-regulation served as a biomarker for early diagnosis of esophageal squamous cell carcinoma. Epigenetics. 2011;6:307–16.
15. Huang ZH, Hu Y, Hua D, et al. Quantitative analysis of multiple methylated genes in plasma for the diagnosis and prognosis of hepatocellular carcinoma. Exp Mol Pathol. 2011;91:702–7.
16. Salvianti F, Orlando C, Massi D, et al. Tumor-related methylated cell-free DNA and circulating tumor cells in melanoma. Front Mol Biosci. 2015;2:76.
17. Liggett TE, Melnikov A, Yi Q, et al. Distinctive DNA methylation patterns of cell-free plasma DNA in women with malignant ovarian tumors. Gynecol Oncol. 2011;120:113–20.
18. Fu D, Ren C, Tan H, et al. Sox17 promoter methylation in plasma DNA is associated with poor survival and can be used as a prognostic factor in breast cancer. Medicine (Baltimore). 2015;94:e637.
19. Fujita N, Nakayama T, Yamamoto N, et al. Methylated DNA and total DNA in serum detected by one-step methylation-specific PCR is predictive of poor prognosis for breast cancer patients. Oncology. 2012;83:273–82.
20. Mirza S, Sharma G, Parshad R, et al. Clinical significance of promoter hypermethylation of ERbeta and RARbeta2 in tumor and serum DNA in Indian breast cancer patients. Ann Surg Oncol. 2012;19:3107–15.

21. Balgkouranidou I, Karayiannakis A, Matthaios D, et al. Assessment of SOX17 DNA methylation in cell free DNA from patients with operable gastric cancer. Association with prognostic variables and survival. Clin Chem Lab Med. 2013;51:1505–10.
22. Yu JL, Lv P, Han J, et al. Methylated TIMP-3 DNA in body fluids is an independent prognostic factor for gastric cancer. Arch Pathol Lab Med. 2014;138:1466–73.
23. Ling ZQ, Lv P, Lu XX, et al. Circulating methylated DNA indicates poor prognosis for gastric cancer. PLoS One. 2013;8:e67195.
24. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14:985–90.
25. Wong IH, Zhang J, Lai PB, et al. Quantitative analysis of tumor-derived methylated p16INK4a sequences in plasma, serum, and blood cells of hepatocellular carcinoma patients. Clin Cancer Res. 2003;9:1047–52.
26. Liggett TE, Melnikov AA, Marks JR, Levenson VV. Methylation patterns in cell-free plasma DNA reflect removal of the primary tumor and drug treatment of breast cancer patients. Int J Cancer. 2011;128:492–9.
27. Hoffmann AC, Vallbohmer D, Prenzel K, et al. Methylated DAPK and APC promoter DNA detection in peripheral blood is significantly associated with apparent residual tumor and outcome. J Cancer Res Clin Oncol. 2009;135:1231–7.
28. Sakakura C, Hamada T, Miyagawa K, et al. Quantitative analysis of tumor-derived methylated RUNX3 sequences in the serum of gastric cancer patients. Anticancer Res. 2009;29:2619–25.
29. Fiegl H, Millinger S, Mueller-Holzner E, et al. Circulating tumor-specific DNA: a marker for monitoring efficacy of adjuvant therapy in cancer patients. Cancer Res. 2005;65:1141–5.
30. Fackler MJ, Lopez Bujanda Z, Umbricht C, et al. Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. Cancer Res. 2014;74:2160–70.
31. Zurita M, Lara PC, del Moral R, et al. Hypermethylated 14-3-3-sigma and ESR1 gene promoters in serum as candidate biomarkers for the diagnosis and treatment efficacy of breast cancer metastasis. BMC Cancer. 2010;10:217.
32. Sharma G, Mirza S, Parshad R, et al. DNA methylation of circulating DNA: a marker for monitoring efficacy of neoadjuvant chemotherapy in breast cancer patients. Tumour Biol. 2012;33:1837–43.
33. Ramirez JL, Rosell R, Taron M, et al. 14-3-3sigma methylation in pretreatment serum circulating DNA of cisplatin-plus-gemcitabine-treated advanced non-small-cell lung cancer patients predicts survival: The Spanish Lung Cancer Group. J Clin Oncol. 2005;23:9105–12.
34. Ponomaryova AA, Rykova EY, Cherdyntseva NV, et al. Potentialities of aberrantly methylated circulating DNA for diagnostics and post-treatment follow-up of lung cancer patients. Lung Cancer. 2013;81:397–403.
35. Schmidt B, Beyer J, Dietrich D, et al. Quantification of cell-free mSHOX2 Plasma DNA for therapy monitoring in advanced stage non-small cell (NSCLC) and small-cell lung cancer (SCLC) patients. PLoS One. 2015;10:e0118195.
36. Mahon KL, Qu W, Devaney J, et al. Methylated Glutathione S-transferase 1 (mGSTP1) is a potential plasma free DNA epigenetic marker of prognosis and response to chemotherapy in castrate-resistant prostate cancer. Br J Cancer. 2014;111:1802–9.
37. Carmona FJ, Davalos V, Vidal E, et al. A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition. Cancer Res. 2014;74:5608–19.
38. Steeg PS, Ouatas T, Halverson D, et al. Metastasis suppressor genes: basic biology and potential clinical use. Clin Breast Cancer. 2003;4:51–62.
39. Guo H, Zhu P, Wu X, et al. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. 2013;23:2126–35.
40. Gravina S, Ganapathi S, Vijg J. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res. 2015;43:e93.

41. Ogunwobi OO, Puszyk W, Dong HJ, Liu C. Epigenetic upregulation of HGF and c-Met drives metastasis in hepatocellular carcinoma. PLoS One. 2013;8:e63765.
42. Chimonidou M, Strati A, Tzitzira A, et al. DNA methylation of tumor suppressor and metastasis suppressor genes in circulating tumor cells. Clin Chem. 2011;57:1169–77.
43. Chimonidou M, Strati A, Malamos N, et al. SOX17 promoter methylation in circulating tumor cells and matched cell-free DNA isolated from plasma of patients with breast cancer. Clin Chem. 2013;59:270–9.
44. Chimonidou M, Kallergi G, Georgoulias V, et al. Breast cancer metastasis suppressor-1 promoter methylation in primary breast tumors and corresponding circulating tumor cells. Mol Cancer Res. 2013;11:1248–57.
45. Friedlander TW, Ngo VT, Dong H, et al. Detection and characterization of invasive circulating tumor cells derived from men with metastatic castration-resistant prostate cancer. Int J Cancer. 2014;134:2284–93.
46. Rosas SL, Koch W, da Costa Carvalho MG, et al. Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. Cancer Res. 2001;61:939–42.
47. Righini CA, de Fraipont F, Timsit JF, et al. Tumor-specific methylation in saliva: a promising biomarker for early detection of head and neck cancer recurrence. Clin Cancer Res. 2007;13:1179–85.
48. Demokan S, Chang X, Chuang A, et al. KIF1A and EDNRB are differentially methylated in primary HNSCC and salivary rinses. Int J Cancer. 2010;127:2351–9.
49. Longo AL, Rettori MM, de Carvalho AC, et al. Evaluation of the methylation profile of exfoliated cell samples from patients with head and neck squamous cell carcinoma. Head Neck. 2014;36:631–7.
50. Subbalekha K, Pimkhaokham A, Pavasant P, et al. Detection of LINE-1s hypomethylation in oral rinses of oral squamous cell carcinoma patients. Oral Oncol. 2009;45:184–91.
51. Pobsook T, Subbalekha K, Sannikorn P, Mutirangura A. Improved measurement of LINE-1 sequence methylation for cancer detection. Clin Chim Acta. 2011;412:314–21.
52. Rettori MM, de Carvalho AC, Bomfim Longo AL, et al. Prognostic significance of TIMP3 hypermethylation in post-treatment salivary rinse from head and neck squamous cell carcinoma patients. Carcinogenesis. 2013;34:20–7.
53. Carvalho AL, Henrique R, Jeronimo C, et al. Detection of promoter hypermethylation in salivary rinses as a biomarker for head and neck squamous cell carcinoma surveillance. Clin Cancer Res. 2011;17:4782–9.
54. Sun W, Zaboli D, Wang H, et al. Detection of TIMP3 promoter hypermethylation in salivary rinse as an independent predictor of local recurrence-free survival in head and neck cancer. Clin Cancer Res. 2012;18:1082–91.
55. Liu Y, An Q, Li L, et al. Hypermethylation of p16INK4a in Chinese lung cancer patients: biological and clinical implications. Carcinogenesis. 2003;24:1897–901.
56. Costa VL, Henrique R, Danielsen SA, et al. Three epigenetic biomarkers, GDF15, TMEFF2, and VIM, accurately predict bladder cancer from DNA-based analyses of urine samples. Clin Cancer Res. 2010;16:5842–51.
57. Eissa S, Swellam M, El-Khouly IM, et al. Aberrant methylation of RARbeta2 and APC genes in voided urine as molecular markers for early detection of bilharzial and nonbilharzial bladder cancer. Cancer Epidemiol Biomarkers Prev. 2011;20:1657–64.
58. Reinert T, Modin C, Castano FM, et al. Comprehensive genome methylation analysis in bladder cancer: identification and validation of novel methylated genes and application of these as urinary tumor markers. Clin Cancer Res. 2011;17:5582–92.
59. Daniunaite K, Berezniakovas A, Jankevicius F, et al. Frequent methylation of RASSF1 and RARB in urine sediments from patients with early stage prostate cancer. Medicina (Kaunas). 2011;47:147–53.
60. Vaissiere T, Cuenin C, Paliwal A, et al. Quantitative analysis of DNA methylation after whole bisulfitome amplification of a minute amount of DNA from body fluids. Epigenetics. 2009;4:221–30.

Chapter 5
Nucleic Acids as Cancer Biomarkers in Circulation

Key Topics

- History of circulating cell-free DNA (ccfDNA) in cancer
- Biology of ccfDNA
- Circulating nucleosomes
- Circulating microbial genomes
- Pre-analytical issues and recommendations in ccfDNA analysis
- Translational potential of circulating DNA as cancer biomarkers
- Technologies for circulating tumor DNA (ctDNA) analysis

Key Points

- The demonstration that nucleic acids, especially DNA, from all cell types in the body are found in body fluids has opened up a new discipline of scientific enquiry, the siphoning of disease biomarkers in circulation.
- These biomolecules circulate in free forms, but are often bound to cell surfaces, proteins, or lipids, and are also encased in circulating extracellular vesicles.
- Circulating tumor DNA analysis provides information about tumor adaptive behaviors due probably to metastatic pressures, or the toxic effects of chemotherapy. This is clinically very useful in making treatment decisions as new cancer mutations are detected in ctDNA.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_5

5.1 Introduction

Cell-free nucleic acids (cfNA), or extracellular nucleic acids (excNA), are found in body fluids of both healthy individuals and people with various diseases including cancer. These nucleic acids are in different fluid compartments. They are in the blood and lymphatic circulation, and these are referred to appropriately as circulating cell-free DNA (ccfDNA), RNA (ccfRNA), or simply nucleic acids (ccfNAs). They are also present in other body fluids, such as saliva, bronchoalveolar lavage fluid, breast fluid including milk, urine, CSF, stool, bile, ascitic fluid, pancreatic juice, urine, semen, prostatic fluid, menstrual flow, and many other body fluids, and can easily be harvested from cell culture media. Since these other body fluids are noncirculating throughout the body, the nucleic acids found in them are simply referred to as cell-free or extracellular nucleic acids (cfNAs or excNAs) in body fluids. Much attention has been focused on ccfNAs, and their potential utility in clinical oncology is emerging. Specifically, ccfNAs derived from cancer cells are referred to as circulating tumor NAs (ctNAs; ctDNA or ctRNA).

Much attention is placed on the translational potential of ctDNA because of the authentic detection and measurement of tumor-specific molecular and genetic alterations in these samples. Indeed, the robust detection of tumor-specific DNA methylation in ctDNA has proven its strong potential for clinical translation. Another area of intense work for translational purposes is detection of tumor genetic changes in circulation. This has several clinical applications, including augmenting diagnostic work-up, but most intensive work has been focused on disease prognosis, therapy selection and monitoring, as well as recurrence monitoring. These potential clinical applications are only being made possible with the development of sensitive technologies, especially for mutation detection in ctDNA, given that ccfDNA from other sources tend to mask their presence in circulation. Moreover, ctDNA levels are very low, especially in early-stage disease and for some tumors. Approaches such as genome-wide profiling of somatic copy number variations/alterations in ctDNA enable the biologic evolution of cancer to be appreciated. The biologic features and biomarker potential of ccfNAs and ctDNA in cancer is addressed in this chapter. Some important technologies for mutational analysis of ctDNA are summarized.

5.2 Historic Recap of ccfDNA in Cancer

Disease exploration including early detection and monitoring of cancer using circulating nucleic acids was explored decades ago and has obtained much attention in recent years with the increased search for noninvasive means of cancer biomarker development. The pioneering work of Mandel and Metais in 1947 revealed the presence of free nucleic acids in the circulation; however, the diagnostic potential of plasma nucleic acids in oncology had to wait several decades before

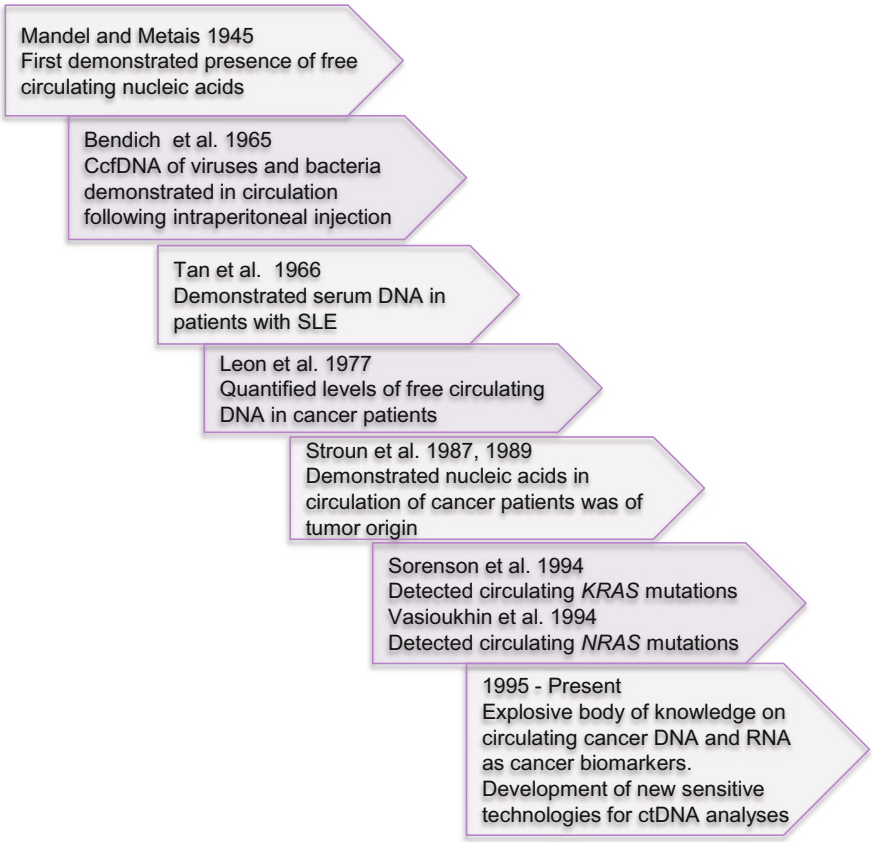

Fig. 5.1 The historical landscape of ccfDNA and ctDNA in cancer

resurgence [1] (Fig. 5.1). Collaborative evidence for the presence of circulating cell-free nucleic acid came from rodent experiments. Active cell-free DNA of infectious viruses and bacteria was observed in mice following intraperitoneal injections [2]. Similarly, in patients with systemic lupus erythematous, serum DNA was demonstrated using immunoassays [3].

With regards to cancer, the early works of Leon and Stroun are noteworthy. In 1977, Leon and coworkers reported on the use of radioimmunoassay to quantitatively measure the levels of ccfDNA in sera from cancer patients and healthy control individuals [4]. In this initial study, all their patients had received some form of treatment; however, ccfDNA levels were still much higher in cancer patients than in controls. Analysis of cancer progression and ccfDNA levels indicated that patients with metastatic disease had much higher levels than those with localized disease. That apart, the levels of free serum DNA in patients correlated with various clinicopathologic parameters. Indeed, total serum DNA ranges from nanograms to micrograms per milliliter in cancer patients compared

to just less than 200 ng/ml in healthy individuals, with the majority of healthy people having below 50 ng/ml to undetectable levels. Whereas the sources of circulating nucleic acids in cancer patients were largely speculative, Stroun and coworkers first demonstrated that these nucleic acids were indeed of tumor origin in cancer patients, because they failed to detect them in healthy individuals [5, 6].

Subsequently, cancer-specific mutations were demonstrated in blood of cancer patients, ushering in the "liquid biopsy" concept of cancer management. *KRAS* mutations in pancreatic cancer patients [7], and *NRAS* mutations in patients with myelodysplastic syndrome [8] were detected in blood from patients. These novel discoveries opened up a whole new area of cancer biomarker exploration, with its potential in personalized oncology. Indeed, these pioneering works served as catalysts that fueled further studies in the field. Microsatellite instability was detectable in plasma samples from breast cancer patients [9]; *KRAS*, *TP53,* and *APC* mutations in sera from colorectal cancer patients [10]; and gene promoter hypermethylation in sera from lung cancer patients [11]. Currently, the available body of evidence suggests early cancer detection and management using genetic information in circulating nucleic acid is an invaluable clinical tool.

5.3 Features of Circulating Tumor Nucleic Acids

Genomic fragments from nuclear, mitochondria, and cancer-causing microbes can be found in the circulation of cancer patients. Double-stranded DNA of various molecular sizes (181 bp–21 kbp) circulates as nucleoproteins in diseased states, but is mostly cell surface bound in healthy individuals. In particular, ctDNA circulates in the form of (Fig. 5.2):

- Cell-free or unbound DNA.
- Nucleosomal DNA, which are DNA-histone protein complexes.
- DNA bound to other proteins (e.g., lipoproteins).
- DNA attached to cell surfaces (cell surface-bound DNA, csbDNA).
- DNA in CTCs.
- DNA in cancer-derived microvesicles/microparticles, including DNA in cancer-derived apoptotic bodies.

With the exception of nucleic acids in CTCs and microvesicles, the other forms of DNA will be considered as ccfNAs (not enclosed by plasma membrane). However, most ctDNA studies have focused on ccfDNA, nucleosomal DNA, and csbDNA.

While measurements of serum proteins had been part of clinical practice in cancer management for decades, the utility of circulating tumor nucleic acids in oncology is at its infancy, but poised to add a wealth of information to cancer management in the future. Circulating tumor nucleic acids are useful for cancer risk assessment, screening, diagnosis, monitoring, and treatment decision-making, staging, prognostication, and monitoring therapy efficacy. Alterations in the genome,

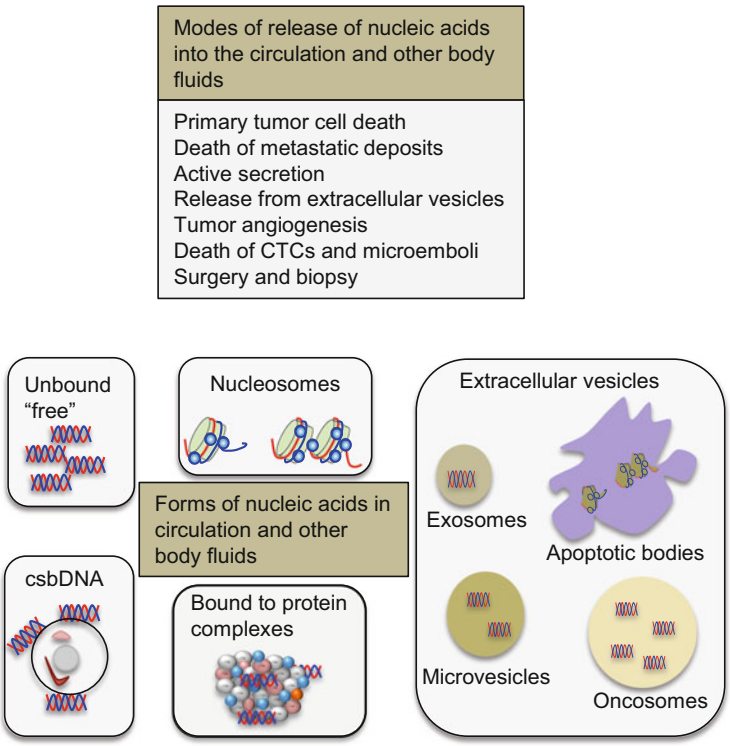

Fig. 5.2 Mechanisms of release and forms of ctDNA. Circulating cancer nucleic acids can be released by various modes including cancer cell necrosis, apoptosis, active secretion, and release from extracellular vesicles. In circulation and body fluids, cancer DNA can be free and unbound or bound to cell surfaces (csbDNA), bound to protein complexes, wrapped around histone proteins, or as vesicle encased

and the processes used for nucleic acid detection in circulation fall under two different categories. These are:

1. *Qualitative* detection of cancer-specific alterations such as:
   (a) Epigenetic modifications; methylation, histone modifications, and altered noncoding RNA expression
   (b) Genetic modifications; LOH, MSA, mutations, SNPs, etc.
   (c) Aberrant tumor-specific transcript expression
2. *Quantitative* measurements including:
   (a) Nonspecific measurements of the levels of total nucleic acids
   (b) Specific measurements of nuclear or mitochondrial RNA or DNA
   (c) Measurement of microbial genomes (e.g., HPV DNA in nasopharyngeal carcinoma patients)

(d) Nucleic acid integrity assays – measurements of the levels of degraded nucleic acids
(e) Measurement of ctDNA levels

Virtually everyone has free circulating nucleic acids. However the levels are much higher in patients with connective tissue diseases, cancer, myocardial infarction, pulmonary embolism, and immunopathologies than in plasma from healthy individuals. In spite of the numerous studies that conclusively indicate the clinical utility of free nucleic acid quantification in disease management, there is no established useful cutoff value or validated data for clinical translation.

5.3.1 Mechanisms of Nucleic Acid Release

Several mechanisms explain how cancer nucleic acids may enter the circulation and other body fluids (Fig. 5.2). These include:

- Death of cancer cells by necrosis or apoptosis.
- Breakdown of micrometastatic cancer cells.
- Active or spontaneous release from cancer cells.
- Tumor-derived exosomal rupture (mainly RNA).
- Other tumor-derived microvesicles (DNA and RNA).
- Apoptotic bodies from cancer cells (DNA and RNA).
- Trauma or surgical procedures.
- Diagnostic or therapeutic procedures (e.g., biopsy procedure).
- Tumor angiogenesis increases DNA release into plasma.
- Formation of serum is associated with release of nucleic acids, and they can emanate from circulating tumor cells.

Because apoptosis is a normal physiologic process of cell turnover, most ccfDNA in healthy individuals are released by this mechanism. This is often primarily of hematopoietic origin but contributed to by death of epithelial and other body cells. In cancer and other pathological conditions, DNA is released mainly by means of cellular necrosis, because cancer cells tend to evade death by apoptosis. Macrophages and other phagocytes scavenge the apoptotic or necrotic cells and subsequently release their DNA and other cellular contents into the circulation or the DNA can passively enter the circulation from the interstitial spaces through porous blood vessels. Noteworthy, both healthy and cancer cells can also actively release DNA into the circulation. In cancer patients, necrosis, apoptosis (to some extent), and direct active release contribute to circulating DNA, but it appears the necrotic source may outweigh the other modes in some cancers. Thus, in general, low molecular weight DNA fragments of almost even sizes (166 bp to 200 bp) and multiples are prevalent in healthy individuals, while cancer patients harbor necrotic high molecular weight DNA (>300 bp) in addition. This thesis is further supported by the increased DNA integrity index (which is the ratio

of high molecular weight to low molecular weight DNA), in some cancer patients and also the findings of cancer mutations in high molecular weight fragments of circulating DNA. But note that informative cancer mutations are associated with even smaller (<100 bp) ccfDNA fragments.

Hypoxia, which occurs as tumors grow rapidly, leads to tumor cell necrosis. It is estimated that a 100 g tumor, which corresponds to 3×10^{10} tumor cells, releases up to 3.3 % tumor DNA daily into the circulation. Circulating disseminated tumor cells can undergo lysis and thus release their contents into the circulation as well. Overall, the levels of circulating tumor nucleic acids are influenced by several factors including tumor type, location, size, degree of differentiation, and stage.

The kinetics of nucleic acid release during cancer development is biphasic. Rodent experiments indicate that DNA release is in phases and from different cell types [12]. An initial tumor-normal adjacent cell interaction induces release of large quantities of non-tumoral DNA into the circulation. The cause of this initial release by non-tumor cell is unclear at this time. The next phase of release is then by tumor cells.

5.4 Features of Cell Surface-Associated Circulating Nucleic Acids

In addition to the nature of circulating NA as free DNA, nucleosomal DNA, or other molecular associations, a good proportion of these nucleic acids are attached to WBCs and RBCs (csbNAs). This association is reversible but can also lead to internalization of the bound nucleic acids. Various mechanisms mediate the possible intake of csbNAs:

- Receptor-mediated process is involved and can lead to endocytic internalization. Implicated receptors in this process include integrins, cytokeratins, scavenger receptors, ezrin, moesin, and albumin.
- Circulating cell-free nucleic acids can also bind through bivalent cation interactions on membrane phospholipids and be internalized.
- Histone proteins (even linker histones) can carry nucleic acids into cells via changes in membrane potential.

Bound nucleic acids vary in sizes and efficiency of interaction with cell surfaces. Most ccfNAs extracted from culture medium range in sizes between 180 bps and 3500 bps. However, surface-bound nucleic acids are much larger (~20 kbp). Extraction of cell surface nucleic acids requires PBS/EDTA or mild trypsinization.

5.5 Features of Circulating Nucleosomes

The human genome is estimated at 3 billion base pairs. This staggering genome is organized on 23 pairs of chromosomes and housed in the minuscule nucleus with an average diameter of ~6 μm. How this arrangement is achieved involves various levels of organization including DNA supercoiling, nucleosomal arrangement, and other layers of compaction. The nucleosome is the initial order of DNA packaging in the nucleus. It is composed of two pairs each of histone H2A, H2B, H3, and H4 that form the core, around which a 146 bp DNA is wrapped around (Fig. 5.3). The nucleosomes are then connected together by a segment of 10–100 bp DNA referred to as the linker region. This organization gives the "beads-on-string" appearance, the tertiary structure of which is referred to as chromatin. The linker regions are associated with histone H1, which contributes to the structural stability of chromatin. In addition to their contribution to DNA packaging in the nucleus, nucleosomes help stabilize DNA and regulate DNA replication, transcription, and repair. Cellular apoptosis leads to fragmentation and release of nucleosomes that can be found in circulation.

5.5.1 Mode of Nucleosomal Release into and Elimination from the Circulation

Nucleosomes are released via some of the mechanisms by which other circulating biomarkers are released. Cells liberate them during differentiation and activated lymphocytes can directly secrete nucleosomes, but with relevance to cancer, most nucleosomes are shed by cell death processes. As cells die, internal endonucleases

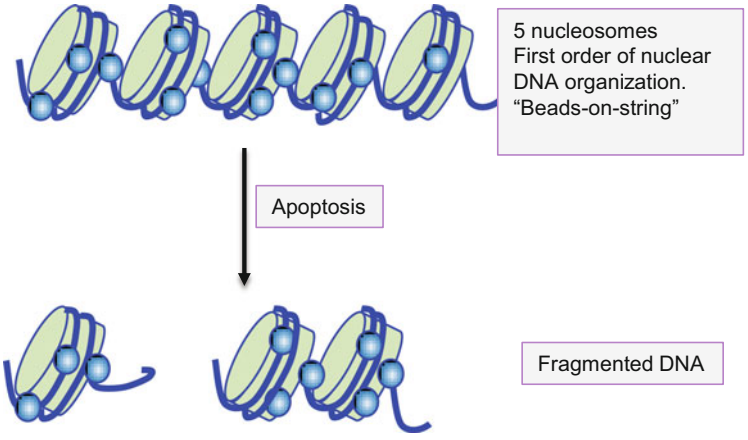

Fig. 5.3 Nucleosomal DNA. The nucleosome is the initial structural organization of nuclear DNA in chromosomes. The DNA molecule is wrapped around special histone proteins

(e.g., caspases, activated DNases) cleave chromatin, usually at the linker regions generating mononucleosomes and oligonucleosomes, which finally enter the blood or lymphatic circulation supplying and draining the tumor. In general, most circulating nucleosomal DNA is of low molecular weight indicative of apoptotic release. However, oncotic (necrotic) cell death in cancer cells suffering from hypoxia can cause the release of larger nucleosomal DNA fragments. Macrophages that engulf necrotic/oncotic cells can release these nucleic acids when they die or get activated.

During cell death, ccfNAs can be observed in the circulation by 24 h, and their release can continue for the next 48 h. There is also considerable variation in tumor-derived circulating nucleosomes between patients (ranges from 3 to 93 % of plasma nucleosomes).

Nucleosomes are cleared from the circulation by various mechanisms, including the following:

- Renal glomerular removal as liposomes.
- Metabolism in organs such as the liver (primary site), kidneys, spleen, lungs, stomach, bowels, and heart.
- Degradation by circulating endonucleases.
- Phagocytosis and digestion by macrophages, monocytes, histiocytes, dendritic cells, and phagocytic endothelial cells.
- Anti-nucleosome binding and removal in immunologic diseases such as SLE.

The rate of elimination also depends on whether DNA is single or double stranded and the amount of nucleosomes in circulation. In general, clearance is slow for double-stranded DNA, and when nucleosomal DNA is in large amounts. Impaired removal also occurs in cases of increased CRP levels as in acute inflammatory conditions. C-reactive protein attaches to the positive charges on histones, impairing their removal from circulation.

5.5.2 Circulating Nucleosomes Mirror Caspase Activity

Caspases are cysteine aspartate proteases involved in cellular apoptosis. They exist as inactive proenzyme forms and are activated by proteolysis. They cause nucleosomal cleavages at the linker regions through effector caspase-dependent activation of DNases. Several cancer therapies cause cancer cell death through apoptosis. Thus, the measurement of circulating apoptotic products including caspase activities in circulation can assess treatment efficacy in cancer patients. In ovarian cancer patients, mean caspase activity in blood prior to chemotherapy is identical to healthy controls. However, upon treatment, caspase activities significantly increased in association with increased circulating nucleosomes [13]. In breast cancer patients, effector caspase 3 and 7 activities in blood were much higher in patients than healthy controls and those with benign conditions, and a trend of

increased caspase activity with tumor progression was observed. Caspase activities are strongly correlated with increased nucleosome levels in circulation [14, 15].

5.6 Circulating Tumor Cells Contribute to Cell-Free Nucleic Acids

In a number of cancers, including melanoma, breast, prostate, head, and neck cancers, CTC and DTC levels correlate significantly with circulating nucleic acids, suggesting some of these nucleic acids are derived from micrometastatic cells that possibly die in the circulation or release nucleic acids by other mechanisms. A number of supporting evidences are provided. For example, in breast cancer, women with DTC had higher levels of circulating DNA than those without DTCs. Cancer gene methylation levels detected in circulating nucleic acids show strong correlations with the presence of CTCs. In prostate cancer, the detection of LOH in circulating nucleic acids is markedly enhanced by the presence of CTCs. The frequency of detection of LOH in cfNAs in bone marrow plasma also increases by the presence of DTCs. Thus, the combined analysis of cfNAs, CTCs, and DTCs may have better clinical value than any of them in isolation.

5.7 Pathogenicity of Circulating Cancer Nucleic Acids

The pathogenic roles of nucleosomes are established for systemic lupus erythematosus (SLE), in which patients have increased levels of anti-nucleosomal antibodies in circulation, even at the early stage of the disease. These antibodies and nucleosomes form immune complexes that get deposited on glomerular basement membranes to mediate disease pathology. In cancer, nucleosomes are also implicated in disease pathology and biology including the development of metastatic disease (genometastasis), escape of tumors from immune surveillance, and tumor angiogenesis.

The genometastasis concept of second primary tumor or distant spread of cancer may be supported by an old experimental observation. Knowledge on the release and eventual uptake and incorporation of “naked” DNA into genomes of other cells isn’t new. Even before our knowledge of the DNA structure and carrier of genetic information, classical work by Griffith demonstrated clearly the biologic importance of cellular transformation (cellular uptake of naked DNA in solution). That, ccfDNA can be taken up and integrated into the genomes of other cells is not surprising then. Multiple evidences have demonstrated the importance of incorporation of cancer genomes, especially oncogenic DNA into other cells. These include:

- Initial rodent experiments proved that ccfDNA from cancer cells can be taken up and integrated into normal cell genomes, leading to the genometastasis hypothesis [16, 17].
- Fibroblasts can take up oncogenic DNA and become transformed with tumor formation potential [18].
- Methylated tumor DNA is more efficient than unmethylated sequences at penetrating and transforming cells [19].
- Oncogenic DNA in plasma from cancer patients can transform normal cells in culture [20].

The genometastasis concept has important clinical implications as postulated by others [21]. That tumors have preferential sites of metastasis is possibly coordinated by specific cell surface receptors for ccfDNA released from a particular tumor. Also, oncogenic ccfDNA could potentially cause cellular transformation in other individuals who receive such blood (i.e., if their system fails to effectively and rapidly remove such DNA from their circulation).

Nucleosomes are also taken up by cancer and immune cells via receptor-mediated mechanisms. The nucleosomes can then undergo intracellular processing in the target cancer cell and used to inhibit recognition and lysis by natural cancer cells. It appears likely that both histones and associated DNA are necessary for this inhibitory function because neither ccfDNA nor histones alone could more effectively perform this function than nucleosomes [22]. Additionally, tumor apoptosis, probably in hypoxic environments, causes the release of nucleosomes to activate NF-kB pathway in endothelial cells. This signaling eventually induces the expression of a potent angiogenic factor, IL-8, which contributes to tumor angiogenesis [23].

5.8 Circulating Microbial Nucleic Acids

Cancer-causing infectious agents include viruses (herpesviruses, hepatitis viruses, retroviruses, polyomaviruses), bacteria (*Helicobacter pylori*), and parasites (schistosomiasis, liver flukes). Circulating genomes of any of these organisms can be detected in patients. However, much research in circulating microbial nucleic acid analyses has focused on Epstein-Barr virus (EBV) and nasopharyngeal carcinoma. Here is a summary of some clinical evidence of the detection of EBV genomes in NPC patients:

- Circulating levels reflect tumor load.
- Higher circulating levels are associated with advanced-stage disease.
- Circulating levels have prognostic value independent of tumor stage.

Over 3 million EBV DNA molecules are released per hour into the circulation to establish a level of 6000 EBV DNA per milliliter. This is the median circulating genomes in early disease. Plasma EBV DNA measurements are adopted for the

detection, monitoring, and prognostication of NPC patients. In view of its strong prognostic evidence, it has been proposed for measurements of circulating EBV DNA levels to be incorporated in risk stratification, to complement the TNM staging system of NPC [24].

5.9 Pre-analytical and Analytical Issues with ccfDNA

The quantity and quality of recoverable ccfDNA is affected by pre-analytical variables that influence subsequence downstream analyses. There are issues with blood sampling and processing including serum preparation, types of anticoagulant used, time between collection and processing, conditions of centrifugation, and storage conditions, times, and temperatures. All these processes need standardization for data quality.

Additionally, the methods of DNA isolation and quantification are equally important variables. Different researchers use different protocols for DNA isolation. These protocols include the use of magnetic bead, and commercial kits, as well as extraction with in-house organic solvents. All these processes recover DNA differently in regards to amounts and quality. For example, the kit-based extraction with silica membrane binding followed by elution can yield templates that vary widely in concentrations of up to 50 % [25]. An additional issue with these kits is the loss of informative small DNA fragments (<100 bp), suggested to harbor tumor mutations [26].

Differences in protocols for ccfDNA quantification have also been an issue with incomparable data. Various methods including PCR amplifications targeting different genes, simple spectrophotometry, and fluorometric methods with dyes such as SYBR green and PicoGreen have all been used for DNA content measurements. However, for total ccfDNA measurements, the sensitive fluorometric approach is recommended as this can detect DNA levels as low as 1 ng/ml.

Partly because of the above issues with ccfDNA analyses, circulating levels have varied considerably even in healthy individuals. They have ranged from as low as 6 ng/ml to as high as 650 ng/ml in healthy people. A meta-analysis of reported studies found the medium ccfDNA in healthy people to be between 2.5 and 27 ng/ml (mean 15 ng/ml) [27].

5.9.1 Recommendations for Pre-analytical Sample Handling for ccfDNA Analysis

In view of the above issues with ccfDNA analyses, El Messaoudi et al. have provided some directions on the pre-analytical sample handling to help improve

both quantitative and qualitative data on ccfDNA in cancer patients [28]. This is a summary of their observations and recommendations:

- There has always been an issue as to whether plasma or serum is the appropriate sample for circulating biomarker analysis. The answer depends on the type of biomarkers and the specific research question being addressed. In the field of ccfDNA quantification, plasma has been recommended as an accurate representative sample of tumor-derived DNA. This is because the clotting process is associated with the release of genomic DNA from blood formed elements and hence ccfDNA levels are much higher in serum than plasma.
- There are various chemicals available for preventing blood coagulation. However, ethylenediaminetetraacetic acid (EDTA) is the recommended anticoagulant compared to citrate and heparin because EDTA can stabilize ccfDNA for up to 6 h. Besides, heparin inhibits the PCR, which is generally used for downstream analysis.
- K3EDTA collection tubes are generally used for ccfDNA sample collection and are recommended. However, cell-free DNA™ blood collection tubes may equally be suitable.
- There is a possibility that ccfDNA levels may increase when stored at room temperature or 4 °C for >6 h after sample collection. Similarly, agitation or rough handling during sample collection can increase ccfDNA and nucleosomal DNA release. Moreover, long storage (>6 h) at room temperature or agitation causes DNA fragmentation that could affect data on DNA integrity index. Hence, gentle precaution must be taken during sample collection, and samples should be processed within 6 h following venipuncture.
- Another parameter that affects the quality of plasma for ccfDNA analysis is centrifugation procedure. Single blood spin is probably not optimal at removing all formed elements from plasma. Thus, centrifugation at 1200–1600× *g* for 10 min followed by a second spin at 16,000× *g* for another 10 min is recommended to generate true ccfDNA. Also optimal at removing all formed elements from plasma is single centrifugation followed by filtration with 0.2 μm pore filter.
- Plasma storage conditions also affect ccfDNA levels and integrity. Serum stored at 37 °C decreases nucleosomal levels after 6 h. Therefore it is recommended that samples are processed within 3 h of collection (or stored at 4 °C for a short time), or else they should be stored at −20 °C or −80 °C for later ccfDNA extraction.
- To prevent DNA fragmentation, samples should be aliquoted for storage and only two freeze-thaw cycles are permitted (if necessary).
- For accurate preservation of ccfDNA levels and integrity, extracted DNA can be faithfully analyzed immediately; otherwise, storage at −20 °C for 3 months is recommended but freeze-thaw cycles must be limited to no more than 3 times.
- Extracted DNA or plasma samples stored at −20 °C or −80 °C for 9 months is acceptable for qualitative analysis such as sequence variation or mutation detection.

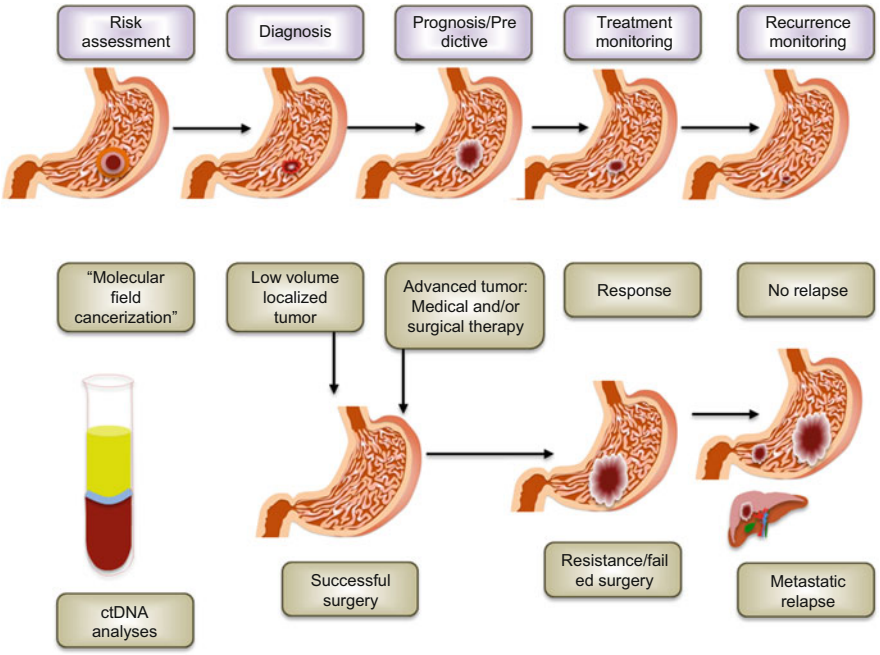

Fig. 5.4 The clinical applications of ctDNA. Circulating tumor nucleic acids are quickly becoming important in clinical management of cancer patients. All aspects of the cancer management cycle can benefit from ctDNA analysis. However, of outmost importance will be their use in tracking disease evolution with treatment and monitoring for early relapse

Circulating nucleosome and ccfDNA levels have been explored for use in cancer management including early detection, diagnosis, staging, prognostication, and treatment predictions, as well as monitoring of therapy response and recurrences (Fig. 5.4). Both quantitative (amounts) and qualitative (e.g., mutation detection) analysis of ccfNAs has been explored in cancer patients. Major issues with standardization of quantitative data have reflected on incomparable reported studies, and hence the slow pace of clinical translation. However, sensitive technologies for ctDNA detection and quantification have proven useful in disease management.

5.10 Translational Potential of Circulating Nucleosomes in Cancer

5.10.1 Diagnosis

In general, the levels of serum/plasma nucleosomes are much higher in cancer patients than healthy controls, indicating their potential use in early cancer

detection. But for this to be feasible, technologies have to be developed that detect cancer-specific nucleosomes. The equally high levels detected in many benign conditions hamper diagnostic utility for cancer. For example, in a subgroup of lung cancer patients, sensitivity and specificity were as high as 90 % and 95 %, respectively, when cancer patients were compared to healthy controls. However, the specificity dropped drastically to 35 % when benign lung conditions were compared [29, 30]. Circulating nucleosome detection may therefore be of limited diagnostic importance, but useful for disease prognosis and treatment response monitoring.

5.10.2 Prognosis

The prognostic utility of circulating nucleosomes is equally limited, but promising. The few studies indicate their potential use in cancer prognostications, at least in lung cancer patients. While high levels of circulating nucleosomes may have prognostic value, an in-depth study of NSCLC indicated their independent utility in prognostication only in univariate analysis. Multivariate scrutiny failed to validate their independent prognostic utility when other prognostic markers were included. However, before the second round of chemotherapy, circulating nucleosomes were of an independent prognostic importance even in multivariate analysis [31]. In SCLC patients, however, pretreatment and, prior to second-line chemotherapy, circulating nucleosomes were of prognostic utility in multivariate analysis [32].

5.10.3 Therapy and Disease Monitoring

Current treatment efficacies are evaluated by various methods including macroscopic reduction in tumor volume, which may take several days. Biomarkers that accurately measure tumor cell death as a consequence of treatment should enable early evaluation to avoid continuance of noneffective medication or adjustment to reduce excess cytotoxicity. Circulating nucleosomes may serve such functions.

Chemotherapy, radiotherapy, and immunotherapy are first associated with increased nucleosome levels in just 2–5 days of therapy, and then the levels suddenly decline. Initial increase is due to massive cell death caused by the therapeutic agent, but this is cleared up quickly. Depending on the strength, chemotherapy and radiotherapy cause apoptotic (weaker stimuli), or oncotic/aponecrotic cancer cell death with release of nucleosomes. Noteworthy, rapidly diving epithelial and hematopoietic cells are also killed and contribute to circulating nucleosome levels. However, most nucleosomes released during cancer treatment are likely from cancer cells. In colorectal cancer patients, for example, nucleosome levels were much higher compared to tumor-free individuals exposed

to radio-chemotherapy [33]. Similarly, in animal models, exposure to chemotherapy revealed changes in circulating nucleosomes only in animals with tumor but not in tumor-naïve animals [34].

Nucleosome levels in circulation of cancer patients are a balance between release and efficacy of clearance. Radiotherapy induces faster release because it causes more direct cellular damage. Chemotherapy is associated with maximal release seen at 24 h after start of treatment. As tumor size decreases due to chemotherapy, radiotherapy, or surgery, levels of circulating nucleosomes decrease toward baseline. These dynamic changes of circulating nucleosomes as a consequence of treatment have been demonstrated for breast, lung, colorectal, and pancreatic cancers as well as in hematologic malignancies.

Because the effect of therapy on circulating nucleosomes is very rapid, studies have been performed to evaluate the use of circulating nucleosomes for assessing treatment efficacy instead of conventional methods such as macroscopic evaluation of tumor volume and imaging. Low levels of circulating nucleosomes and high percentage reductions before and after the initial treatment correlate with therapy response in NSCLC. Nonresponders to therapy usually have very high initial levels that are inefficiently cleared from circulation possibly due to reduced activity of endonucleases in such patients.

5.11 Translational Potential of Circulating Cell-Free (Unbound) DNA

Aside from nucleosomal analyses for clinical applications, ccfDNA in general has been extensively studied in various neoplastic conditions. The focused here has been both quantitative DNA content analyses and examination of tumor-specific alterations as ctDNA. The latter analyses had been mostly qualitative, but can be used to quantify ctDNA fraction in ccfDNA.

5.11.1 Diagnosis

The diagnostic potential of ccfDNA content in a number on cancers including those of the lung, breast, and prostate has only been modest compared to conventional diagnostic biomarkers, although in general the area under the receiver operating characteristic curve (AUROC) has been >0.70. Diagnostic sensitivities have been limited, and specificities are hampered by confounding variables such as inflammation, exercise, and other nonmalignant diseases that are associated with increase release of DNA into the circulation. It is unlikely that all solid tumors release substantial amounts of DNA into the circulation. For example, ctDNA was detectable in >75 % of people with cancers of the ovaries, pancreas, colorectum, bladder,

stomach, esophagus, breast, liver, head and neck, and skin (melanoma), but less so (<50 %) in those with renal, brain, thyroid, or prostate cancers. Irrespective of the differences in the levels and frequencies of release, a general utility of circulating tumor-specific DNA in multiple cancers was uncovered by this study [35]. Thus, a meta-analysis including 17 studies of ccfDNA in lung cancer diagnosis only found the diagnostic performances to be inadequate as a stand-alone screening tool. The NLR of 28 %, for instance, was high. Factors such as lung inflammation caused by smoking will lead to elevated ccfDNA levels. Notwithstanding these shortcomings, the performance of ccfDNA levels in lung cancer detection was comparable to conventional biomarkers. It was therefore suggested to be used as an adjunct to other diagnostic biomarkers, possibly in a nomogram [36].

5.11.2 Prognosis

Consistent with its use in cancer diagnosis, a place for ccfDNA levels in prognostic predictions has been demonstrated. In general, higher ccfDNA levels have been an independent prognostic predictor of overall and disease-free survival in cancers such as those of the breast, lung, liver, prostate, and ovary. Expectedly, others have not been able to find any such prognostic associations in other cohorts of cancer patients. In addition to known confounding factors such as pre-analytical variables, the use of different cutoff levels for ccfDNA in prognostication has been problematic. Standardization is once again important for unlocking the prognostic utility of ccfDNA analysis in cancer patients.

5.11.3 Therapy and Disease Monitoring

The seminal work of Leon in 1977 suggested the value of ccfDNA in treatment response monitoring of cancer patients. These findings were later demonstrated in xenograft mouse models of human cancer. In nude mice with established ovarian cancer, the levels of ccfDNA increased, mirroring increasing tumor burden. Twenty-four hours following single-dose docetaxel chemotherapy, ctDNA increased 63 %, followed by a decline to 20 % below baseline in 72 h, and were 83 % lower than baseline at 10 days after therapy [37]. Rago et al. demonstrated similar finding a year later, with an initial therapy-induced increase in ccfDNA, followed by a decline [38]. These initial promising observations provided evidence for the use of ccfDNA levels as predictive and response monitoring of cancer. As an example in clinical studies, successful surgical treatments of breast, colon, esophagus, kidney, and lung cancers were associated with decreases in ccfDNA levels below presurgical levels. The lack of decline was suggestive of incomplete resection or the presence of systemic disease. Also a decline followed by a rise was indicative of disease relapse. Similar findings have been observed in cancer patients

on chemotherapy and radiotherapy. As with all ccfDNA analysis, others have failed to demonstrate such fluctuations [39].

5.12 Translational Potential of DNA Integrity Index in Cancer

Cancer cells are resilient to apoptotic cell death (except when therapy induced) and hence frequently die by necrosis. Apoptosis is associated with DNA fragmentation that yields smaller-sized DNA than necrosis. Therefore, assays measuring the ratios of larger to smaller DNA fragments in circulation (referred to by Wang et al. [40] as DNA integrity index, DII) have found increased values in cancer patients. Other investigators studying similar phenomenon have used the inverse ratio, named the apoptotic index. In the pioneering study by Wang et al., DNA concentration and integrity were measured in plasma from women with multiple cancers including breast, ovarian, endometrial, and cervical cancers, as well as peritoneal serous carcinomas compared with age-matched women with various gynecologic and other non-neoplastic conditions as well as healthy women [40]. They targeted 100 bp and 400 bp DNA fragments of *ACTB*, and DII proved to be very accurate in this series for discriminating cancer from non-neoplastic conditions with AUROC of 0.911, at a sensitivity of 62 % and a specificity of 100 %. Thereafter, several others have used ratios of DNA lengths between 201 bp to 618 bp (longer fragment) and 100 bp to 200 bp (shorter fragment) as a measure of DII to characterize the sources of ccfDNA. While some studies target structural genes including housekeeping genes, analyses of DII have also focused on examination of noncoding nuclear DNA (ALU/SINE and LINE1) repeat sequences.

Probably as inherent in the lack of analytical standardization and tumor heterogeneity, and pathology, the discriminatory ability of DII for various cancers has been contradictory. Aside from the initial study that revealed their utility in breast and gynecological malignancies, others have found its diagnostic potential for cancers of the head and neck (including nasopharyngeal carcinoma), esophagus, colon, kidneys, and prostate. But other studies of lung, bladder, prostate, and testicular cancers have failed to demonstrate a diagnostic utility. Similarly, the prognostic and predictive value of DII has shown promise in nasopharyngeal, breast, bladder, and prostate cancer. Standardized protocols used in multicenter studies should help resolve these discrepancies.

5.13 Translational Potential of Circulating Tumor-Specific DNA

Clinically, ctDNA has been used to explore epigenome and genome alterations in cancer patients for various possible applications (Fig. 5.4). With sensitive technologies, it has been possible to detect ctDNA even in asymptomatic individuals with very low ctDNA percentage among ccfDNA (even <1 % ctDNA fraction). Besides, ctDNA has great potential for cancer patient management compared to the use of traditional approaches including tumor biopsy and imaging.

5.13.1 Circulating Tumor DNA Detection Methods

Various approaches have been developed for assaying cancer mutations and epigenetic alterations in ccfDNA. Some approaches such as whole genome and exome sequencing do not require prior knowledge of the genetic changes in the primary tumor and hence can detect novel mutations as they evolve with therapy or metastasis. These untargeted approaches, although less sensitive analytically, play important role in tumor biology and management due to intratumor heterogeneity and tumor evolution with time. There are also newer, very sensitive, targeted approaches that enable the detection of very low levels of ctDNA. Currently, methods at detecting ctDNA target the epigenome and genome of the cancer cell.

There are several methods for examining the genomic pattern and content of methylation, as well as specific gene promoter methylations in ccfDNA. These include methylation-specific PCR (MSP), methylation CpG-island amplification (MCA), amplification of intermethylated sites (AIMS), restriction landmark genomic scanning (RLGS), bisulfite sequencing, methylation-sensitive restriction endonuclease PCR (MSRE-PCR), MethyLight, HPLC, and HPLE. However, the most commonly used method for methylation analyses of ctDNA has been MSP. Additionally, cancer-type specific methylation platforms such as the microarray-based multiplex assay (MethDet56) for ccfDNA methylation analysis in ovarian cancer are useful [41]. The various studies of methylation alterations in ctDNA have shown high specificity (mostly >90 %) for various cancers. However, sensitivity has been dismal, demonstrating slight improvements in panel studies.

Considerable work has focused on detection of cancer genome alterations in ctDNA. Several cutting-edge untargeted technologies have all proven useful, though with different strengths and weaknesses, in ctDNA analyses. Array-CGH analyses of plasma enabled tumor-specific copy number variation (CNV) detection [42]. Similarly, next-generation sequencing efforts enabled increased resolution of CNV, which have equally been identified in ctDNA using bisulfite DNA sequencing [43] and MiSeq Illumina platform [44]. The Affymetrix SNP 6.0 array for CNV and LOH analyses enabled focal identification of high-level DNA amplifications in ctDNA from breast cancer patients [45]. Leary's massively parallel sequencing

(MPS) of whole genome coupled with the development of personalized analysis of rearranged ends (PARE) also enabled the detection of translocations, rearrangements, and amplifications of *ERBB2* and *CDK6* in colorectal and breast cancer patients [46]. Because of high cost, many of these sequencing techniques have involved low depth coverage, but have still proven useful in CNV analyses of ctDNA. For the detection and management of early cancers, and those with low ctDNA fractions, high depth coverage is warranted. Murtaza et al. approached this issue by performing exome sequencing of plasma samples from cancer patients [47]. Although cost was reduced, this effort enabled the detection of resistant mutations arising de novo or enriched for due to therapy in ctDNA. As sequencing technologies become economical, ctDNA analyses will be clinically applicable.

Other technologies for targeted ctDNA mutation analyses with enhanced analytical sensitivities (can detect mutations in ctDNA fractions <1 %) include digital PCR [48, 49]; beads, emulsions, amplification, and magnetics (BEAMing) [26]; amplification refractory mutation system (ARMS) [50]; tagged-amplicon deep sequencing (TAm-Seq) [51]; cancer personalized profiling by deep sequencing (CAPP-Seq) [52]; and safe-sequencing system (Safe-SeqS) [35]. The very sensitive Safe-SeqS technology could detect one mutant template among DNA from 5 ml of blood. Similarly, CAPP-Seq could detect ctDNA as low as 0.02 %, at 96 % specificity for detected mutations. This level of resolution enabled ctDNA in all stage II–IV and 50 % of stage I NSCLC patients to be detected [52]. However, an issue with these approaches is the targeting of only a few genomic regions, supposedly hotspots of cancer driver mutations. But even passenger mutations can be informative biomarkers. To increase coverage, Forshew et al. developed the TAm-Seq, which includes 5995 bases to be interrogated at relatively high depth [51]. Indeed, allele frequencies as low as 2 % could be detected for all cancer mutations, but resolution was even higher at 0.2 % when considering a limited set of known cancer mutations.

5.13.2 Early Detection

Because, in general, it takes 6–10 mutations for cellular transformation and, more so, several years for the appearance of overt tumors, healthy-looking individuals could harbor circulating tumor DNA without any symptoms. This thesis has been demonstrated in some proof-of-principle studies. In a population-wide screening of 1098 volunteers, *KRAS* mutations were detected using restriction-enriched PCR in circulating DNA from 13 individuals. Within 25 months, five and one developed bladder and head and neck cancers, respectively [53]. In another screening study of 1318 volunteers, EBV (known risk for NPC) DNA was present in 69 asymptomatic individuals. Subsequently, 3 were diagnosed with NPC [54]. There are a number of issues with this type of screening. First, unless analytical costs are down (possibly with development of point-of-care diagnostics), this may not be practical for population applications. Second, while early detection should translate into curative

interventions and save lives, the issues of overdiagnosis and overtreatments need to be considered. Third, the psychological effects of early detection of genetic defects must be appropriately managed. Individuals testing positive should be provided with genetic counseling to understand the meanings and implications of their results.

Aside from population screening, the early detection potential of ctDNA to complement clinical diagnosis has been demonstrated. In contrast to the mostly published studies in patients with advanced-stage disease, whereby there is increased release of tumor DNA into circulation, such that mutant copies could be as high as 50 % of wild-type DNA, ctDNA has been demonstrated in patients with localized low-volume tumors. A few proof-of-principle studies indicate the presence of abundant ctDNA in early-stage cancer patients. In 12 tumor types, ctDNA was detected in 55 % of plasma samples. Indeed mutant DNA copies were up to 135,000 per 5 ml of blood [35]. In breast cancer patients with early low-volume cancer, *PIK3CA* mutations were detected in 93.3 % of preoperative plasma samples. Postoperative monitoring of plasma mutations may be useful for monitoring completion of tumor resection or early disease recurrence [55]. Diehl et al. had revealed detectable ctDNA in patients with localized colorectal cancer and even in those with adenomas. In these patients, mutant copies were very low, with mutant to wild-type allele frequencies ranging from 0.001 to 0.012 % [56].

5.13.3 Prognosis

Because some cancer genetic alterations confer outcomes, ctDNA analyses have shown potential for disease prognostic predictions. These are examined in detail under the specific cancers. However, a few examples here suffice an illustration of their importance in clinical oncology. In colorectal cancer patients with no evidence of *KRAS* hotspot mutations or *CDKN2A* promoter hypermethylation in ctDNA, the 2-year survival rate was 100 % [57]. The presence of high levels of ccfDNA content and *KRAS* mutant ctDNA conferred dismal outcome for patients with metastatic colorectal cancer [50]. Similarly, after surgery, the presence of ctDNA predicted relapse [26]. It has even been demonstrated that ctDNA analyses may be superior to CTC enumeration as prognostic biomarkers [58, 59].

5.13.4 Treatment Prediction

Circulating tumor DNA, as it may represent all tumor cells or clones, has been explored for longitudinal monitoring of patients with advanced cancers on different therapeutic regimens. Thus, a number of investigations demonstrate the ability to track disease evolution over time. Proof-of-principle studies have been conducted in a number of solid tumors including colorectal and lung cancers. The emergence

of new tumor clones with novel *KRAS* mutations have been demonstrated in colorectal cancer patients. Analyses of pre- and post-therapy plasma samples revealed novel mutations in MAPK pathway genes in 95.8 % colorectal cancer patients on anti-EGFR inhibitors. Fifty percent of these mutations involved *KRAS* codon 12 [35]. Using BEAMing technology, *KRAS* mutations and amplifications were uncovered in ctDNA from 60 % of CRC patients with disease progression despite cetuximab-targeted therapy. These drug-resistant mutations, which were also found in metastatic lesions, occurred several months prior to radiographic detection of disease progression [60]. Moreover, in patients with chemorefractory tumors on panitumumab therapy, 38 % of them with initial wild-type *KRAS* were found to harbor *KRAS* mutations 6 months later in ctDNA [61]. Another evidence of ctDNA analysis for early detection of evolving mutant clones was demonstrated in NSCLC patients on anti-EGFR (erlotinib) therapy. While the resistant mutation (T790M) was undetectable in all patient plasma samples prior to treatment, in 6/9 patients, this mutation levels increased in ctDNA 4–24 weeks before detection of progression by response evaluation criteria in solid tumors (RECIST) [62].

5.14 Digital PCR Technology

Because ctDNA can be diluted in body fluids, especially in circulation of early disease patients, technologies that enable sensitive analytical detection are warranted. One such technology that has recently been optimized and extensively used in biomarker analyses in body fluids is digital PCR. The concept of being able to determine the copy number of a desired target in a sample was initially reported in 1992 by Sykes and colleagues [63]. Using limiting dilution and Poisson statistics, the two rearranged immunoglobulin heavy chain (IgH) genes from leukemic cells could be quantified in the presence of 160,000 copies of IgH from normal lymphocytes. This method simply enables quantification of total number of initial targets in a sample rather than PCR products at the end of the reaction. Seven years later, Vogelstein and Kinzler optimized and named this method dPCR and used it to quantify colorectal cancer mutations in stool samples [49].

In principle, dPCR involves sample dilution and subsequent partitioning into hundreds or even millions of separate reaction wells (Fig. 5.5). Each of these wells contains a copy or none of the target of interest. Following PCR amplification, there will be positive (presence of target molecule) and negative (absence of target molecule) chambers. The counting of these wells enables the number of target copies in the initial sample to be exactly determined. Absolute quantification by dPCR has been advanced using nanofabrication and microfluidic technologies. There are commercial platforms that are able to generate millions of partitions each with contents as low as picoliter amounts. Reagents can even be carried in separate droplets (droplet dPCR).

Digital PCR has a number of advantages and uses. Digital PCR can quantify low copy number targets, even among complex mixtures in body fluids without the need

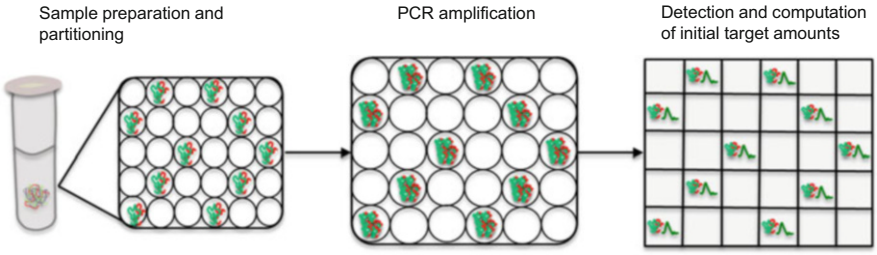

Fig. 5.5 Digital PCR technology. This is a powerful tool for ctDNA analysis. This method simply involves sample dilution, partitioning, PCR amplification, and data readout and analysis. Knowledge from amplified and unamplified wells enables determination of the quantities of the target in the initial sample

for reference standards. Additionally, it can detect small fold changes and easily overcome PCR inhibitors. The uses of this technology encompass several genetic analyses. Some of the uses in cancer biomarker studies are:

- Detection and quantification of rare cancer mutations. This is especially relevant in analysis of ctDNA where mutation load is low.
- Detection and quantification of copy number variations.
- Differential gene expression analysis.
- Validation of next-generation sequencing libraries and data.

5.15 BEAMing Technology

A variation of the digital PCR technology, BEAMing, uses emulsion PCR with magnetic bead capture followed by flow cytometry for sensitive detection and quantification of mutant alleles. Reported by Diehl et al., this technology has been considerably advanced for sensitive detection of tumor mutations in ctDNA [64]. It has an analytical sensitivity of ~0.01, indicating the ability to detect 1 mutant copy among 10,000 wild-type genomes in a sample. The principles of BEAMing technology include the following (Fig. 5.6):

- Isolated ccfDNA is pre-amplified using conventional PCR with primers that contain tag sequences for the specific mutation.
- Pre-amplified DNA is mixed with beads coated with primers (via streptavidin-biotin) to capture the tagged sequences.
- Sample partitioning and emulsion droplet digital PCR is then performed. The emulsion droplets contain PCR reagents and the second primer for PCR. Following amplification, each positive emulsion droplet contains beads coated with 1000s of copies of the initial *single* DNA molecule in the droplet (after dilution and partitioning).

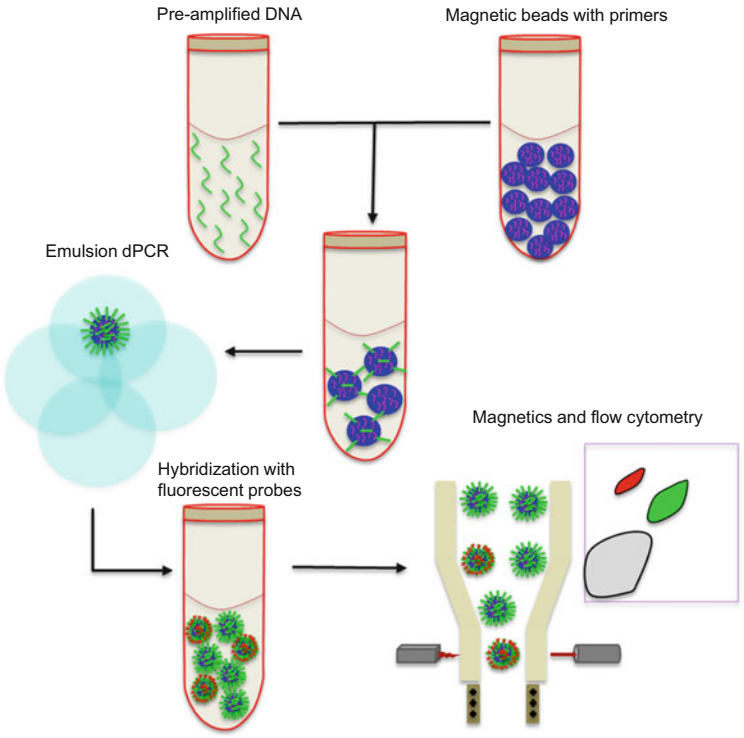

Fig. 5.6 BEAMing digital PCR technology. Mutant allele in ctDNA is pre-amplified, captured with primers coated on magnetic beads, and subjected to emulsion droplet dPCR. Fluorescent labeling followed by flow cytometry enables the detection of mutations that can be quantified

- Targeted fluorescent probes for mutant alleles are then hybridized to amplified DNA on beads.
- Finally, magnetics, isolation, and flow cytometry are used for detection of mutant copies that can then be quantified.

This technology enables the noninvasive detection and quantification of tumor genotype in real time, and hence the readout of potentially driver or actionable mutations for many clinical applications.

5.16 Summary

- Cancer and normal cells release nucleic acids into the circulation and other body fluids. This circulating tumor DNA (ctDNA) has a half-life of about 2 h.
- The nucleic acids circulate in various forms as unbound free forms, attached to cell surfaces, complexed with histone (nucleosomes) or nonhistone proteins, associated with lipids, or enclosed in extracellular vesicles.

- The levels of circulating nucleic acids are much higher in cancer patients than healthy individuals. Therefore, they have been assayed as potential diagnostic and prognostic biomarkers of cancer.
- Due to pre-analytical and analytical variables, quantitative data on circulating nucleic acids have been incomparable between studies, which have been a limitation in their validation for clinical applications.
- However, detection and quantitation of ctDNA hold tremendous potential for cancer detection and management. Circulating tumor epigenome changes and various types of mutations form part of personalized oncology.
- Sensitive technologies such as digital PCR and BEAMing are paving the way for the clinical application of ctDNA.

References

1. Mandel P, Metais P. C R Seances Soc Biol Fil. 1948;142:241–3.
2. Bendich A, Wilczok T, Borenfreund E. Circulating DNA as a possible factor in oncogenesis. Science. 1965;148:374–6.
3. Tan EM, Schur PH, Carr RI, Kunkel HG. Deoxybonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. J Clin Invest. 1966;45:1732–40.
4. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37:646–50.
5. Stroun M, Anker P, Lyautey J, et al. Isolation and characterization of DNA from the plasma of cancer patients. Eur J Cancer Clin Oncol. 1987;23:707–12.
6. Stroun M, Anker P, Maurice P, et al. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology. 1989;46:318–22.
7. Sorenson GD, Pribish DM, Valone FH, et al. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev. 1994;3:67–71.
8. Vasioukhin V, Anker P, Maurice P, et al. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. Br J Haematol. 1994;86:774–9.
9. Shaw JA, Smith BM, Walsh T, et al. Microsatellite alterations plasma DNA of primary breast cancer patients. Clin Cancer Res. 2000;6:1119–24.
10. Wang JY, Hsieh JS, Chang MY, et al. Molecular detection of APC, K- ras, and p53 mutations in the serum of colorectal cancer patients as circulating biomarkers. World J Surg. 2004;28:721–6.
11. Fujiwara K, Fujimoto N, Tabata M, et al. Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res. 2005;11:1219–25.
12. Garcia-Olmo DC, Samos J, Picazo MG, et al. Release of cell-free DNA into the bloodstream leads to high levels of non-tumor plasma DNA during tumor progression in rats. Cancer Lett. 2008;272:133–40.
13. Wimberger P, Roth C, Pantel K, et al. Impact of platinum-based chemotherapy on circulating nucleic acid levels, protease activities in blood and disseminated tumor cells in bone marrow of ovarian cancer patients. Int J Cancer. 2011;128:2572–80.
14. Roth C, Pantel K, Muller V, et al. Apoptosis-related deregulation of proteolytic activities and high serum levels of circulating nucleosomes and DNA in blood correlate with breast cancer progression. BMC Cancer. 2011;11:4.

15. Roth C, Kasimir-Bauer S, Pantel K, Schwarzenbach H. Screening for circulating nucleic acids and caspase activity in the peripheral blood as potential diagnostic tools in lung cancer. Mol Oncol. 2011;5:281–91.
16. Garcia-Olmo D, Garcia-Olmo DC. Functionality of circulating DNA: the hypothesis of genometastasis. Ann N Y Acad Sci. 2001;945:265–75.
17. Garcia-Olmo DC, Ruiz-Piqueras R, Garcia-Olmo D. Circulating nucleic acids in plasma and serum (CNAPS) and its relation to stem cells and cancer metastasis: state of the issue. Histol Histopathol. 2004;19:575–83.
18. Bergsmedh A, Szeles A, Henriksson M, et al. Horizontal transfer of oncogenes by uptake of apoptotic bodies. Proc Natl Acad Sci USA. 2001;98:6407–11.
19. Skvortsova TE, Vlassov VV, Laktionov PP. Binding and penetration of methylated DNA into primary and transformed human cells. Ann N Y Acad Sci. 2008;1137:36–40.
20. Garcia-Olmo DC, Dominguez C, Garcia-Arranz M, et al. Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells. Cancer Res. 2010;70:560–7.
21. Chen Z, Fadiel A, Naftolin F, et al. Circulation DNA: biological implications for cancer metastasis and immunology. Med Hypotheses. 2005;65:956–61.
22. Le Lann-Terrisse AD, Fournie GJ, Benoist H. Nucleosome-dependent escape of tumor cells from natural-killer-mediated lysis: nucleosomes are taken up by target cells and act at a postconjugation level. Cancer Immunol Immunother. 1997;43:337–44.
23. Tanner JE. Nucleosomes activate NF-kappaB in endothelial cells for induction of the proangiogenic cytokine IL-8. Int J Cancer. 2004;112:155–60.
24. Leung SF, Zee B, Ma BB, et al. Plasma Epstein-Barr viral deoxyribonucleic acid quantitation complements tumor-node-metastasis staging prognostication in nasopharyngeal carcinoma. J Clin Oncol. 2006;24:5414–8.
25. de Kok JB, Hendriks JC, van Solinge WW, et al. Use of real-time quantitative PCR to compare DNA isolation methods. Clin Chem. 1998;44:2201–4.
26. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14:985–90.
27. van der Vaart M, Pretorius PJ. Is the role of circulating DNA as a biomarker of cancer being prematurely overrated? Clin Biochem. 2010;43:26–36.
28. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: preanalytical considerations. Clin Chim Acta. 2013;424:222–30.
29. Holdenrieder S, Stieber P, Bodenmuller H, et al. Nucleosomes in serum of patients with benign and malignant diseases. Int J Cancer. 2001;95:114–20.
30. Holdenrieder S, Stieber P, von Pawel J, et al. Circulating nucleosomes predict the response to chemotherapy in patients with advanced non-small cell lung cancer. Clin Cancer Res. 2004;10:5981–7.
31. Holdenrieder S, Stieber P. Therapy control in oncology by circulating nucleosomes. Ann N Y Acad Sci. 2004;1022:211–6.
32. Holdenrieder S, Nagel D, Schalhorn A, et al. Clinical relevance of circulating nucleosomes in cancer. Ann N Y Acad Sci. 2008;1137:180–9.
33. Kremer A, Holdenrieder S, Stieber P, et al. Nucleosomes in colorectal cancer patients during radiochemotherapy. Tumour Biol. 2006;27:235–42.
34. Trejo-Becerril C, Perez-Cardenas E, Trevino-Cuevas H, et al. Circulating nucleosomes and response to chemotherapy: an in vitro, in vivo and clinical study on cervical cancer patients. Int J Cancer. 2003;104:663–8.
35. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6:224ra224.
36. Zhang R, Shao F, Wu X, Ying K. Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis. Lung Cancer. 2010;69:225–31.
37. Kamat AA, Bischoff FZ, Dang D, et al. Circulating cell-free DNA: a novel biomarker for response to therapy in ovarian carcinoma. Cancer Biol Ther. 2006;5:1369–74.

38. Rago C, Huso DL, Diehl F, et al. Serial assessment of human tumor burdens in mice by the analysis of circulating DNA. Cancer Res. 2007;67:9364–70.
39. Jung K, Fleischhacker M, Rabien A. Cell-free DNA in the blood as a solid tumor biomarker--a critical appraisal of the literature. Clin Chim Acta. 2010;411:1611–24.
40. Wang BG, Huang HY, Chen YC, et al. Increased plasma DNA integrity in cancer patients. Cancer Res. 2003;63:3966–8.
41. Melnikov A, Scholtens D, Godwin A, Levenson V. Differential methylation profile of ovarian cancer in tissues and plasma. J Mol Diagn. 2009;11:60–5.
42. Heitzer E, Auer M, Hoffmann EM, et al. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. Int J Cancer. 2013;133:346–56.
43. Chan KC, Jiang P, Chan CW, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci USA. 2013;110:18761–8.
44. Heitzer E, Ulz P, Belic J, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Med. 2013;5:30.
45. Shaw JA, Page K, Blighe K, et al. Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. Genome Res. 2012;22:220–31.
46. Leary RJ, Sausen M, Kinde I, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med. 2012;4:162ra154.
47. Murtaza M, Dawson SJ, Tsui DW, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013;497:108–12.
48. Taly V, Pekin D, Benhaim L, et al. Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. Clin Chem. 2013;59:1722–31.
49. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci USA. 1999;96:9236–41.
50. Spindler KL, Pallisgaard N, Vogelius I, Jakobsen A. Quantitative cell-free DNA, KRAS, and BRAF mutations in plasma from patients with metastatic colorectal cancer during treatment with cetuximab and irinotecan. Clin Cancer Res. 2012;18:1177–85.
51. Forshew T, Murtaza M, Parkinson C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4:136ra168.
52. Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20:548–54.
53. Gormally E, Vineis P, Matullo G, et al. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. Cancer Res. 2006;66:6871–6.
54. Chan KC, Hung EC, Woo JK, et al. Early detection of nasopharyngeal carcinoma by plasma Epstein-Barr virus DNA analysis in a surveillance program. Cancer. 2013;119:1838–44.
55. Beaver JA, Jelovac D, Balukrishna S, et al. Detection of cancer DNA in plasma of patients with early-stage breast cancer. Clin Cancer Res. 2014;20:2643–50.
56. Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci USA. 2005;102:16368–73.
57. Lecomte T, Berger A, Zinzindohoue F, et al. Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis. Int J Cancer. 2002;100:542–8.
58. Bidard FC, Madic J, Mariani P, et al. Detection rate and prognostic value of circulating tumor cells and circulating tumor DNA in metastatic uveal melanoma. Int J Cancer. 2014;134:1207–13.
59. Punnoose EA, Atwal S, Liu W, et al. Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. Clin Cancer Res. 2012;18:2391–401.
60. Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012;486:532–6.

61. Diaz Jr LA, Sausen M, Fisher GA, Velculescu VE. Insights into therapeutic resistance from whole-genome analyses of circulating tumor DNA. Oncotarget. 2013;4:1856–7.
62. Oxnard GR, Paweletz CP, Kuang Y, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin Cancer Res. 2014;20:1698–705.
63. Sykes PJ, Neoh SH, Brisco MJ, et al. Quantitation of targets for PCR by use of limiting dilution. Biotechniques. 1992;13:444–9.
64. Diehl F, Li M, He Y, et al. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. Nat Methods. 2006;3:551–9.

Chapter 6
Noncoding Transcripts as Cancer Biomarkers in Circulation

Key Topics

- Biology of noncoding RNA (ncRNA)
- MiRNA and cancer
- Genomics of miRNA
- MiRNA biogenesis and its defects in cancer
- MiRNA release, stability in body fluids, and uptake
- Oncogene and tumor suppressor miRNAs
- Precautions in miRNA analysis
- Long noncoding RNA (lncRNA)
- Structure and classification of lncRNAs
- Functions of lncRNA
- LncRNA aberrant expressions and mutations in cancer
- LncRNA in body fluids of cancer patients

Key Points

- An important component of genomic output has hitherto been referred to as "dark matter." Expanding knowledge now assigns critical gene regulatory functions to this "dark matter."
- Both short and long noncoding RNAs control gene expression and are themselves under regulation by various gene networks. The complex interplay between coding and noncoding transcripts drives the noted hallmarks of cancer.
- Noncoding RNAs are stable in body fluids and serve as attractive noninvasive cancer biomarkers.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_6

6.1 Introduction

With the completion and publication of the human genome reference sequence, it immediately became clear that what had been known through genetic and evolutionary studies about the coding patterns of the human genome are true. The majority of the human genome has been demonstrated to be noncoding. Indeed, ~99 % of the ~3.3 billion nucleotides of the human genome is non-protein coding. Thus, only a minuscule amount (1.5 %) of the genome constitutes protein-coding sequences.

The vast majority of the noncoding nucleotides are made up of repeat sequences. Actually, up to 50 % of the nucleotides in the genome are repeat sequences. Additionally, the noncoding sequences are binding sites for transcription factors and other gene regulatory elements, as well as sources of noncoding transcripts. Biochemical and genome-wide association studies suggest >75 % of the genome reveal some functional transcriptional activity. This finding suggests that outside the 1.5 % protein-coding genes, the majority of the output from the human genome is noncoding transcripts. Thus, about 75 % of different RNA molecules can be detected in the genome, though most are of low abundance.

Noncoding genes are by definition genes that give rise to transcripts, which are not translated into proteins, due partly to poor or absent open reading frames. Our expanded knowledge now conclusively proves that despite the lack of protein production from these genomic regions, the noncoding RNAs (ncRNAs) have profound functions in gene regulation. Indeed, miRNA alone controls ~50 % of mammalian protein-coding genes.

A simplistic classification dichotomizes these ncRNAs into small/short and long ncRNAs, using the criteria as to whether they are <200 nucleotides (nts) (small/short) or >200 nts (long) in length. The class of small ncRNAs includes miRNA, piRNA, tiRNA, tRNA, rRNA, snoRNA, and snRNA. In regards to diseases, especially cancer, miRNA has been the most extensively studied small ncRNA. Knowledge on miRNA biology and as biomarkers of disease diagnosis, prognosis, and management and as targets and vehicles of therapeutic intervention is well elucidated, and some inroad into the clinic is being achieved. Their resilience to degradation and presence in protected exosomes and other extracellular vesicles make them attractive biomarker targets in body fluids. While the other small ncRNAs (e.g., snRNA, snoRNA) may be important in cancer biology, their roles are yet to be appreciated. Similarly, although more numerous than miRNA, the importance of lncRNA in cancer biology has lagged behind the attention given to miRNA. However, recent advances unravel some biologic functions of importance, and the biomarker potential of these low abundant lncRNAs in diseases including cancer is expanding. This chapter explores the biology and biomarker utility of ncRNA (primarily miRNAs and lncRNAs) and, importantly, their noninvasive measurements in body fluids for cancer management.

6.2 "Dark Matter" RNA

With technological advancement in the past decade, the genome has been interrogated at much greater detail than ever. Massive parallel sequencing and tilling arrays are able to uncover expressed unannotated genomic regions. Thus, the functions of transcripts from these regions have hitherto been largely unknown. These unannotated RNAs and whose functions were illusive to us had been referred to as "dark matter" RNA. Included in dark matter RNAs are noncoding RNA, which were initially thought to constitute a minor part of the cell's total RNA. However, work by Kapranov et al. [1], which involved total nuclear RNA sequencing, suggests "dark matter" RNA forms as much as 50–65 % of the entire repertoire of non-ribosomal and non-mitochondrial RNA by mass and that these RNAs play important roles in neoplasia. The full elucidation and demonstration of the functions of all the cellular (nuclear and mitochondrial) noncoding RNA in cancer is awaited.

6.3 Biology of Noncoding Transcripts

Transcripts from genes with no open reading frames and hence no protein-coding capabilities are referred to as ncRNAs. Though about 90 % of the human genome is transcribed, a mere ~1.5 % of these transcripts possess protein-coding structural motifs (i.e., have mRNA sequences). Thus, most of the human genome gives rise to ncRNA molecules, which are now recognized to have important regulatory functions. They function in normal development and cell physiology, and abnormal expression is observed in cancer and other pathologic conditions. There are two sources of noncoding transcripts in the human genome: nuclear and mitochondria.

6.3.1 Noncoding Nuclear Transcripts

There are two categories of nuclear ncRNAs based on transcript sizes. These are small or short and long ncRNAs (Fig. 6.1). Many small ncRNAs are usually ~18–22 nt long. Included in this class of small ncRNAs are well-known transcripts such as tRNAs, small nucleolar RNAs (snoRNA), small nuclear RNAs (snRNA), small interfering RNAs (siRNAs), and miRNAs. Long ncRNAs range in sizes from ~200 nt to even 100 kb. They are mRNA-like molecules without or with poorly defined open reading frames. Unlike small ncRNAs, long ncRNAs are exclusively intergenic. Included in the long ncRNAs are 18SrRNA, 28SrRNA, and transcribed pseudogenes. There are over 1000 known miRNAs that control >50 % of protein-coding genes. In contrast, 7000–23,000 long ncRNAs are found in the human

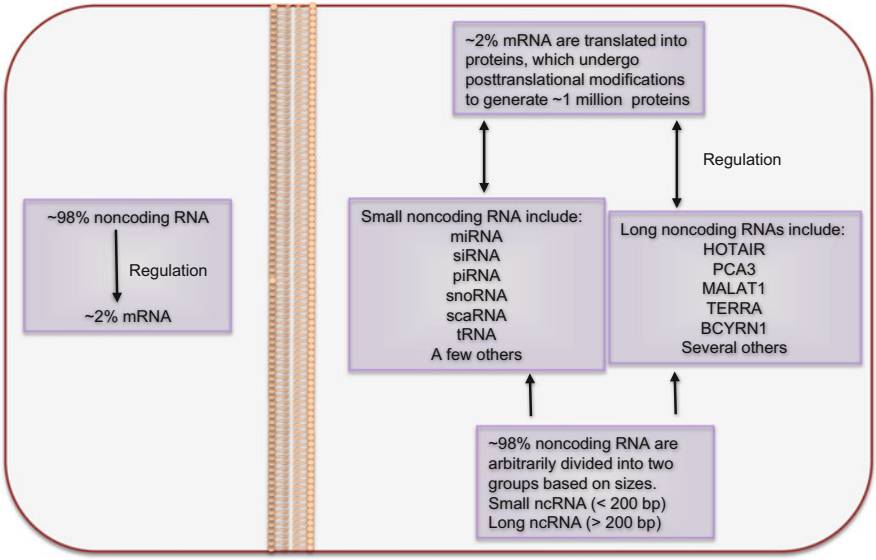

Fig. 6.1 Types of nuclear noncoding RNAs. Much of the genome does not contain protein-coding genes. These noncoding transcripts fall into two categories, short and long ncRNAs. Both classes exert tremendous control on protein-coding genes and hence cellular functions

genome. The emergence of the roles of these molecules in cancer biology is noteworthy and is detailed here.

6.3.2 Noncoding Mitochondrial Transcripts

The mitochondrial genome (mtgenome) encodes noncoding transcripts of biologic importance. Currently, a number of groups have demonstrated the presence of noncoding mitochondrial RNAs (ncmtRNA). Initially reported in the mouse, the human equivalent of the ncmtRNA was uncovered in 2007 [2]. This molecule is composed of an inverted repeat sequence that is covalently linked to the 5′ end of the mitochondrial 16SrRNA sense transcript. The human long noncoding RNAs are ~2374 nt long that is organized into a stem-loop structure composed of ~820 bp double-stranded stem and a 40 nt loop (Fig. 6.2a). The stem is resistant to RNase A digestion. Putatively, this transcript forms complexes with cytochrome c and endonuclease G in mitochondria, and its expression is associated with proliferating cells. Subsequent analyses indicated proliferating cells express both sense and antisense transcripts [3]. The antisense ncmtRNAs are two different inverted repeats linked to the 5′ ends of the antisense 16SrRNA transcribed from the L-strand.

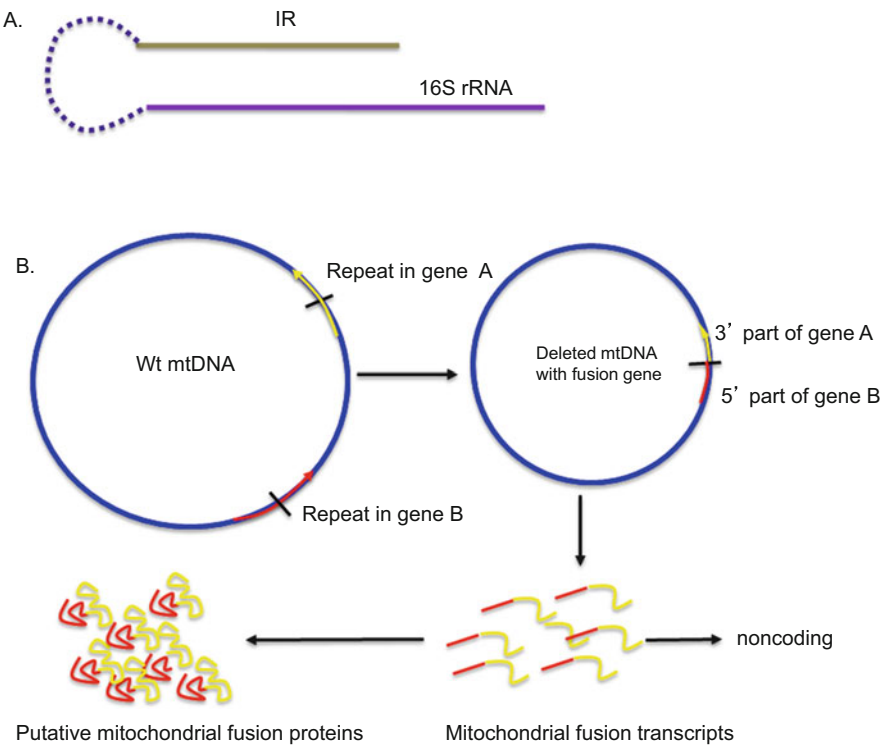

Fig. 6.2 Examples of noncoding mitochondrial genes. (**a**) Construct of the human ncmtRNA of 2374 nucleotides. The inverted repeat (IR) is linked to the 5′ end of 16SrRNA transcript. (**b**) Large-scale mitochondrial DNA deletion. The wild-type mtgenome is cleaved at repeats in genes A and B to produce the monomeric molecule that now contains a novel gene from the remaining parts of genes A and B. Transcripts from the novel fusion gene may be coding or noncoding

The biologic importance of these novel molecules has been demonstrated in a number of follow-up studies. Normal proliferating cells express both transcripts (sense and antisense). However, analyses of 15 different cancer cell lines and 273 cancer biopsies from 17 different cancers indicate the complete absence of the antisense transcript, with expression of only the sense ncmtRNA. This suggests the downregulated expression of the antisense transcript is a tumor-specific biomarker [4]. It has also been shown in human renal cells that these molecules translocate into the nucleus and associate with heterochromatin. In cancer cells, as expected, only the sense molecules form complexes with heterochromatin [5]. In a pilot study of urinary samples, both transcripts were not detected in exfoliated cells in urine from healthy individuals. However, urine samples from bladder cancer patients expressed the sense transcript in association with loss of the antisense noncoding transcripts, and this aberrant expression was able to discriminate between samples from cancer patients and healthy control subjects [6]. The decreased expression of the antisense molecule appears to be a hallmark of cancer.

Indeed, it is demonstrated that high-risk human papillomavirus oncoproteins modulate the expression of these ncmtRNAs. Human papillomavirus E2 oncoprotein represses the expression of the antisense transcript, while E6 and E7 induce the expression of yet another sense transcript referred to as SncmtRNA-2. This novel transcript shares 3′ sequence homology with SncmtRNA-1, but has a different 5′ sequence structure.

The mtgenome also encodes lncRNAs. Rackham et al. [7] uncovered three novel lncRNAs of mitochondrial origin by close analysis of deep sequencing data. Their presence was authenticated using Northern blotting and qRT-PCR approaches. These molecules are mainly encoded by the light strand of the mtgenome and are complementary to *MTND5*, *MTND6*, and *MTCYTB*. The expression levels of these lncRNAs were at 58 %, 34 %, and 14 %, respectively, to their traditionally known heavy-strand coding complements. These molecules are also partners of the nuclear-mitochondrial communication network, because nuclear encoded genes that regulate RNA processing are also involved in their processing. Mitochondrial RNase P protein 1 (MRPP1) has been identified to play a role in their processing as well. These molecules form intermolecular duplexes that resist RNase 1 digestion, suggestive of their possible functions in the regulation of their complementary coding mRNA. Putatively, the *MTND6* lncRNA molecule protects the non-polyadenylated *MTND6* coding molecule through this duplex formation. On the flip side, this duplex formation also reduces the levels of translated products from *MTND6*, which is established to be less abundant compared to other mitochondrial proteins.

Further complexity of mtgenome is the recent finding of a novel class of abundant small noncoding RNAs, analogous to miRNA [8]. Referred to as mitosRNA, these molecules range in sizes from 20 to 40 nt long, with many composed of 30 nts. Their abnormal expression was associated with aberrant mitochondrial transcription, suggestive of a regulatory role by mitosRNA. While these small ncmtRNAs resemble miRNA, it appears they are not necessary processed by DICER. The polymerases involved in their expression as well as their processing molecular machinery are yet to be uncovered.

Another class of novel genes has been uncovered from the mitochondrial genome. These genes are created from two adjacent heavy-strand coding genes following large-scale deletions in the mitochondrial DNA (Fig. 6.2b). The mtgenome is studded with perfect and imperfect nucleotide repeats that serve as breakpoints. As biomolecules bathed in an organelle with high free radical production, these genomes sustain numerous large-scale deletions that occur at these repeats. Following a deletion, the mtgenome recircularizes. Thus, the resultant molecules, in their simplest forms, will be a large mtgenome that maintains the origins of replication, and a smaller one devoid of these replication sites. The large genome can then replicate, leading to its clonal expansion. Of interest, the deletion usually results in the fusion of the remaining 5′ upstream gene with the remaining 3′ downstream gene at the junctional point of recircularization. Surprisingly, the thousands of different deletions that occur in this genome give rise to fusion genes with open reading frames. Subsequent work has demonstrated that they are

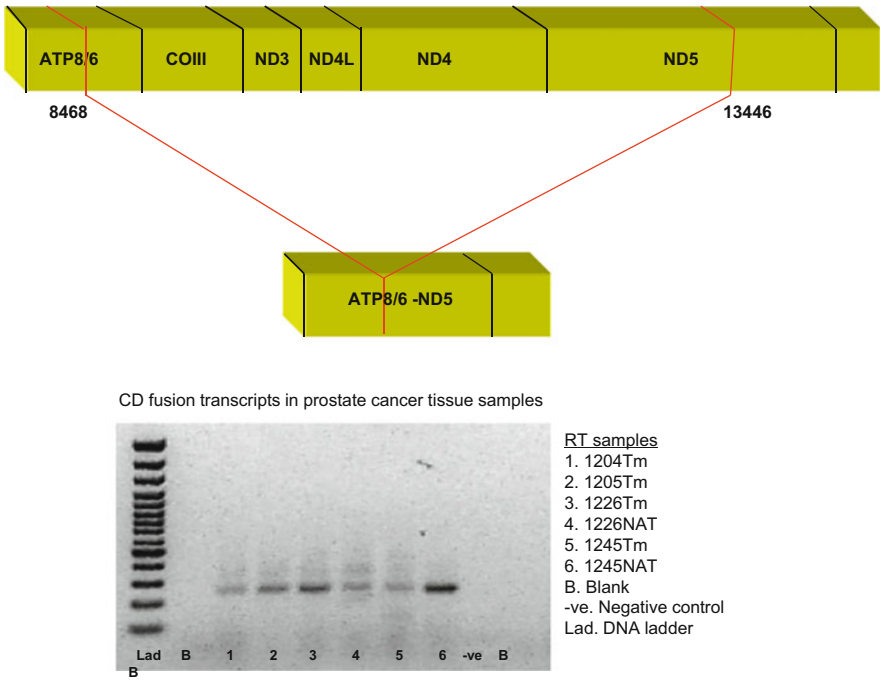

Fig. 6.3 The expression of the mitochondrial common deletion fusion gene transcript. The common (4977 bp) deletion removes the intervening genes (*MTCOIII*, *MTND3*, *MTND4L*, and *MTND4*) and parts of *MTATPase* and *MTND5*. The newly reconstituted gene has an ORF, and shown are expressions of the polyadenylated transcript in prostate cancer samples

transcribed and polyadenylated (Fig. 6.3). While the differential expression is demonstrated in several cancers, it is still unclear whether these molecules are translated. Therefore, it remains to be determined whether these are coding or noncoding transcripts. However, in a heteroplasmic state, whereby there are sufficient wild-type molecules, intramolecular complementation should make available adequate tRNA and rRNAs for their translation. The functional role of these novel molecules will be of interest in molecular pathology.

6.4 MicroRNA

The expansive body of knowledge on miRNA mostly pertains to their role in posttranscriptional gene regulation. Having a plethoric role in controlling over 50 % of coding genes, they are critical to animal development and viability, and yet the roles of some are dispensable in their developmental regulatory functions. The reason for this lack of effect in loss of functions of some miRNAs is due to the share high degree of redundancy built into their regulatory functions. A single

miRNA can control thousands of genes with many having overlapping partners. They also control all the cardinal hallmarks of cancer biology, such that they are emerging as potential therapeutic targets.

6.4.1 Historic Recap of MicroRNAs

In 1993, work by scientists from Harvard University identified a gene called *lin-4* that had profound influence on the development of *Caenorhabditis elegans* (*C. elegans*). *Lin-4* is a heterochronic gene that regulates the timing of *C. elegans* early larval development. Ambrose's group isolated *lin-4* [9]. By sequence comparison of lin-4 genes from four *C. elegans* species, as well as site-directed mutagenesis of open reading frames, it was uncovered that *lin-4* was a non-protein-coding gene. However, *lin-4* produces two small transcripts of 22 and 61 nt long. The 61 nt transcript folds into a stem-loop structure that predictively generates the shorter RNA. Subsequent work indicated this gene had no associated encoded protein and hence was appropriately referred to as noncoding small RNA.

Lin-4 negatively regulates another gene, *lin-14*, such that there is a temporal decrease in *lin-14* protein beginning at the first larval stage. It had previously been known that *lin-4* regulates a region of the 3′ UTR of *lin-14* [10]. Interestingly, Ambrose and Ruvkum's laboratories uncovered that *lin-4* transcripts had antisense complementary loci in these 3′ UTRs of *lin-14*. Subsequent work revealed that *lin-4* regulates *lin-14* protein levels without obvious changes in *lin-14* mRNA. The conclusion therefore was that, *lin-4* repressed the translation of *lin-14* at the level of the message.

The eventual realization that these small noncoding molecules are not limited to the nematode came with the discovery of yet another *C. elegans* heterochronic gene, *let-7*. This gene also encodes a 22 nt RNA that controls late *C. elegans* development, from larval to adult cell types [11]. Pasquinelli and coworkers [12] then revealed the presence of *let-7* homologues in humans and fly genomes and the RNA was detectable in several animal species. They were initially named small temporal RNAs because of their role in timing larval development. However, with the discovery of several other members with different functions, they were renamed microRNA. Cloning efforts soon uncovered numerous miRNAs in living organisms, and currently over 1500 miRNAs from the human genome have been uncovered. The miRNA family is highly conserved across species [13] and is even found in ancient species such as archaea and eubacteria.

6.4.2 MicroRNA and Cancer

Calin et al. first reported on aberrant miRNA expression in human cancer [14]. In B-cell chronic lymphocytic leukemia (CLL), miR-15 and miR-16 were

deregulated. Interestingly, these miRNAs are located on chromosome 13q14, a fragile site deleted in ~50 % of these neoplastic cells. Noteworthy, ~50 % of all miRNAs are located at fragile sites in the human genome [15], and these sites are often either deleted (sites of tumor suppressor miRNAs, tumor suppressormirs) or amplified (sites of oncogenic miRNAs, oncomirs). The two tumor suppressormirs (miR-15 and miR-16) deregulated in B-cell CLL target and decrease anti-apoptotic BCL2 expression. Therefore, their loss of function increases BCL2 expression to initiate cellular transformation.

Three years later, the next milestone in miRNA deregulation in cancer was reported. Lu et al. first developed a high-throughput profiling platform comprised of 217 different miRNAs for whole genome miRNA (miRNome) interrogation [16]. Using this platform, over a 100 tumors could be accurately classified in regards to their tissue of origin. This finding was the first indication that miRNA profiling has a potential utility in identifying the correct tissues or organs in people with tumors or cancers of unknown primaries (CUP).

The next phase in these developments, and to extend these findings to patient benefits, was the commercial development of miRNA profiling platform to identify tissues of origins in patients with cancers of unknown primaries (Rosetta Genomics). The first generation of this platform comprised 48 miRNAs used to identify primaries in patients with CUP. In a study, 253 tumors from 22 different cancers were typed at an overall accuracy of 90 %. The second-generation platform (Cancer Origin Test™) has 64 miRNAs that can classify 49 different cancer origins, using formalin-fixed and paraffin-embedded (FFPE) tissue samples.

With regards to miRNAs in circulating and body fluids of cancer patients, the landmark work of Lawrie et al. is noteworthy [17]. They were first to demonstrate the differential levels of miRNA in patients with B-cell lymphoma (BCL) compared to healthy controls. Elevated miR-21 and miR-155 were associated with BCL and could accurately distinguish patients from controls. In a couple of years later, Weber et al. reported on the presence of hundreds of miRNAs in 12 different normal body fluids, including plasma, CSF, tears, saliva, breast milk, colostrum, bronchial fluid, pleural fluid, peritoneal fluid, urine, seminal fluid, and amniotic fluid [18]. These findings certainly set the stage for miRNA exploration in body fluids for noninvasive clinical applications.

6.4.3 Genomics of MicroRNAs

These evolutionarily conserved small noncoding RNAs are composed of 18–24 nts. Many miRNAs are intergenic, being located far from annotated genes. Some occur in clusters and are transcribed as polycistronic units. Such clusters are very common in flies. Human examples include miR-15a/miR-16 cluster on chromosome 13, a region that contains a tumor suppressor gene. Although the worm *lin-4* and *let-7* genes are separated, they are arranged in clusters in the fly and human genomes. A few human miRNAs are located in introns with similar orientations as exons. These

miRNAs are transcribed en bloc with the genes and subsequently spliced and processed.

There are cell- and tissue-specific expressions of miRNAs. For example, miR-122 is expressed in the liver, miR-1 mainly in the mammalian heart, and miR-223 in mouse bone marrow granulocytes and macrophages. MiRNAs are also expressed in a temporal manner during embryonic development. There are copy number expression variations in cells. In some cells, the copy number is much higher than mRNA, and these offer an advantage over mRNA profiling in body fluids.

MiRNAs regulate sequence-specific posttranscriptional gene expression by endonucleolytic mRNA cleavage, translational repression, degradation, or deadenylation. These posttranscriptional processes occur in cytoplasmic processing bodies or P bodies. Thousands of coding transcripts are targeted by miRNAs. A single miRNA may control multiple genes, and several miRNAs may regulate a particular gene, accounting in part for the redundancy in their functions. They have diverse cellular functions including control of developmental biologic activities, and cellular physiology. Thus, many are deregulated in pathologic conditions. Mutations, methylations, and polymorphisms in miRNA genes, or their target genes, also act as another level of genomic regulation. Similarly, other genes control expressions of several miRNAs.

6.4.4 MicroRNA Biogenesis

MicroRNA genes are transcribed and processed into mature functional forms (Fig. 6.4). The nascent molecules, called primary miRNAs (pri-miRNAs), are much longer than the conserved stem-loop miRNAs. Primary miRNAs are capped and polyadenylated and contain single or multiple hairpin structures. They are transcribed using regulatory elements of genes (intronic miRNAs), or their own promoters (isolated intergenic miRNAs). Many are transcribed using RNA polymerase II, but a few are controlled by polymerase III.

Primary miRNAs can be as long as 1 kb. The stem is composed of ~30 imperfect base pairs with a distal single-stranded RNA segment beginning at a point referred to as the single-stranded/double-stranded RNA junction. These junctions are cleaved to release a ~60–70 nt stem-loop precursor miRNA (pre-miRNA). RNase III endonuclease DROSHA cleaves the pri-miRNA at sites near the base, releasing the pre-miRNA with staggered ends having a 5′ phosphate and 3′ ~2 nt overhang. DROSHA is part of a microprocessor complex that includes the RNA-binding protein, DiGeorge syndrome critical region gene 8 (DGCR8), and other proteins. DiGeorge syndrome critical region gene 8 recognizes the ss:dsRNA junction, while DROSHA cleaves the pri-miRNA at about 11 bp from the junction. The pre-miRNA is actively transported out of the nucleus by the guanine triphosphatase (Ran-GTP-dependent dsRNA-binding protein) and EXPORTIN-5 receptor. The 3′ overhang and double-stranded stem of at least 16 bps are minimal requirements for

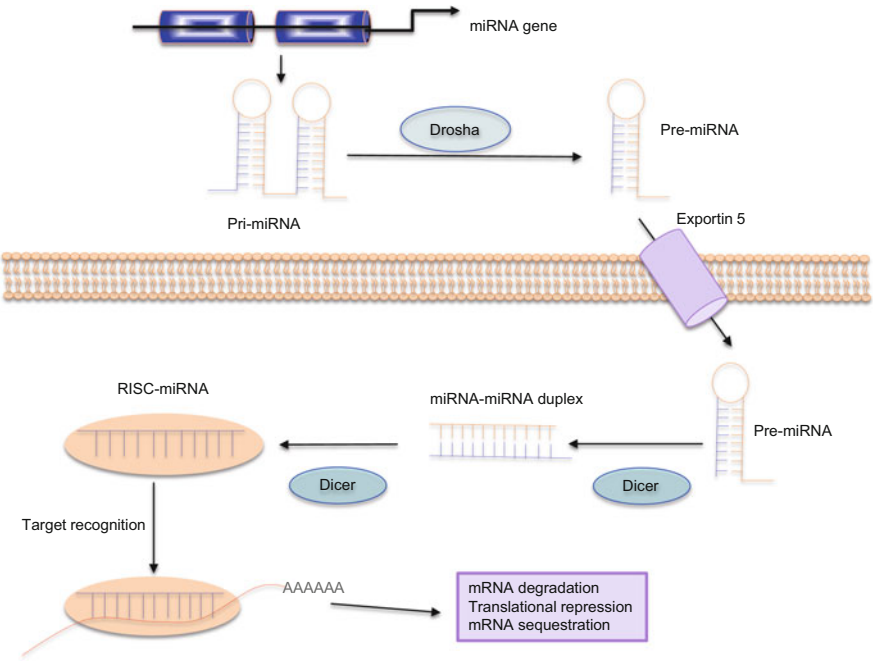

Fig. 6.4 MicroRNA biogenesis. MiRNAs genes are transcribed in the nucleus into pri-miRNAs, which are processed into pre-miRNA and then exported into the cytoplasm where they are finally converted into their matured forms

recognition by EXPORTIN-5. In the cytosol, pre-miRNA is delivered to another RNase III endonuclease, DICER, which is part of the RNA-induced silencing complex (RISC). DICER cleaves the pre-miRNA to release the mature miRNA. DICER was initially discovered for its role in producing small interference RNA used in RNA interference studies and applications. DICER recognizes the double-stranded portion of pre-miRNA, facilitated by its affinity for the 5′ phosphate and 3′ nt overhang. At about 2 helical turns (~22 nt) from the base, DICER cuts the DNA, releasing a miRNA-miRNA* duplex. As a characteristic of RNase III endonuclease activity, DICER similarly leaves a 5′ phosphate and 3′ nt overhangs; thus, the mature miRNA has 3′ 2 nt overhang and 5′ phosphate at both ends. The RISC loading complex includes DICER, members of the AGONUATE (AGO) family, HIV transactivation response (TAR) RNA-binding protein (TRBP), and kinase-activating protein (PACT). In most cases, only a single strand (known as the guide strand, miRNA) of the miRNA-miRNA* associates with AGO in the complex to form miRISC or programmed RISC, while the other strand (appropriately known as the passenger or passive strand, miRNA*) is released and degraded. Strand selection is determined by a number of factors including features of the duplex, such as base pairing, sequence composition, and importantly the thermodynamic stability

of the ends of the miRNA duplex. Because in some cases both the guide and passenger strands participate in gene silencing, these rules are a bit relaxed.

Non-canonical pathways of miRNA biogenesis have been reported. Short intronic hairpins (known as mirtrons) are spliced out of mRNA transcripts without DROSHA involvement. MiR-451 is cleaved by AGO2 at the 3′ arm, not by DICER. Here, the needed 3′ feature or overhang is generated by 3′–5′ exonuclease trimming. The complete knowledge of miRNA biogenesis awaits further investigation.

6.4.5 *Defects in MicroRNA Biogenesis in Cancer*

Cellular expression of miRNA is under multiple levels of control, namely, transcriptional regulation, genomic alterations (e.g., point mutations and deletions), and the miRNA processing machinery. While transcriptional control and genomic alterations often affect expression of one or just a few miRNAs, defects in members involved in miRNA biogenesis (e.g., DROSHA) often exert global effects on miRNA expression. These defects can occur at all levels of miRNA processing, from pri-miRNA to the eventual formation of the mature miRNA. A few evidences suggest these defects occur in some cancer cells.

Deregulated expression and functions of DROSHA/DGCR8 are associated with cancer. Both appear to function contextually in different cancers as either oncogenes or tumor suppressors. As an oncogene, loss of DGCR8 function is associated with decreased cell proliferation and reduced stem cell development. In this context, high expression is observed and has been demonstrated in colorectal cancer [19]. The tumor suppressor functions of DROSHA/DGCR8 are, however, observed in lung cancer. In this cancer, decreased DGCR8 levels are associated with increased cellular transformation, and loss of DICER1 enhanced tumor development in *Kras*-induced mouse model of lung cancer [20]. Similar increased and decreased expression of DROSHA is demonstrated in many malignancies, suggestive of cancer-specific or cellular-specific functions in miRNA biogenesis.

DROSHA functions less efficiently in the absence of two RNA-binding proteins, p68/DEAD-BOX5 (p68/DDX5) and p72/DDX75. Mutant p53 impedes association of p68/DDX5 with DROSHA, thereby reducing miRNA biogenesis [21, 22]. An example of the role of p68/DDX5 in miRNA biogenesis is the finding that decreased levels of this RNA-binding protein reduces expression of pre-miRNA-21 and mature miR-21, although the transcript levels of pri-miRNA-21 are normal. This mode of miR-21 regulation by p68/DDX5 appears to be mediated by TGFβ/BMP signaling pathways [23, 24]. Similarly, loss of let-7a is associated with Wilms tumor, and one mechanism by which let-7a levels are low in Wilms tumor is via mutations and copy number loss of DROSHA/DGCR8 and SIX1/2 [25]. Similarly, loss of DROSHA/DGCR8 is associated with decreases in the levels of miR-200 family members, which play important roles in stem cell maintenance and mesenchymal-to-epithelial transition (EMT) in cancer. Reduced levels of DROSHA in association with global decreases in miRNA expression confer

favorable survival in NSCLC patients [26]. On the contrary, Merritt et al. [27] associated high levels of DROSHA with increased survival, with low levels conferring poor survival in multivariate analysis.

Some cancer cells also exploit the transport mechanism to achieve their phenotypes. Loss of EXPORTIN-5 C-terminal binding region for pre-miRNA leads to the nuclear accumulation of pre-miRNA [28]. Inactivating mutations in EXPORTIN-5 are associated with decreased expression of some miRNAs, including miR-26a, miR-200 family, and let-7a. MiR-138 levels correlate with tumor progression and metastasis. This is partly achieved through loss of CDH1/E-cadherin, as well as promoting EMT through the control of EZH2, ZEB2, and VIM. One mode of achieving these effects by miR-138 is through its repression of EXPORTIN-5 stability and thus reduced expression of tumor suppressormirs.

DICER is another biogenic target demonstrating aberrant expression in different types of cancer, and the levels are of prognostic potential in some cancers. In breast, lung and ovarian cancers, increased DICER expression is associated with good prognosis [27]. Similarly, decreased expression is associated with cytogenetic aberrations and poor prognosis in chronic lymphocytic leukemia [29], and in advanced neuroblastoma, reduced DICER and DROSHA confer poor prognosis [30]. On the contrary, elevated DICER levels confer poor prognosis in colorectal and prostate cancer patients. In prostate cancer, elevated DICER levels correlate with tumor stage, lymph node status, and Gleason score [31]. Elevated DICER expression is associated with poor cancer-specific survival and progression-free survival in colorectal cancer (CRC) patients [32], while DICER is elevated early in rectal cancer development, and expression in normal mucosa from CRC patients is associated with poor survival [33]. Similar to DROSHA and DGCR8, DICER deregulation may be associated with deregulated expression of sets of miRNA in specific cancer types. DICER mutations in association with aberrant miRNA expression are found in some tumors. For example, in children with pleuropulmonary blastoma, loss of function mutations in DICER1 results in aberrant miRNA expression and mesenchymal cell proliferation [34]. Noteworthy, DICER1 mutations are associated with a group of events referred to as Dicer1-related disorders, which include embryonal rhabdomyosarcoma, multinodular goiter, and Sertoli-Leydig cell tumors. Finally, decreased expression of the dsRNA-binding protein, TRBP, impairs miRNA processing by DICER in cancer cell lines.

6.4.6 MicroRNA Release and Uptake

In both physiologic and pathologic states, intercellular communication networks are partly accomplished via miRNA. Cells constantly release miRNAs into body fluids, and their biologic functions can occur in the immediate vicinity in a paracrine fashion. Alternatively, they can be carried in the circulation to distant sites to act on other cell types in an endocrine manner. Moreover, miRNAs can be packaged in extracellular vesicles (e.g., exosomes) and be secreted or released in

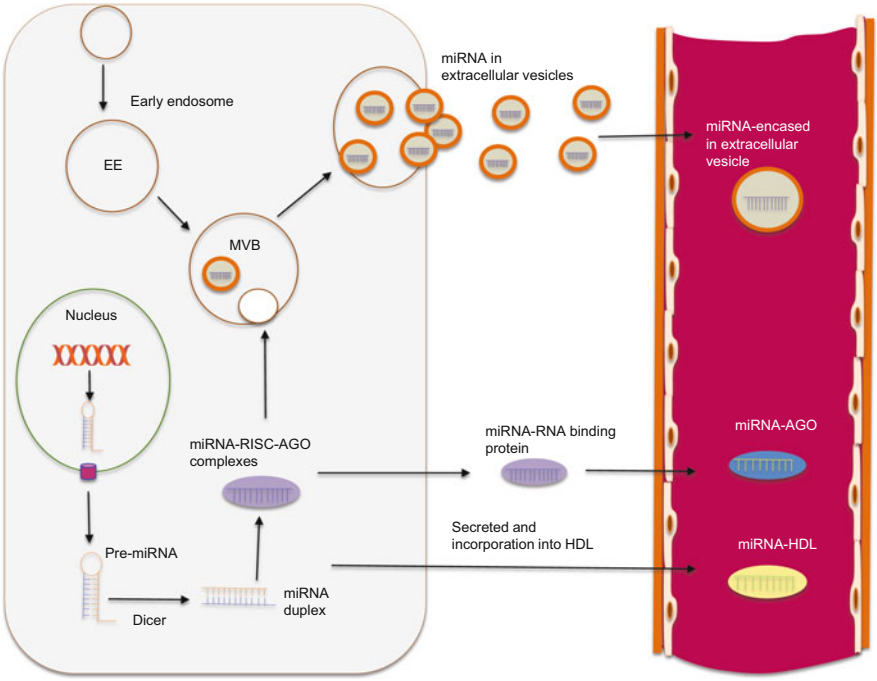

Fig. 6.5 MiRNA release and protective mechanisms in body fluids. MiRNAs are released by secretion, enclosed in extracellular vesicles, mostly exosomes, and with RNA-binding protein complexes. The association of miRNAs with these structures, as well as HDL, partly contributes to their protection in circulation

association with RNA-binding protein complexes (Fig. 6.5). A few examples illustrate the dynamic cellular functions of miRNAs:

- Microvesicles from embryonic stem cells contain numerous miRNAs that are transferable to impact gene expression in neighboring cells.
- MicroRNA from EBV-infected cells is known to spread to uninfected cells through exosomes.
- Another possible but less likely mode of miRNA release is via apoptotic bodies, which are membrane-bound microvesicular structures shed during cellular apoptosis. For example, in atherosclerosis, miR-126 enriched in epithelial cell-derived apoptotic bodies mediate paracrine signaling to recipient cells leading to the production of chemokine CXC motif ligand 12 (CXCL12).

The mechanism of release and uptake of miRNA are not well understood. Evidence, however, indicates that they are at least also released via ceramide-dependent secretory pathways [35]. Neutral sphingomyelinase 2 (nSMase 2), involved in exosomal release, also controls miRNA secretion. Neutral sphingomyelinase 2 is a rate-limiting enzyme (an important enzyme) in ceramide biosynthesis and thus controls the ceramide-dependent secretory pathway. Thus,

inhibition of nSMase 2 impairs miRNA export [36, 37]. Moreover, unlike exosomes, the endosomal-sorting complex for transport (ESCRT) system is dispensable in miRNA release from cells. Thus, overlapping and non-overlapping mechanisms operate in the release miRNAs and exosomes.

6.4.7 MicroRNA Target Recognition and Gene Silencing

MicroRNAs must identify and interact with their targets to alter gene expression. Thus, simple diffusion processes transport them to target locations, and partial complementarity of miRNAs to target mRNA enables desired interaction and effects. There are no identified target recognition signals in miRNA, which may partly explain their functional redundancy. However, initial perfect match and pairing of at least 2 nt (usually 2–8 nts) of the miRNA to target sequences in the 3′ UTR of the mRNA is critical for target selection. Most probably, the main mode of gene silencing is mRNA destabilization and degradation. However, translational repression or arrest, mRNA deadenylation, and, to a lesser extent, cleavage are other noted mechanisms of posttranscriptional gene silencing.

6.4.8 MicroRNA Stability in Body Fluids

In spite of the abundance of ribonucleases in serum, miRNAs are very stable, being resistance to digestion by these enzymes. Hash treatment conditions such as boiling, exposure to very low or high pH conditions, extended storage at room temperature, and several freeze-thaw cycles that easily destroy mRNA and exogenous or synthetic miRNA fail to degrade endogenous miRNA. They are even preserved in archived 10-year old samples and unrefrigerated dried serum samples.

Cells are known to communicate using miRNA in microvesicles and exosomes. MicroRNAs are also abundant in apoptotic bodies and prostate ductal epithelial-derived particles called prostasomes. All these membrane enclosures provide protection against degradation. It may appear however that very small amounts (~10 %) of miRNAs are vesicle-associated, with the majority being bound to protein and lipid complexes. Their inclusion in lipid and lipoprotein complexes including HDL, LDL, AGO2, and nucleophosmin 1 (NPM1) is one mechanism by which they may be protected in circulation (Fig. 6.5). AGO2 is recognized as a primary protective carrier molecule of miRNAs in circulation. Another important carrier and transporter of miRNAs is HDL. Additionally, the RNA-binding protein, NPM1, is involved in packaging, exporting, and protecting miRNAs. All these protective mechanisms are important in liquid biopsy applications targeting circulating miRNA.

6.4.9 Physiologic and Pathologic Variations in Circulating MicroRNA

MicroRNA levels appear to differ in normal people and under different physiologic conditions. Normal levels of circulating miRNA vary with age, gender, race, intense exercise, and in other physiologic states such as pregnancy, and environmental stressors. For example, circulating levels of miR-18b and miR-130 are higher in males than females, while miR-145 and miR-181a-2 levels are increased by intense exercises. Serum miR-520d-5p, miR-526a, and miR-527 are highly elevated in pregnancy and are considered putative pregnancy biomarkers. MicroR-141, miR-149, miR-517a, and miR-299-5p emanate from the placenta and are measurable in maternal plasma. Their levels fall postpartum and hence may have potential utility in prenatal diagnosis. Owing to their release, transport, and actions at distant sites, miRNAs are considered to have physiologic functions such as hormones.

Almost all pathologic conditions are modulated by miRNA. For example, pathologic conditions such as myocardial infarction (MI), liver damage, and inflammatory reactions can cause changes in the serum levels of miRNAs. The circulating levels of cardiac-specific miR-208 and miR-499 are elevated in rat model of MI, and they are similarly elevated in sera from patients with acute MI. Liver injury resulting from drug reaction in mouse models causes increased plasma levels of miR-122 and miR-192. Sepsis is associated with reduced levels of miR-146a and miR-223, while rheumatoid arthritis and osteoarthritis are associated with reduced miR-132. In cancer, circulating miRNAs are associated with disease evolution, clinicopathologic features, and prognosis and are predictive of drug response and disease relapse. For example, miR-141 and miR-375 are markers of prostate cancer progression and can delineate metastatic from localized prostate cancer. Similarly, miR21, miR-73, miR-195, and let-7a are associated with breast cancer and miR-21, miR-155, and miR-210 with diffuse B-cell lymphoma, while miR-1, miR-30d, miR-486, and miR-499 predict outcomes in NSCLC.

6.4.10 MicroRNA Alterations in Cancer

MicroRNAs are aberrantly expressed in all types of cancers. Various signaling networks, DNA methylation, mutations, and SNPs control their expression. Lu et al. [16] demonstrated the generalized deregulated miRNA profiles in many cancers, and work by He et al. [38] unveiled that miRNA overexpression can initiate the development of cancer. Almost all cancers harbor aberrant expressions of miRNAs. These cancer-specific upregulated or downregulated miRNAs serve important functions in cancer development by altering oncogenes, tumor suppressor genes, and therefore cancer-specific signaling pathways. MicroRNA expression profiles also serve as important risk assessment, screening or diagnostic, treatment

selection, prognostic, and monitoring biomarkers of cancer. In some instances, the expression profiles of miRNAs are so specific to cancers, such that they can inform the primary sites/tissues of origin in patients with cancers of unknown primaries.

MicroRNAs are expressed in the germline as well. Aberrant germline expressions may predispose individuals to specific cancers. SNPs are rare in miRNA genes; however, the presence of SNPs in miRNA or their target genes are associated with increased risk of cancer evolution, and such biomarkers are important in cancer risk stratification. These risk-associated miRNAs are commonly found in cancer-associated genomic regions (CAGRs). Genomic aberrations including loss of heterozygosity, deletions, breakpoints, and other rearrangements affect miRNA expression. As examples, SNP rs895819 in pre-miR-27a decreases, while SNP rs11614913 in pre-miR-196a-2 increases the risk of breast cancer. Moreover, in African-American women, SNPs rs2008591 and rs887205 in miR-185 are inversely associated with breast cancer risk, while rs4938723 in miR-34b/34c and rs6920648 in miR-206 may be survival predictors [39]. Additionally, SNP rs2910164 in the 3′ strand of miR-146a increases the risk for developing papillary thyroid cancer; SNP rs531564 in pri-miR-124-1 increases the risk for developing bladder and esophageal cancers; and SNP in the 3′ UTR of *KRAS* that affects the binding of let-7 increases the risk for lung cancer in moderate smokers. Several of these risk alleles have been identified. Thus, validation studies are required to establish these associations and their roles in cancer biology.

6.4.11 Oncogenic and Tumor Suppressor Functions of MicroRNAs

MicroRNAs can function as oncogenes or tumor suppressors. Oncogenic miRNAs are upregulated in cancer cells and target the degradation of tumor suppressor genes to promote tumor formation. Elevated expressions of tumor suppressor miRNAs on the other hand prevent tumor growth by inactivating oncogenes and hence are in general downregulated in cancer cells (Fig. 6.6).

There are numerous deregulated miRNAs in cancer. Just a few are provided here to illustrate their important roles in cancer. MiR-21 for instance is upregulated in a number of cancers and targets several genes including the *PTEN* and *PDCD4* tumor suppressor genes. Similarly, miR-10b is upregulated in tumors and suppresses *HOXD10* to promote tumor growth and metastasis. MiR-17-92 cluster enhances tumor cell proliferation, angiogenesis, and survival by targeting a number of genes including *PTEN* and *RAB14*. Tumor suppressor miRNAs include let-7 family members. They inhibit cancer growth by regulating oncogenes such as *HMGA2* and *RAS*, while miR-34 family are part of the *TP53* regulatory network involved in cell cycle control and apoptosis. Thus, miR-34 inhibits *SIRT1* to enhance p53 levels and hence increased expression of cell cycle regulator p21 and PUMA involved in cellular apoptosis.

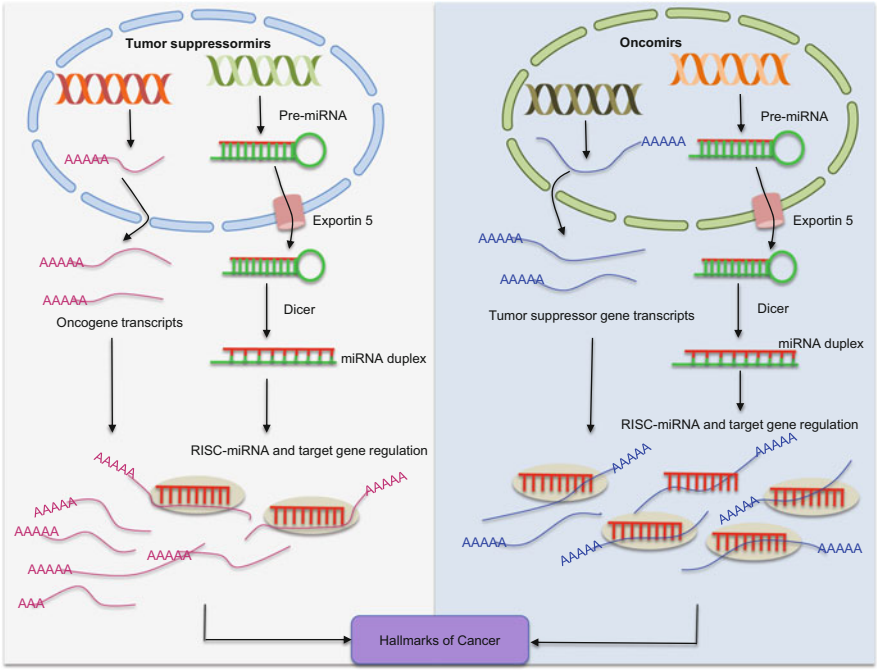

Fig. 6.6 Oncomir and tumor suppressormir functions of miRNAs. In cancer cells, some miRNAs are overexpressed (oncomirs) and target the repression of tumor suppressor genes, while others are downregulated (tumor suppressormirs) in order to prevent their repressive effects of oncogenes

The designated oncogenic and tumor suppressor functions of miRNAs is contextual. The same miRNA may function as an oncogene or tumor suppressor based on specific cell or cancer type. MiR-125b is a classic example. In oral squamous cell carcinoma and breast, thyroid, and ovarian cancers, it has a tumor suppressor function and is thus downregulated. As a tumor suppressor, miR-125b inhibits cell proliferation and cell cycle progression in those tumors and their cell lines. MiR-125b is on the other hand oncogenic in prostate cancer, neuroblastoma, glioblastoma, and thyroid cancers, in which it is overexpressed. When upregulated, miR-125b inhibits apoptosis in a p53-dependent fashion and enhances cell proliferation and invasion. Other miRNAs with such dual functions include miR181a, miR181c, and miR-220 [40]. Again the plethoric targets and absence of targeting signals, coupled with the overlapping and redundant functions of miRNAs, may partly explain such functional plasticity.

6.4.11.1 MicroRNA Targets in Cancer

MicroRNAs target a large variety of genes to mediate their molecular pathologic effects in cancer. Established miRNA target genes are deposited at the miRTarBase

[41]. Currently, in humans, there are 38,113 miRNA-target interactions orchestrated by 587 mRNAs and 12,194 target genes.

The conserved miRNA, let-7 (tumor suppressormir) directly targets *RAS* oncogene. Thus, levels of let-7 are markedly reduced in several cancers including those of the skin, stomach, colon, lung, ovary, and in benign uterine tumors (myoma), and these levels correlate with increased *RAS* expression. Reduced let-7 levels promote colony formation of cancer cell lines in vitro and are also associated with significant reduction in survival of lung cancer patients. The chromosome X miRNA 221 and miRNA 222 family directly target and inhibit *PTEN* and *CDKN1B* tumor suppressor genes to promote cell cycle progression, cell growth, and invasiveness. These miRNAs are upregulated in glioblastoma, hepatocellular carcinoma, and ovarian cancer. MiR-21, miR-155, miR-196a, and miR-210 are elevated in plasma of pancreatic cancer patients and have biologic implications. MiR-196a has prognostic utility in pancreatic cancer. MiR-21 targets *PTEN* and *PDCD4*. MiR-155 represses the expression of *TP53INP1*, while miR-210 is induced by hypoxia through HIF-1-α-mediated mechanisms. MiR-1 is downregulated in lung cancer, and its targets include *MET* oncogene and *HDAC4*. These alterations are associated with increased cell proliferation and migration.

6.4.12 Precautionary Measures in MicroRNA Analyses

Circulating miRNAs have a strong clinical potential in translational research and product development. However, discrepancies exist between the various studies of miRNA in body fluids. There are a number of reasons that account for this. Among them is the lack of standard operating procedures for miRNA analyses using body fluids. Thus, analytical and even biological (see physiologic variations of miRNA) issues may partly be blamed for lack of data uniformity. Below are some of these issues that require careful consideration in miRNA analysis.

- Commonly used samples in terms of frequency are serum, plasma, whole blood, and peripheral blood mononuclear cells (PBMCs). The choice of sample can influence the study outcome. While there is a good correlation in miRNA levels between serum and plasma, it should be noted that blood formed elements contain miRNA that can mask disease-specific miRNAs in circulation. Thus, whole blood and PBMCs ought to be avoided in situations where the research question does not need such interference.
- Diligent blood processing is necessary to prevent hemolysis and platelet activation or contamination. Hemolysis is associated with increases in miR-15b and miR-16, while platelet contamination will aberrantly increase miR-223 levels.
- Standard operating procedures should be established for RNA extraction and other downstream analyses. Different RNA extraction protocols introduce variability in detected miRNAs. For example, TRIzol® reagent (Thermo Fisher

Scientific) extraction causes loss of miRNAs with low guanine and cytosine content.

- MicroRNA sequence and structure posse unique challenges in regard to qPCR and microarray profiling studies. MicroRNAs from the same family with minimal sequence variation are of particular concern because as few as 1–2 nt differences at the 3′ end can spuriously cause differences in expression levels. Moreover, sequence differences also affect enzymatic reactions (e.g., ligases), and annealing temperatures of primers and probes, which could preferentially favor some miRNAs over others.
- Platform differences are another consideration in miRNA analyses. MiRNome studies rely on initial analysis of small sample sizes between cases and controls, followed by prioritized biomarkers for validation and studies using large sample sizes. There are various and diverse platforms used in these initial discovery studies, including microarrays, next-generation sequencing, and microfluidic arrays. These platforms differ in their analytical sensitivities and specificities and hence will generate different data sets, even when using the same samples.
- The promising biomarkers from miRNome studies must be validated using the gold standard, qPCR, following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [42]. An outstanding unresolved issue is normalization. Different investigators use different methods. These include spiking exogenous miRNAs into the reaction, using other small RNAs (not miRNAs), and using miRNAs that show no differential expressions between samples.

Standardization is thus required in miRNA analysis using body fluids. This needed leadership should enable the real potential of circulating miRNAs to be harnessed.

6.5 Nonhuman MicroRNA in Circulation

While the field of miRNA analyses has flourished with potential clinical applications, recent evidence suggests a possible contamination from exogenous sources. MicroRNA from ingested plants has been demonstrated to cross the gut-blood barrier into the circulation [43]. Additionally, these miRNAs are not passive molecules, but were shown to regulate physiologic functions. MiR-168a, abundant in rice, is commonly found in sera from Chinese subjects. Functional studies suggest miR-168a binds to lipoprotein receptor adaptor protein 1 (e.g., LDLRAP1) transcripts, thus reducing the levels of LDLRAP1. The consequence of this is elevated plasma LDL, since its clearance by LDLRAP1 is reduced [43]. These findings have been collaborated by other investigators. Indeed, using next-generation sequencing, it has been estimated that ~40 % of circulating RNA is of nonhuman source. They primarily are of microbial and common food sources. Thus, clinically useful circulating miRNA must be demonstrated to be of disease

origin. The fact that ingested miRNA can be absorbed opens up the possibility of delivering ingestible and nondigestible therapeutic noncoding RNAs.

6.6 Other Small ncRNAs in Circulation

Other small ncRNAs show differential levels in circulation of cancer patients compared to controls. The levels of small nuclear RNA (snRNA), U2, are elevated in blood from women with ovarian cancer compared to controls, and this was associated with response to chemotherapy [44]. Six small nucleolar RNAs (snoRNA) were elevated in plasma from NSCLC patients [45]. The actual relevance of these small ncRNAs in cancer awaits further elucidation.

6.7 Long Noncoding RNA in Cancer

Noncoding RNAs of over 200 nts in length are referred to as long noncoding RNAs (lncRNA). Compared to the ~2000 miRNAs, about 16,000 genes transcribe the ~28,000 human lncRNAs currently catalogued. Yet, unlike miRNAs that have received extensive studies, our knowledge on lncRNAs is just emerging. Indeed, we are still uncertain as to the mechanisms of their biogenesis and release into body fluids. However, knowledge has been garnered on the differential expressions of lncRNAs in diseases, as well as some pathogenic mechanisms of these biomolecules. LncRNAs tend to exhibit cell-type specific expression with subcellular compartmentalization, although many are localized in the nucleus where they control gene expression. They are involved in normal cellular functions, in metazoan development, and are deregulated in many disease states including cancer.

6.7.1 Genomic Structure of LncRNAs

Similar to miRNAs, lncRNAs are transcribed by RNA polymerase II. They are 5′ capped, 3′ polyadenylated and variably spliced. Compared to mRNAs, they tend to have fewer exons with negligible, absent, or short open reading frames (ORFs). The primary and secondary structures of these molecules are currently being deciphered and annotated. Of noteworthy, functional lncRNAs share some primary sequence structure, including low GC content, low number of introns, and poor start codon and ORFs.

LncRNAs have diverse functions that require their interactions with various biomolecules, including DNA, RNA, and proteins. Thus, some sequence motifs in lncRNAs enable these biomolecular interactions to occur. Indeed, part of the mechanism of protein-coding gene regulation by several lncRNAs is through

provision of binding sites for miRNA and half-STAU1, to control gene expression. In addition to the primary sequence motif, lncRNAs also demonstrate secondary structures that are necessary for their functions. The lncRNA, steroid receptor RNA activator (SRA), has complex structural organization that includes four distinct domains with secondary structures, which are critical for their interactions with a number of proteins. The maturation of the lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) includes the formation of a cloverleaf secondary structure at its 3′ end. Moreover, the lncRNA HOX transcript antisense RNA (HOTAIR) forms multiple double stem-loop secondary structures that interact with histone-modifying complexes including lysine-specific demethylase 1 and polycomb repressor complex 2 (PRC2).

6.7.2 Classification of LncRNAs

The large body of lncRNAs necessitates their classification, which is important in their functional characterization and annotation. Currently, they are classified using various criteria, summarized herein:

1. The most common classification scheme considers the length of the primary transcript. Although defined as transcript lengths >200 nts, some lncRNAs can be as long as 1 Mb. This class includes:
   (a) Long intergenic ncRNAs (lincRNAs). Specific examples include HOTAIR and HOTTIP.
   (b) Very long intergenic ncRNAs (vlincRNAs). Specific examples include HELLP and vlinc_21.
   (c) Promoter-associated long RNA (PALR).
   (d) MacroRNAs. Specific examples include KCNQOT1 and Airn.
2. Another commonly used classification scheme (used by GENCODE) considers the location of the lncRNA in relationship to protein-coding genes. Using this criterion, they are variously referred to as:
   (a) Intronic.
   (b) Sense.
   (c) Antisense (natural or mirror).
   (d) Circular exonic.
   (e) Chromatin interlinking.
   (f) NcRNA from the 3′ UTR.
   (g) Transcript start site-associated RNAs.
3. Less commonly used classification schemes are based on other criteria such as:
   (a) Being associated with repeats (e.g., *PTENP1* and *KRASP1* transcribed pseudogenes).
   (b) Being associated with biochemical pathways (e.g., *H19*).

(c) Being associated with DNA of known functions (e.g., telomeric repeat-containing RNA, *TERRA*).
(d) Being associated with subcellular structures (e.g., chromatin-associated RNA, *CAR*, and PCR2-associated RNAs).
(e) Their resemblance to structural/coding transcripts (e.g., *HOTAIR*, *HOTTIP*, *XIST*, *ANRIL*, and *lincRNA-p21*).
(f) Conservation in their sequence and structure (e.g., transcribed-ultraconserved regions, *UCR106*).
(g) Their involvement in specific functions (e.g., lncRNA with enhancer-like functions).
(h) Induced expression under certain physiologic conditions (e.g., hypoxia-induced noncoding ultraconserved transcript, *HINCUT*).

6.7.3 *Functions of LncRNAs*

LncRNAs control gene expression at multiple levels (Fig. 6.7). These include regulating epigenetic processes (e.g., by modifying chromatin structure), gene transcriptional processes, mRNA structural stability, posttranscriptional processes, and translation. They achieve these functions through different means of controlling the regulators of gene expression, as well as their interactions with DNA, RNA (including miRNA), and various proteins. Some specific examples of the mechanisms used in gene regulation are summarized herein.

LncRNA can alter the epigenome to silence gene expression by recruiting chromatin remodeling complexes (e.g., PRC2) to form condense chromatin. Indeed, a major function of lncRNA is in influencing epigenetic modifications by enabling recruitment of chromatin remodeling complexes [46]. Nearly 38 % of all human lncRNAs interact with PRC2, and/or chromatin-modifying proteins including CoREST and SMCX [47], as well as trithorax chromatin-activating complexes [48]. The lncRNA high expression in hepatocellular carcinoma (HCC) (*HEIH*) achieves this possibly via its interaction with enhancer of zest homolog 2 (EZH2), which is a component of PRC2. The lncRNAs can also bind to chromatin and recruit histone-modifying enzymes to alter chromatin structure. Other well-characterized lncRNAs that regulate chromatin to control gene expression are *HOTAIR*, *XIST*, *KCNQ1OT1*, and *ANRIL*.

Another mechanism of gene regulation is to enhance or repress gene transcription through a number of mechanisms. To promote gene expression, they can demethylate gene promoters or act as transcriptional coactivators. On the contrary, they can repress gene expression by sequestration of transcription factors and RNA-binding proteins or through direct interaction and inhibition of promoters. As an example, the *PTEN* pseudogene, *PTENP1*, is a lncRNA that has a well-conserved 3′ UTR, which can repress miRNAs, that represses *PTEN* expression. Thus, *PTENP1* expression enhances *PTEN* expression [49].

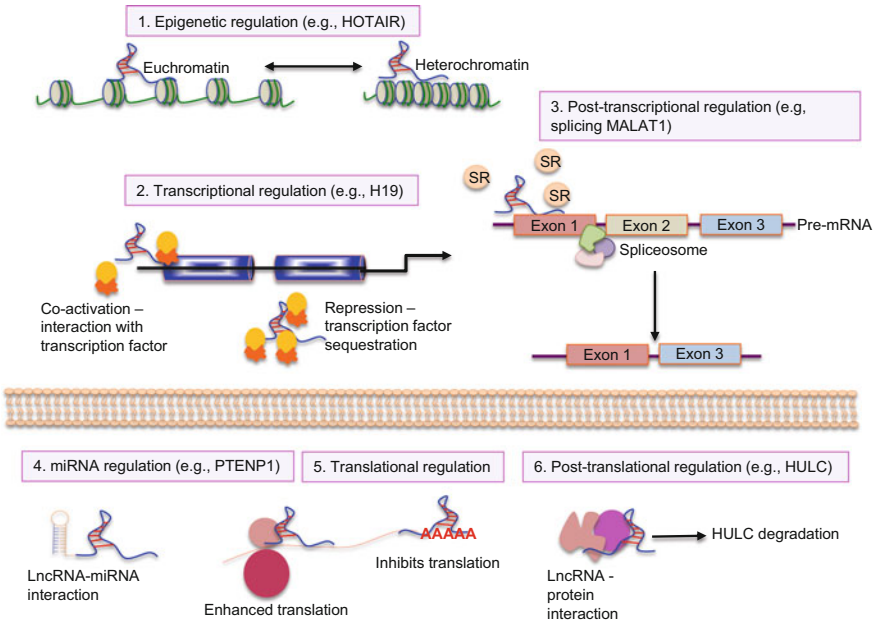

Fig. 6.7 The functional mechanisms of lncRNA in controlling cancer progression. Long noncoding RNAs control gene functions at several levels from chromatin structural modification, transcription, posttranscriptional modification, translation, and posttranslational modification

The lncRNAs can also modify the activity of the splicing machinery to regulate alternative pre-mRNA splicing. *MALAT1*, for example, can interact with the serine-/arginine-rich family of nuclear phosphoproteins (the SR proteins) to change their distribution and phosphorylation states, in order to regulate alternative pre-mRNA splicing.

Finally, they can also alter gene expression at the mRNA level. LncRNAs can decrease mRNA stability through STAU1-mediated mRNA degradation or enhance their stability by hindering mRNA degradation through the formation of perfect base pairing. They are also able to promote translation of mRNA, which is usually mediated by antisense lncRNAs. These antisense transcripts can bind to the 5′ ends of the target mRNA to enhance their interactions with polyribosomes. Some lncRNAs can also directly bind to transcripts, thus obstructing ribosomal interactions and hence hinder translation.

6.7.4 Expression of LncRNAs in Cancer

A number of lncRNAs demonstrate deregulated expression in a number of cancer types. While some are fairly cancer-type specific (e.g., *PCA3* and prostate cancer), making them attractive diagnostic biomarkers, many show expressional changes in

a broad range of cancers, making them useful prognostic or therapy selection or monitoring, rather than diagnostic biomarkers.

In gastric cancer, the lncRNA AA174084 is downregulated (~3.18 times) in 71 % of cancer tissues from patients compared to normal gastric tissue [50]. The reduced levels are associated with tumor aggressiveness. As a diagnostic biomarker, AA174084 achieved a sensitivity of 57 %, specificity of 73 %, and a fairly good AUROC of 0.676. A number of lncRNAs inducing *HOTAIR*, *MVIH*, and *MALAT1* are deregulated in liver cancer. However, a promising biomarker in HCC is the lncRNA, highly upregulated in liver cancer (*HULC*). *HULC* overexpression in HCC has a diagnostic AUROC of 0.86, and elevated levels are associated with aggressive tumors [51]. Similarly, increased *MVIH* expression is associated with higher tumor-node-metastasis and reduced recurrence-free survival and overall survival [52]. Moreover, *HOTAIR* overexpression in HCC predicts tumor recurrence after liver transplantation [53].

The lncRNA, urothelial carcinoma associated 1 (*UCA1*) is a potential urothelial cancer-specific biomarker. It is highly expressed in tumor tissues and can discriminate between bladder cancer and nonurothelial cancers such as renal cell carcinoma [54]. The lncRNA prostate cancer antigen 3 (*PCA3*/*DD3*) is highly expressed in primary prostate cancer tissues compared to nonmalignant prostate, and this appears specific to prostate cancer because it is undetectable in tumors from non-prostatic origins [55]. As a diagnostic biomarker, tissue-expressed *PCA3* was able to achieve a AUROC of 0.98 [56]. This biomarker has since been developed as a urinary diagnostic biomarker for prostate cancer, although with lower performance than the tissue assay.

HOTAIR and *MALAT1* demonstrate deregulated expression in many types of cancer. *MALAT1* is overexpressed in lung, gastric, liver, and colorectal cancers, and high levels are associated with poor prognosis in colorectal cancer and clear cell renal cancer [57, 58]. In a similar fashion, *HOTAIR* is overexpressed in breast, lung, and gastrointestinal cancers. The high levels in these tumor tissues are predictive of metastatic potential and poor prognosis [59, 60].

6.7.5 LncRNA Mutations in Cancer

In addition to the expressional changes of lncRNA, mutations in the primary sequences of these transcripts are associated with cancer. Cancer risk loci unraveled by genome-wide association studies (GWAS) are mostly in noncoding genomic regions. Indeed comprehensive analyses of GWAS catalogue identified 301-cancer risk SNPs. Of the 301 cancer SNPs, only 12 (3.3 %) involved exons with the majority found in noncoding intergenic (44 %) and intronic (40 %) sequences, indicating an increased probability of their involvement in ncRNAs [61]. Some specific SNPs involving lncRNA in a few cancers are provided as examples.

In HCC, SNP rs7763881 in *HULC* is significantly associated with reduced risk of HCC [62]. Additionally, a 4 bp insertion/deletion polymorphism, rs10680577 in

lncRNA, *RERT* contributes to the progression of liver cancer. The papillary thyroid cancer (PTC) risk loci, 14q13.3, contain the tumor suppressor lncRNA PTC susceptibility candidate 3 (*PTCSC3*). *PTCSC3* regulates the expression of genes involved in DNA replication and repair, cell death, cell motility, and tumor morphology. A specific SNP, rs944289, in *PTCSC3* interferes with the interaction of its promoter with C/EBP, leading to its reduced expression. Thus, in papillary thyroid cancer, this SNP causes strong downregulation of *PTCSC3* [63].

An important GWAS "hotspot" is located at 126 kb lncRNA, *ANRIL*, which is implicated in a number of disease conditions including vascular diseases, type 2 diabetes, glaucoma, and cancer (e.g., breast cancer). The *ANRIL* locus is adjacent to the p14/ARF gene locus. This lncRNA interacts with polycomb proteins to epigenetically modify histones to repress genes including p15/CDKN2B, p16/CDKN2A, and p14/ARF to halt cell cycle progression. SNPs in *ANRIL* that may impede its interaction with DNA and proteins have been identified. Of interest, these SNPs appear to cluster in a disease-specific fashion that may explain differential functions in different tissues. For example, SNPs at the 3′ end are associated with vascular pathologies, whereas those at the 5′ end, which may be involved in cell cycle control, are cancer risk-associated alterations [61].

6.7.6 LncRNAs in Cancer Biology

The biologic functions of deregulated lncRNAs in cancer have been addressed using both in vivo and in vitro model systems. Indeed they influence all aspects of the hallmarks of cancer, namely, sustained proliferation, refractory from growth suppression, evasion of apoptosis, invasion and metastasis, angiogenesis, and altered cancer cell metabolism.

Although *HOTAIR* can stimulate cancer cell proliferation, it appears to play a major role in the metastatic cascade as demonstrated by in vitro and in vivo models. *HOTAIR* strongly induces epithelial-to-mesenchymal transition (EMT) [64], and in several cancer cell lines, overexpression enhances invasive phenotypes [65, 66]. In an in vivo model, overexpression of *HOTAIR* in non-metastatic cell lines increased their degree of metastasis. The lncRNA growth arrest-specific 5 (*GAS5*) is an inducer of apoptosis. Thus, to evade growth suppression, *GAS5* is downregulated in several cancers [67, 68].

Unlike *HOTAIR*, *MALAT1*, overexpressed in multiple cancers, is a strong regulator of cancer cell proliferation and growth via control of the cell cycle. In vitro downregulation in several cancer cell lines caused cell cycle arrest at G2/M transition with eventual apoptosis [69–71]. These growth inhibitory functions of *MALAT1* are also demonstrated using in vivo xenograft models and corroborated by overexpression studies [70]. *MALAT1* may also contribute to cancer cell invasion and metastasis as it promotes angiogenesis and EMT.

PCGEM1 is a prostate cancer-associated lncRNA that is overexpressed in prostate cancer tissues. Although its biomarker potential is yet to be explored,

PCGEM1 confers on these cells the altered cancer cell metabolic phenotype critical for their survival. Not only does this lncRNA enhance glucose uptake, but it also acts at the transcriptional level to control the metabolism of glucose and glutamine, as well as the factors involved in tricarboxylic acid, fatty acid, pentose phosphate, and nucleotide biosynthesis pathways [72].

Interplay between p53 and the lincRNA-p21 exists that suppresses tumor growth. In response to DNA damage, p53 directly binds to lincRNA-p21 promoter to activate its expression. LincRNA-p21 then recruits and localizes heterogeneous nuclear ribonucleoprotein K to promoters of p53 responsive genes [73]. LincRNA-p21 is thus a transcriptional repressor that responds to p53 tumor suppressor gene signaling.

Cancer cells increase telomerase expression in order to maintain normal telomeres and hence continue to proliferate and grow. The lncRNA *TERRA* transcribed from telomeric regions of the genome can bind and inhibit telomerase activity [74]. Thus, as a tumor suppressor, *TERRA* is downregulated in several cancer types.

6.7.7 Circulating LncRNA in Cancer Patients

The deregulated lncRNAs in a variety of cancer cells is reflected in their levels in body fluids as well. They have been mostly quantified in the peripheral circulation and urine as cancer biomarkers.

Plasma lncRNA *H19* levels were significantly higher in gastric cancer patients compared to healthy controls [75]. In HCC, the diagnostic potential of a number of circulating lncRNAs has been explored. *HULC* that is overexpressed in tumor tissues is equally elevated in plasma of HCC patients [51]. Other noninvasive diagnostic lncRNA biomarkers of HCC include elevated plasma levels of *RP11-160H22.5*, *XLOC_014172*, and *LOC149086* [76]. As a panel, these three lncRNAs achieved a diagnostic performance with a sensitivity of 82 %, specificity of 73 %, and AUROC of 0.896. Additionally, *LOC149086* and *XLOC_014172* are possible metastasis prediction biomarkers with discriminating accuracy of 91 %, 90 %, and 0.934 in sensitivity, specificity, and AUROC, respectively. Another serum lncRNA *AF085935* could distinguish HCC patients from healthy controls and HBV-infected people with AUROC of 0.96 and 0.86, respectively [77].

Circulating lncRNA as prostate cancer biomarkers has also shown promise. Circulating *PCA3* appears to be a prognostic biomarker, because elevated levels correlate with aggressive disease as indicated by Gleason scores [78, 79]. Levels of plasma *MALAT1* are also associated with prostate cancer. High copy number of these transcripts in plasma of prostate cancer patients achieved a diagnostic sensitivity, specificity, and AUROC of 58.6 %, 84.8 %, and 0.863, respectively [80]. Plasma samples from patients with CLL and multiple myeloma harbor differential levels of *MALAT1*, *HOTAIR*, *GAS5*, *TUG1*, and *linc-p21* [81].

6.7.8 LncRNA in Cancer Tissue-Associated Body Fluids

Besides the peripheral circulation, lncRNAs have been assayed in tissue-associated body fluids from cancer patients. The levels of six lncRNAs including *HOTAIR* and *MALAT1* were different in saliva from patients with oral squamous cell carcinoma compared to healthy controls [82]. The gastric cancer-associated lncRNA *AA174084* has been assayed in gastric juice samples as a diagnostic biomarker. The levels of this lncRNA were significantly higher in cancer patient samples than samples from healthy controls and patients with other gastric mucosal lesions. While *AA174084* is nondiscriminatory when assayed in plasma, its differential levels in gastric juice achieved a sensitivity, specificity, and AUROC of 46 %, 93 %, and 0.848, respectively [50]. The lncRNA that has been extensively studied in urinary samples as prostate cancer biomarker is *PCA3*, which is currently approved by the FDA as an ancillary prostate cancer diagnostic biomarker (Progensa® PCA3, Hologic). Two independent meta-analyses of the numerous urinary studies of *PCA3* reveal identical performances. Xue et al. achieved summary sensitivity, specificity, and AUROC of 62 %, 75 %, and 0.75, respectively [83], while corresponding parameters by Hu et al. were 57 %, 71 %, and 0.7118, respectively [84].

6.7.9 LncRNA in Cancer-Derived Extracellular Vesicles

Although we are yet to fully understand their biogenesis, lncRNAs may also be selectively loaded into extracellular vesicles and used for intercellular communications, similar to miRNAs. Using deep sequencing, Huang et al. characterized RNA content in exosomes from plasma obtained from healthy volunteers and revealed that lncRNA constituted 3.36 % of the entire exosomal RNA content in this cohort [85]. Indeed, *PCA3* had previously been localized in urinary exosomes from prostate cancer patients [86].

Exosomal lncRNA as cancer biomarkers have been demonstrated. In gastric cancer patients, the lncRNA *LINC00152* levels show statistically insignificant differences between plasma samples and plasma-derived exosomes, indicating that almost all this lncRNA is circulating in exosomes and not as free or naked RNA [87]. As a diagnostic biomarker of gastric cancer, *LINC00152* had a dismal sensitivity of 48.1 %, with a much better specificity of 85.2 % and AUROC of 0.66. Also, exosomal lnc*RNA-p21* is elevated in men with prostate cancer, and this could discriminate men with prostate cancer from those with benign prostatic hyperplasia [88]. Exosomal lncRNA may be a prognostic biomarker as well. Serum-derived exosomal *HOTAIR* is elevated in patients with laryngeal squamous cell carcinoma who had lymph node metastasis compared to patients without. The increasing levels of exosomal *HOTAIR* also mirrored disease progression.

Exosomal lncRNAs may have similar functions to those of miRNA. Thus, exosomal lncRNAs are involved in intercellular communications that mediate

transfer of chemoresistance and modulation of tumor microenvironment to enhance metastasis. Two examples are provided using in vitro HCC cell lines. The long intergenic ncRNA ROR (*linc-ROR*) confers sorafenib chemoresistance to HepG2 HCC cell line. Exosomes from chemoresistant HepG2 cells contain *linc-ROR*. Non-chemoresistant HepG2 cells acquired resistant phenotypes upon being treated with exosomes derived from chemoresistant HepG2 cells. Interestingly, TGFβ signaling may modulate lncRNA content in exosomes from HepG2 cells, because lncRNA content changed upon in vitro TGFβ stimulation of these cells [89]. In another study, the CD90+ HCC cell line was demonstrated to secrete exosomes containing *HULC*, *linc-ROR*, *HOTAIR*, and *H19*. In vitro coculture experiments demonstrated the rapid uptake of CD90+-derived exosomes by endothelial cells, leading to phenotypic changes. These endothelial cells with CD90+ exosomes not only increased VEGF and VEGFR-1 expression but also reorganized their cytoskeleton to become tubular-like in shapes, suggestive of the proangiogenic effects of these exosomes. There were also changes in adhesion molecule expression on these endothelial cells that could be part of the metastatic cascade [90].

6.8 Summary

- Much of the human genome encodes transcripts that are not translated into proteins. These non-protein-coding RNAs are referred to as noncoding RNA (ncRNA).
- NcRNAs are of two types, short (<200 nt) and long (>200 nt) ncRNA.
- Apart from the nuclear genome, the mitochondrial genome also produces ncRNAs, whose functions are yet to be clarified.
- A class of small ncRNA that has been extensively studied is miRNAs. MiRNAs are ~20 nt long. They have profound epigenetic control on gene expression.
- Some miRNAs target and inactivate oncogenes to prevent cancer from developing or progressing and are called tumor suppressor miRNAs (tumor suppressormirs). Other miRNAs (oncomirs) target tumor suppressor genes to promote the development of cancer. However, this dichotomized function is contextual, because the same miRNA may function as an oncomir or tumor suppressormir depending on the type of tumor.
- MiRNAs have plethoric and redundant functions. They are aberrantly expressed in tumor tissues, and their circulating levels also differ between cancer and healthy individuals. They therefore serve as potential cancer biomarkers.
- Although numerically more abundant than miRNA, lncRNAs are less well understood. However, emerging evidence show they perform similar functions in gene regulation. It appears though that the functions of lncRNAs are more in the nucleus than cytoplasm.
- LncRNAs also demonstrate aberrant expression in tumors and can be measured in the circulation and other body fluids.

References

1. Kapranov P, St Laurent G, Raz T, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. BMC Biol. 2010;8:149.
2. Villegas J, Burzio V, Villota C, et al. Expression of a novel non-coding mitochondrial RNA in human proliferating cells. Nucleic Acids Res. 2007;35:7336–47.
3. Burzio VA, Villota C, Villegas J, et al. Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. Proc Natl Acad Sci USA. 2009;106:9430–4.
4. Landerer E, Villegas J, Burzio VA, et al. Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. Cell Oncol (Dordr). 2011;34:297–305.
5. Rivas A, Burzio V, Landerer E, et al. Determination of the differential expression of mitochondrial long non-coding RNAs as a noninvasive diagnosis of bladder cancer. BMC Urol. 2012;12:37.
6. Villota C, Campos A, Vidaurre S, et al. Expression of mitochondrial non-coding RNAs (ncRNAs) is modulated by high risk human papillomavirus (HPV) oncogenes. J Biol Chem. 2012;287:21303–15.
7. Rackham O, Shearwood AM, Mercer TR, et al. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA. 2011;17:2085–93.
8. Ro S, Ma HY, Park C, et al. The mitochondrial genome encodes abundant small noncoding RNAs. Cell Res. 2013;23:759–74.
9. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75:843–54.
10. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 1993;75:855–62.
11. Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403:901–6.
12. Pasquinelli AE, Reinhart BJ, Slack F, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000;408:86–9.
13. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet. 2004;5:396–400.
14. Calin GA, Liu CG, Sevignani C, et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc Natl Acad Sci USA. 2004;101:11755–60.
15. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci USA. 2004;101:2999–3004.
16. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435:834–8.
17. Lawrie CH, Gal S, Dunlop HM, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol. 2008;141:672–5.
18. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56:1733–41.
19. Kim B, Lee JH, Park JW, et al. An essential microRNA maturing microprocessor complex component DGCR8 is up-regulated in colorectal carcinomas. Clin Exp Med. 2014;14:331–6.
20. Kumar MS, Lu J, Mercer KL, et al. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet. 2007;39:673–7.
21. van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. Nat Rev Cancer. 2011;11:644–56.
22. Suzuki HI, Yamagata K, Sugimoto K, et al. Modulation of microRNA processing by p53. Nature. 2009;460:529–33.
23. Davis BN, Hilyard AC, Nguyen PH, et al. Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. Mol Cell. 2010;39:373–84.

24. Davis BN, Hilyard AC, Lagna G, Hata A. SMAD proteins control DROSHA-mediated microRNA maturation. Nature. 2008;454:56–61.
25. Walz AL, Ooms A, Gadd S, et al. Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. Cancer Cell. 2015;27:286–97.
26. Diaz-Garcia CV, Agudo-Lopez A, Perez C, et al. DICER1, DROSHA and miRNAs in patients with non-small cell lung cancer: implications for outcomes and histologic classification. Carcinogenesis. 2013;34:1031–8.
27. Merritt WM, Lin YG, Han LY, et al. Dicer, Drosha, and outcomes in patients with ovarian cancer. N Engl J Med. 2008;359:2641–50.
28. Melo SA, Moutinho C, Ropero S, et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. Cancer Cell. 2010;18:303–15.
29. Zhu DX, Fan L, Lu RN, et al. Downregulated Dicer expression predicts poor prognosis in chronic lymphocytic leukemia. Cancer Sci. 2012;103:875–81.
30. Lin RJ, Lin YC, Chen J, et al. microRNA signature and expression of Dicer and Drosha can predict prognosis and delineate risk groups in neuroblastoma. Cancer Res. 2010;70:7841–50.
31. Chiosea S, Jelezcova E, Chandran U, et al. Up-regulation of dicer, a component of the MicroRNA machinery, in prostate adenocarcinoma. Am J Pathol. 2006;169:1812–20.
32. Faber C, Horst D, Hlubek F, Kirchner T. Overexpression of Dicer predicts poor survival in colorectal cancer. Eur J Cancer. 2011;47:1414–9.
33. Stratmann J, Wang CJ, Gnosa S, et al. Dicer and miRNA in relation to clinicopathological variables in colorectal cancer patients. BMC Cancer. 2011;11:345.
34. Hill DA, Ivanovich J, Priest JR, et al. DICER1 mutations in familial pleuropulmonary blastoma. Science. 2009;325:965.
35. Kosaka N, Iguchi H, Yoshioka Y, et al. Secretory mechanisms and intercellular transfer of microRNAs in living cells. J Biol Chem. 2010;285:17442–52.
36. Mittelbrunn M, Gutierrez-Vazquez C, Villarroya-Beltri C, et al. Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. Nat Commun. 2011;2:282.
37. Kogure T, Lin WL, Yan IK, et al. Intercellular nanovesicle-mediated microRNA transfer: a mechanism of environmental modulation of hepatocellular cancer cell growth. Hepatology. 2011;54:1237–48.
38. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435:828–33.
39. Bensen JT, Tse CK, Nyante SJ, et al. Association of germline microRNA SNPs in pre-miRNA flanking region and breast cancer risk and survival: the Carolina Breast Cancer Study. Cancer Causes Control. 2013;24:1099–109.
40. Fabbri M, Ivan M, Cimmino A, et al. Regulatory mechanisms of microRNAs involvement in cancer. Expert Opin Biol Ther. 2007;7:1009–19.
41. Chou CH, Chang NW, Shrestha S, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2016;44:D239–47.
42. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611–22.
43. Zhang L, Hou D, Chen X, et al. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. Cell Res. 2012;22:107–26.
44. Kuhlmann JD, Baraniskin A, Hahn SA, et al. Circulating U2 small nuclear RNA fragments as a novel diagnostic tool for patients with epithelial ovarian cancer. Clin Chem. 2014;60:206–13.
45. Liao J, Yu L, Mei Y, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol Cancer. 2010;9:198.
46. Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol. 2001;18:1611–30.

47. Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA. 2009;106:11667–72.
48. Dinger ME, Amaral PP, Mercer TR, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 2008;18:1433–45.
49. Poliseno L, Salmena L, Zhang J, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465:1033–8.
50. Shao Y, Ye M, Jiang X, et al. Gastric juice long noncoding RNA used as a tumor marker for screening gastric cancer. Cancer. 2014;120:3320–8.
51. Xie H, Ma H, Zhou D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. Biomed Res Int. 2013;2013:136106.
52. Yuan SX, Yang F, Yang Y, et al. Long noncoding RNA associated with microvascular invasion in hepatocellular carcinoma promotes angiogenesis and serves as a predictor for hepatocellular carcinoma patients' poor recurrence-free survival after hepatectomy. Hepatology. 2012;56:2231–41.
53. Yang Z, Zhou L, Wu LM, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011;18:1243–50.
54. Wang XS, Zhang Z, Wang HC, et al. Rapid identification of UCA1 as a very sensitive and specific unique marker for human bladder carcinoma. Clin Cancer Res. 2006;12:4851–8.
55. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999;59:5975–9.
56. de Kok JB, Verhaegh GW, Roelofs RW, et al. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res. 2002;62:2695–8.
57. Zhang HM, Yang FQ, Chen SJ, et al. Upregulation of long non-coding RNA MALAT1 correlates with tumor progression and poor prognosis in clear cell renal cell carcinoma. Tumour Biol. 2015;36:2947–55.
58. Zheng HT, Shi DB, Wang YW, et al. High expression of lncRNA MALAT1 suggests a biomarker of poor prognosis in colorectal cancer. Int J Clin Exp Pathol. 2014;7:3174–81.
59. Cai B, Wu Z, Liao K, Zhang S. Long noncoding RNA HOTAIR can serve as a common molecular marker for lymph node metastasis: a meta-analysis. Tumour Biol. 2014;35:8445–50.
60. Kogo R, Shimamura T, Mimori K, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011;71:6320–6.
61. Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. Br J Cancer. 2013;108:2419–25.
62. Liu Y, Pan S, Liu L, et al. A genetic variant in long non-coding RNA HULC contributes to risk of HBV-related hepatocellular carcinoma in a Chinese population. PLoS One. 2012;7, e35145.
63. Jendrzejewski J, He H, Radomska HS, et al. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. Proc Natl Acad Sci USA. 2012;109:8646–51.
64. Padua Alves C, Fonseca AS, Muys BR, et al. Brief report: The lincRNA Hotair is required for epithelial-to-mesenchymal transition and stemness maintenance of cancer cell lines. Stem Cells. 2013;31:2827–32.
65. Zhang J, Zhang P, Wang L, et al. Long non-coding RNA HOTAIR in carcinogenesis and metastasis. Acta Biochim Biophys Sin Shanghai. 2014;46:1–5.
66. Liu YW, Sun M, Xia R, et al. LincHOTAIR epigenetically silences miR34a by binding to PRC2 to promote the epithelial-to-mesenchymal transition in human gastric cancer. Cell Death Dis. 2015;6:e1802.
67. Pickard MR, Williams GT. Regulation of apoptosis by long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy. Breast Cancer Res Treat. 2014;145:359–70.

68. Pickard MR, Mourtada-Maarabouni M, Williams GT. Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. Biochim Biophys Acta. 1832;2013:1613–23.
69. Yang MH, Hu ZY, Xu C, et al. MALAT1 promotes colorectal cancer cell proliferation/migration/invasion via PRKA kinase anchor protein 9. Biochim Biophys Acta. 1852;2015:166–74.
70. Hu L, Wu Y, Tan D, et al. Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma. J Exp Clin Cancer Res. 2015;34:7.
71. Hirata H, Hinoda Y, Shahryari V, et al. Long noncoding RNA MALAT1 promotes aggressive renal cell carcinoma through Ezh2 and interacts with miR-205. Cancer Res. 2015;75:1322–31.
72. Hung CL, Wang LY, Yu YL, et al. A long noncoding RNA connects c-Myc to tumor metabolism. Proc Natl Acad Sci USA. 2014;111:18697–702.
73. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142:409–19.
74. Redon S, Zemp I, Lingner J. A three-state model for the regulation of telomerase by TERRA and hnRNPA1. Nucleic Acids Res. 2013;41:9117–28.
75. Li H, Yu B, Li J, et al. Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. Oncotarget. 2014;5:2318–29.
76. Tang J, Jiang R, Deng L, et al. Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. Oncotarget. 2015;6:4505–15.
77. Lu J, Xie F, Geng L, et al. Investigation of serum lncRNA-uc003wbd and lncRNA-AF085935 expression profile in patients with hepatocellular carcinoma and HBV. Tumour Biol. 2015;36:3231–6.
78. Merola R, Tomao L, Antenucci A, et al. PCA3 in prostate cancer and tumor aggressiveness detection on 407 high-risk patients: a National Cancer Institute experience. J Exp Clin Cancer Res. 2015;34:15.
79. Chevli KK, Duff M, Walter P, et al. Urinary PCA3 as a predictor of prostate cancer in a cohort of 3,073 men undergoing initial prostate biopsy. J Urol. 2014;191:1743–8.
80. Ren S, Liu Y, Xu W, et al. Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. J Urol. 2013;190:2278–87.
81. Isin M, Ozgur E, Cetin G, et al. Investigation of circulating lncRNAs in B-cell neoplasms. Clin Chim Acta. 2014;431:255–9.
82. Tang H, Wu Z, Zhang J, Su B. Salivary lncRNA as a potential marker for oral squamous cell carcinoma diagnosis. Mol Med Rep. 2013;7:761–6.
83. Xue WJ, Ying XL, Jiang JH, Xu YH. Prostate cancer antigen 3 as a biomarker in the urine for prostate cancer diagnosis: a meta-analysis. J Cancer Res Ther. 2014;10(Suppl):C218–21.
84. Hu CC, Gan P, Zhang RY, et al. Identification of prostate cancer LncRNAs by RNA-Seq. Asian Pac J Cancer Prev. 2014;15:9439–44.
85. Huang X, Yuan T, Tschannen M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics. 2013;14:319.
86. Nilsson J, Skog J, Nordstrand A, et al. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. Br J Cancer. 2009;100:1603–7.
87. Li Q, Shao Y, Zhang X, et al. Plasma long noncoding RNA protected by exosomes as a potential stable biomarker for gastric cancer. Tumour Biol. 2015;36:2007–12.
88. Isin M, Uysaler E, Ozgur E, et al. Exosomal lncRNA-p21 levels may help to distinguish prostate cancer from benign disease. Front Genet. 2015;6:168.
89. Takahashi K, Yan IK, Kogure T, et al. Extracellular vesicle-mediated transfer of long non-coding RNA ROR modulates chemosensitivity in human hepatocellular cancer. FEBS Open Bio. 2014;4:458–67.
90. Conigliaro A, Costa V, Lo Dico A, et al. CD90+ liver cancer cells modulate endothelial cell phenotype through the release of exosomes containing H19 lncRNA. Mol Cancer. 2015;14:155.

Chapter 7
Mitochondrial Genome Changes as Cancer Biomarkers in Circulation

Key Topics

- Features of mitochondrial genome (mtgenome)
- Mitochondrial genetics and cancer
- Mitochondrial dysfunctions and cancer
- Mtgenome changes in circulation of cancer patients
- Mtgenome changes in tissue-associated body fluids

Key Points

- The mitochondrion is the master of the cellular orchestra that harmonizes the hydra-headed functions central to malignant transformation.
- Similar to the nuclear genome, mtgenome changes are demonstrable biomarkers in body fluids of cancer patients.
- The polyploid (high copy number) nature of the mtgenome should make it attractive as an analytically sensitive molecule to assay in body fluids.

7.1 Introduction

Mitochondrial functions contribute immensely to cellular homeostasis. Not only do they supply most of the energy required for cellular functions, but also they serve numerous other roles important in cellular transformation and cancer development. These functions include their roles in regulating metabolism, apoptosis, the cell cycle, growth and differentiation, as well as many signaling networks. There is a bidirectional communication between the mitochondrion and the nucleus, which is one mechanism exploited for efficient tumor progression.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_7

The mitochondrion has its own semiautonomous minuscule genome, the mitochondrial genome (mtgenome). This genome is the remnant of the α-proteobacterium (now mitochondrion) during the symbiotic relationship with the primitive eukaryotic cell, whereby most of the proteobacterial genes were downloaded onto nuclear chromosomes. The mtgenome expresses 13 proteins required for mitochondrial functions as well as ribosomal and transfer RNAs required for mitochondrial translation. Unlike the nuclear genome, the mtgenome lacks histone and other protective proteins and hence is relatively naked. Yet, this molecule exists in the mitochondrial matrix, where a substantial amount of cellular ROS is produced as by-products of the electron transport chain (ETC). Thus, the mtgenome sustains lots of mutations (has a high mutation rate), but these are often compensated for by the presence of many mitochondria and mtgenomes per cell, as well as efficient free radical scavenging systems. Notwithstanding, a critical amount of mutations can manifest as disease. Thus, there are several diseases with mitochondrial DNA defects as the primary cause. It is also becoming evident that cancer cells have altered mtgenomes. While mostly studied in tissue and biopsy samples, mtgenome alterations in circulation and other body fluids have been pursued.

Given the low tumor cell numbers, and nucleic acid content in circulation and other body fluids, the high copy number of the mtgenomes per cell and its noted early imprints in cancer make this molecule more sensitive than nuclear targets to measure. In spite of the limited data available, mtgenome changes in body fluids from cancer patients have been demonstrated in a number of cancers, with some potential clinical relevance.

7.2 The Mitochondrial Genome

Mitochondria are specialized eukaryotic organelles acquired several billion years ago by the eukaryotic cell in a symbiotic relationship with an α-proteobacterium. This relationship led to functional specialization such that the mitochondrion became the site of most energy production required for normal cell physiology. In addition to being the only organelle with its own genome (the mitochondrial genome, mtgenome), the mitochondrion has several unique features worthy of note:

- The mtgenome is small and circular in nature, which contrasts with the large and linear nuclear genome.
- The mitochondrial genes are compactly organized without intervening introns as seen in the nuclear genome; hence, there is no splicing of primary mitochondrial transcripts. The polycistronic transcript is processed to release mitochondrial genes.
- It is primarily inherited in a uniparental fashion, being passed on from mother to offspring.

- The genomes are polyploidy, with each cell possessing hundreds to thousands of mtgenomes.
- As a genome with such high copy number, alterations in some genomes give rise to the concept of heteroplasmy, which refers to the coexistence of mutated and wild-type genomes in the same organelle or cell.
- For mitochondrial disease to manifest, mutant copies must outweigh wild-type genomes in proportion, and this is referred to as the threshold effect. This level is the critical percentage of mutant genomes needed for manifestation of clinical signs and symptoms.
- This genome has a high mutation rate as a consequence of its residence in an organelle that generates most of the reactive oxygen species in the cell.
- The translation of mitochondrial mRNAs relies on mitochondrial-specific genetic code that is slightly different from the nuclear genetic code.
- Finally, mitochondria segregate in a stochastic fashion to daughter cells during cell division.

All these unique features have implications in mitochondrial diseases and genetic technologies such as cloning and production of stem cells.

Nass and Nass [1, 2] first reported on the mitochondrion having its own genetic material, and the Sanger's group [3] sequenced this genome nearly two decades later. Structurally, the mtgenome is a compact prokaryotic-like double-stranded molecule composed of 16,568 nucleotide base pairs (Fig. 7.1). This genome encodes 37 molecules involved in its functions. These molecules are the 2 rRNAs and 22 tRNAs used for mitochondrial translation and the 13 proteins that complement the nuclear-encoded proteins involved in the electron transport chain and oxidative phosphorylation. Except for mitochondrial-encoded *NADH dehydrogenase 6* (*MTND6*), all the 13 mitochondrial polypeptides are encoded using the heavy or H-strand. Additionally, the H-strand encodes 14 of the 22 tRNAs and the 2 rRNAs. *NADH dehydrogenase 6* and the remaining eight tRNAs are encoded by the light or L-strand. There are two noncoding mitochondrial regions. The first is a large segment referred to as the D-loop, which is composed of 1121 bp and extends from nucleotide position (np) 16,024 to np 576. This region houses regulatory elements for mtgenome transcription and translation. Within the L-strand origin of replication is another noncoding region composed of just 30 nucleotides. The structural complexity of this genome belies its involvement in many diseases including cancer.

7.3 The Electron Transport Chain and ROS Production

The mitochondrial respiratory activity produces most of the energy required for cellular functions. The respiratory machinery, which resides in the inner mitochondrial membrane, is organized into five protein complexes (Fig. 7.2). The mitochondrial genome encodes 7 (NADH dehydrogenase subunits 1–6 and 4 L) of the ~46

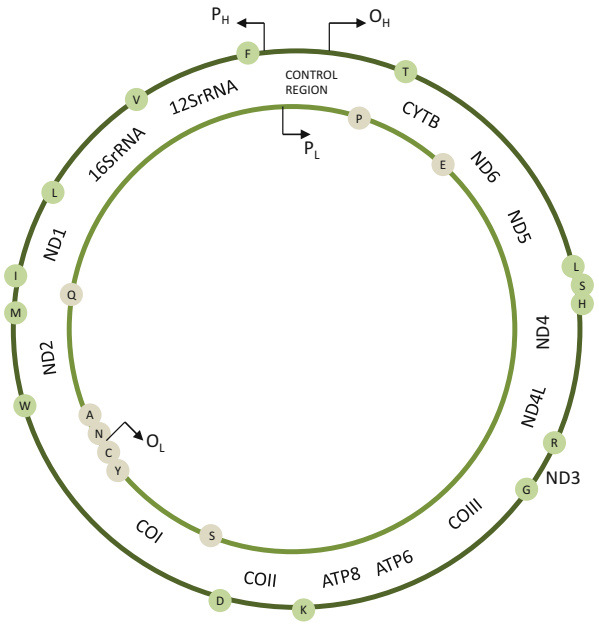

Fig. 7.1 The mitochondrial DNA molecule. The mtgenome is a circular double-stranded mtDNA molecule composed of 13 protein-coding genes that are transcribed and translated within the mitochondrion. Transcription factors interact with the heavy-strand (P_H) and light-strand (P_L) promoters and induce the production of polycistronic transcripts, which are processed into 22 tRNAs (*small circles*), 2 rRNAs, and 13 protein complements of the nuclear-encoded OXPHOS proteins. Replication of the molecule starts at the heavy-strand origin of replication (O_H) and proceeds in a clockwise direction producing a daughter H-strand molecule. After copying about 66 % of the H-stand, the light-strand origin of replication (O_L) is encountered, and the parental H-strand is displaced exposing the O_L, which then commences replication in the opposite direction

subunits of complex I, 1 (cytochrome b) of the 11 subunits of complex III, 3 (cytochrome c oxidase subunits I–III) of the 13 subunits of complex IV, and 2 (ATP synthetase subunits 6 and 8) of the ~18 subunits of complex V. Nuclear genes encode all of the four subunits of complex II and the remaining subunits of the other complexes. In addition to the five complexes, ubiquinone (Coenzyme Q10) and cytochrome c participate in electron shuttle through the respiratory chain.

Energy in the form of ATP is generated from calories derived from food. Acetyl-CoA from glycolysis and beta-oxidation is oxidized in the Krebs cycle to generate electrons in the form of NADH and succinate ($FADH_2$), which are received by complexes I and II, respectively. The electrons are sequentially transported through redox groups to the final acceptor, oxygen, which converts them to water. In the process of electron shuttle, protons are pumped from the matrix to the intermembrane space by complexes I, III, and IV. Complex V then uses the electrochemical gradient established to pump the protons back into the matrix, and this process is coupled with the synthesis of ATP.

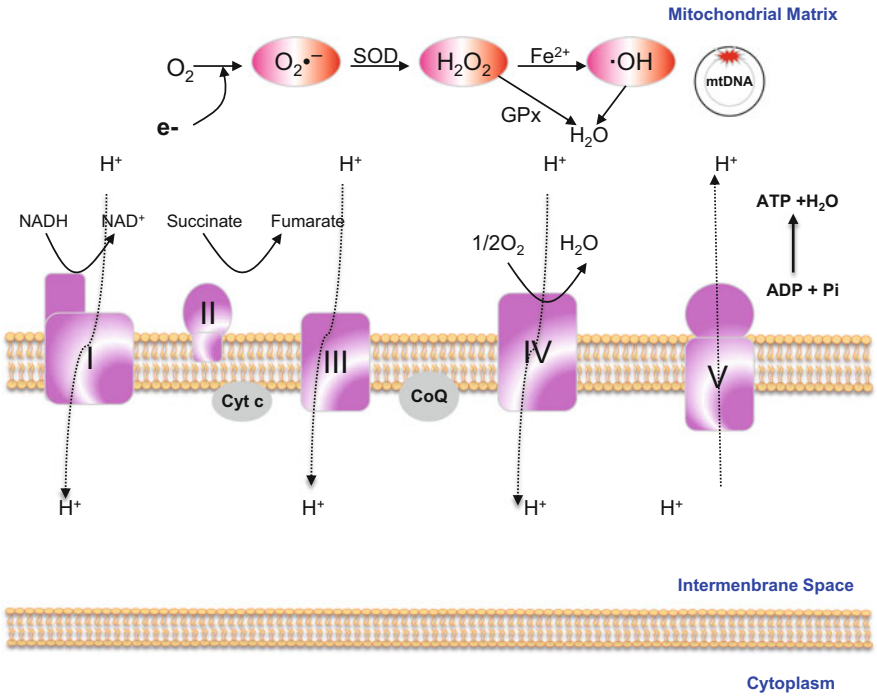

Fig. 7.2 The electron transport chain and ROS production. ATP production by the respiratory chain complex is associated with the generation of ROS as electrons leak and interact with oxygen during electron transport. In normal cells, the levels of the various free radicals are kept in balance by the activities of various antioxidant mechanisms (maintenance of redox homeostasis). However, excess production of free radicals can overwhelm the detoxifying mechanisms leading to damage to nearby mtDNA molecules, mitochondrial membranes, and proteins

The activities of the respiratory chain are inevitably coupled with the generation of superoxide anion ($O_2^{\bullet -}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($^{\bullet}OH$) (Fig. 7.2). Electrons can leak from the respiratory chain and interact with oxygen to produce the first family member, $O_2^{\bullet -}$, which can be converted to H_2O_2 by superoxide dismutase (SOD). The H_2O_2 generated is usually detoxified by the antioxidant mechanisms such as the activities of cytosolic glutathione peroxidase and peroxisomal catalase. However, in the presence of reduced transition metals (e.g., Fe^{2+}), H_2O_2 can be converted to one of the most reactive forms of the free radicals, the $^{\bullet}OH$. Functionally important mtDNA mutations in several diseases including cancer are associated with elevated mitochondrial reactive oxygen species (ROS) production and a vicious cycle of mtDNA damage. Reactive oxygen species production increases in conditions of excess electrons such as will occur in states of increased caloric intake and/or defective respiratory chain activity.

7.4 Mitochondrial Genetics and Cancer

Mitochondrial dysfunction contributes to cellular transformation and maintenance of the hallmarks of the cancer cell (for detailed synthesis of the role of mitochondrial genetics in cancer, consult "Mitochondrial Genetics and Cancer, G. D. Dakubo, Springer 2010). Altered intermediary metabolism, increased oxidative stress, and resistance to apoptosis are key features of cancer cells. Several decades ago, Otto Warburg (1931) observed that in contrast to normal cells that consumed less glucose and produced less lactate when utilizing oxygen to generate energy via the most efficient mode of energy production (the respiratory chain), cancer cells metabolized excess glucose and produced excess lactate. This phenomenon is referred to as aerobic glycolysis or the Warburg effect. In normal cells, pyruvate generated from glycolysis enters the mitochondria and is subsequently oxidized in the Krebs cycle. However, defective cancer cell mitochondria are incapable of removing pyruvate, which therefore accumulates and is converted to lactic acid by lactate dehydrogenase. Excess lactate creates an acidic tumor microenvironment that favors tumor invasiveness. Elucidation of the molecular mechanisms of the Warburg effect is currently being appreciated in cancer biology. This metabolic switch, however, appears to result in part from the inefficient respiratory chain activity, probably from mutations in the mitochondrial genome, and is possibly an adaptive mechanism that permits solid tumors to thrive in adverse environments such as reduced oxygenation.

An important evidence for the involvement of mitochondria in the malignant transformation process also came from the discovery that mutations in nuclear genes that encode mitochondrial proteins were associated with some types of human tumors. Uterine leiomyomas and renal cell carcinomas have been linked to mutations in the mitochondrial enzyme, fumarate hydratase, while mutations in succinate dehydrogenase (*SDH*) subunits B, C, or D, three of the four subunits of the respiratory chain complex II, are associated with paragangliomas. Biochemical evidence suggests that mutations in these genes can lead to at least two major defects. First, the reduced metabolism of succinate and fumarate should result in a buildup of these metabolites in the cytosol, which can stabilize the transcription factor, hypoxia-inducible factor 1 alpha (HIF-1α), that, in turn, can induce the expression of genes involved in tumor progression. Moreover, tumor hypoxic microenvironment could also induce expression of HIF-1α, which can potentially stimulate the switch of tumor cell metabolism toward aerobic glycolysis. Second, the primary function of mitochondrial complex II is to receive and pass electrons from succinate to ubiquinone. Thus, mutations that disrupt the normal functioning of complex II can impede electron flow through the respiratory chain leading to excess electron buildup in mitochondria. Leakage of electrons from the respiratory chain will reduce oxygen to superoxide anion, thus initiating the cascade of ROS production. Because ROS can induce and promote tumor growth, alterations in either mitochondrial complex I or II through mutations in nuclear- and/or mitochondrial-encoded subunits can trigger or enhance tumor development.

The resistance of cancer cells to apoptosis or programmed cell death is one of the intriguing alterations of cancer cell mitochondria. Apoptosis is a homeostatic mechanism by which cells are killed without inducing inflammatory reactions. Intrinsic or mitochondrial apoptotic pathway is initiated by activation of the mitochondrial permeability transition pore (mtPTP), a complex composed of cyclophilin D, BCL2, BAX, inner membrane adenine nucleotide translocator, and outer membrane voltage-dependent anion channel. Activated mtPTP leads to the release of mitochondrial proapoptotic factors including cytochrome c, procaspase 9, SMAC/DIABLO, apoptosis-inducing factor (AIF), and endonuclease G into the cytosol. Cytochrome c interacts with APAF1 and procaspase 9 to form an apoptosome, which then activates caspase 9 and subsequently the executioner caspase (caspase 3), while SMAC/DIABLO activates caspase 3 via inhibition of inhibitors of apoptosis (IAPs). Apoptosis-inducing factor and endonuclease G, however, translocate into the nucleus to induce DNA fragmentation. An important link between mitochondria and cancer is the finding that mitochondria from cancer cells are resistant to induction of mitochondrial outer membrane permeabilization required for the initiation of apoptosis.

The involvement of the mtgenome in cancer biology has been known for several decades. Clayton and Vinograd [4] observed mtDNA dimers and higher oligomers in circulating leukocytes of leukemic patients, but not in normal mature human leukocytes. Interestingly, the levels of these abnormal forms of mtDNA molecules declined with chemotherapy. Following this initial report, several groups have documented mtDNA alterations in other types of human cancers. However, Horton et al. [5] first provided a direct causal role for mtDNA mutations in cancer, with the finding of a 294 bp deletion in NADH dehydrogenase subunit 1 associated with renal cell carcinomas. This deletion was present in 50 % of the mitochondrial genomes of the tumor and was likely to be functionally relevant to the etiology of the tumor, because it resulted in truncated protein product that could cause defective respiratory chain activity. Subsequent to this work, the literature on mtDNA mutations in cancer has expanded considerably. Indeed, it has been proposed that cancer should be added to the list of mitochondrial diseases.

Mitochondrial DNA mutation detection will be important for personalized medicine. This is because the genetic signature of the disease is different from the individual's inherited mitochondrial genetic signature. The comparison of these signature genomes present within a person is used to determine the mutation load and identify early the presence of personalized disease genotype. For example, the genetic species contained by tumor cells arises from the inherited "native" patient's mtDNA and is genetically distinct, and that difference is measurable. The intra-genotype comparison of tissues within the same person rules out potential aging-related mutations in the mitochondrial genome.

7.4.1 Mitochondrial Dysfunction and Cancer Metastasis

Studies of the contributions of mtgenome alterations to cancer biology have primarily relied on genetic experiments involving the use of transmitochondrial cybrids and xenograft mouse models. The classic experiments by Ishikawa et al. [6] involving the induction of metastasis by replacing mitochondria of poorly metastasizing cells with those from highly metastasizing cells are noteworthy. This phenotype was associated with increased oxidative stress and overexpression of genes involved in survival and angiogenesis. Also cybrids with mtDNA mutations demonstrated higher frequency of lung metastasis compared to those with wild-type genomes, and this was partly due to protection of these cells from stress-induced death via PI3K pathway activation [7]. Additional evidence for the involvement of mitochondrial-induced redox stress in carcinogenesis is the finding that the increase in antioxidant enzyme, catalase, in breast cancer mouse model decreased mitochondrial oxidative stress in association with markedly reduced metastatic burden in these mice [8].

Another molecular mechanism by which cancer cells may sustain metastatic potential is through the reversal of the Warburg phenomenon. Metastatic cancer cells in lymph nodes secrete excess hydrogen peroxide to rather induce oxidative stress and hence Warburg effect in normal stromal cells. The excess lactate and ketone bodies produced by the stromal cells are then used to rather reprogram metastatic cancer cells back to oxidative mitochondrial metabolism [9]. Additional support for this cooperation between stromal cells and metastatic cancer cells was the finding that in triple-negative breast cancer tissue microarrays, high stromal staining of the glycolytic marker, MCT4, correlated with decreased patient survival. Surprisingly, however, tumor positivity for MCT4 had no prognostic significance [10]. Also of interest to cancer biology is the finding that proliferative CD133-expressing cancer stem cells rely mainly on glycolysis for their energy requirements, which is associated with increased Glut-1 expression [11]. Targeting this transporter may be a usefully therapeutic approach.

7.5 Mitochondrial Genome Changes in Body Fluids of Cancer Patients

Analyses of mtDNA changes in body fluids can be used for early cancer detection, monitoring of disease progression and response to therapy, or predictive stratification of patients for treatment. It is becoming obvious that mitochondrial genome alterations may modulate the behavior of, or be associated with, some tumors. For example, mtDNA content changes correlate with clinicopathologic features of head and neck, ovarian, and endometrial cancers, D-loop mutations are associated with poor outcomes in breast and colorectal cancers, and levels of circulating

Table 7.1 Mitochondrial genome changes assayed in various body fluids of cancer patients

Body fluids	Primary tumor	MtDNA changes in body fluids
Plasma/serum	Liver cancer	Somatic mutations
	Melanoma	Somatic mutations
	Colorectal cancer	Somatic mutations
	Prostate cancer	Somatic mutations
	Prostate cancer	Increased mtDNA and mtRNA associated with poor outcomes
	Ovarian	High mtDNA in cancer patients
	Testicular	High mtDNA in cancer patients
Saliva	Head and neck cancer	Somatic mutations
	Head and neck cancer	Increased mtDNA content in cancer patients; decreased mtDNA content with therapy
Bronchoalveolar lavage/sputum	Lung cancer	Somatic mutations
Nipple aspirate fluid	Breast cancer	Somatic mutations
Urine	Bladder cancer	Somatic mutations
	Renal Cancer	Somatic mutations
	Prostate cancer	Somatic mutations
CSF	Medulloblastoma	Somatic mutations
Pancreatic juice	Pancreatic cancer	Somatic mutations

mitochondrial genomes appear to have prognostic significance in prostate cancer and modulate with therapeutic responses [12].

The elegant demonstration of mtDNA mutations in cancer by the Vogelstein's group, followed by the seminal work by the Sidransky's group of detecting mitochondrial genome changes in body fluids, convincingly proofed that analysis of mtDNA mutations in body fluids is an easy, feasible, and practical means of tracking early genomic alterations in cancer [13, 14]. Currently, several other investigators have shown that tumor-specific mutations in mtDNA are detectable in body fluids representative of the specific organ associated with the cancer. Mitochondrial genome changes detected in body fluids of various cancers are summarized in Table 7.1 and discussed herein.

7.5.1 Tumor mtDNA in Circulation

The possibility of detecting tumor-derived mtDNA mutations in circulating body fluids has been explored in melanoma, colorectal, and liver cancers. Takeuchi et al. examined mtDNA mutations in the D-loop region of melanoma samples and matched plasma [15]. Forty-two percent of the tumors had D-loop mutations, of which two mutations were detected in matched plasma samples.

Mitochondrial genome changes are detectable in circulation of colorectal, liver, prostate, testicular, and epithelial ovarian cancers. Following a successful demonstration of *KRAS* and *TP53* mutations in plasma of colorectal cancer patients, Hibi and coworkers extended their studies to include mtDNA D-loop alterations [16]. Analysis of seven colorectal cancer patients with mtDNA mutations identified only one mutation in matched serum sample. Several groups have demonstrated mitochondrial genome mutations in colorectal cancer. Therefore, analysis of exfoliated colonocytes in stool samples may increase the sensitivity of detecting colorectal tumor-derived mtDNA mutations. Nomoto et al. examined mtDNA alterations in hepatocellular carcinoma (HCC) and matched plasma samples [17]. Tumor-specific mtDNA mutations were successfully detected in 80 % of plasma samples. In a follow-up study, Okochi et al. uncovered tumor-specific mutations in only 33 % of plasma samples from HCC patients [18]. These studies imply freely circulating mtDNA contain tumor-specific genomic alterations.

Employing the sensitive mutant-specific mismatch ligation assay, Jeronimo et al. were able to detect all three primary prostate cancer mutations in matched plasma [19]. Because the majority of men diagnosed with prostate cancer have a favorable outcome, an important requirement for prostate cancer diagnosis and management is the ability to predict tumor behavior. It has been shown that the levels of mitochondrial nucleic acids in circulating blood might have a prognostic utility in managing prostate cancer patients [20, 21]. A study of men with advanced prostate cancer found that those with high levels of plasma mitochondrial DNA and RNA had a poor 2-year survival compared to men with low levels. Importantly, the levels of mitochondrial RNA appeared to be an independent predictor of survival. Ellinger et al. amplified both large and short cell-free circulating mtDNA in 100 patients with prostate cancer and 18 with benign prostatic hyperplasia [20]. Similar levels were observed in both groups. However, patients with short mtDNA fragments had an increased risk of biochemical recurrence after prostatectomy.

Ellinger et al. further evaluated the potential utility of cell-free circulating nucleic acids in the management of patients with testicular cancer [22]. In an earlier study, *ACTB* DNA fragments were analyzed and DNA integrity defined as the ratio of the large to the smaller fragment assessed. The study subjects comprised of 39 patients with seminomas, 35 with nonseminoma testicular cancer, and 35 healthy controls. Cell-free DNA fragments were significantly increased in circulation of patients compared to controls. In addition, DNA integrity was significantly decreased in patients compared to controls. This assay could differentiate patients from healthy controls with a sensitivity and specificity of 87 % and 97 %, respectively. The levels of circulating mtDNA also tended to be more elevated in high-grade tumors. In a follow-up study of these same subjects, mtDNA fragments comprised of 79 bp and 220 bp fragments in *12SrRNA* were amplified, and DNA integrity defined as previously (levels of the ratio of 220 bp fragments to 79 bp fragments). Again, the levels of both fragments were significantly elevated in cancer patients compared to controls. However, mtDNA integrity was similar in both groups. Apoptotic mechanism involves nucleosomal fragmentation with the

release of small DNA fragments. It is unlikely to be the mechanism of mtDNA fragmentation, given the absence of histone proteins. The assay was accurate at distinguishing between patients and controls with a sensitivity of 59.5 % and a specificity of 94.3 % with the area under a receiver operating characteristic curve of 0.787. Importantly, in both nuclear and mitochondrial DNA analysis, 31 patients with normal levels of testicular cancer serum markers (alpha-fetoprotein, human chorionic gonadotropin, placental alkaline phosphatase, and lactate dehydrogenase) could be distinguished from healthy subjects with a very high degree of accuracy. Thus, cell-free circulating nucleic acids appear to be valuable in testicular cancer detection, especially in people with normal levels of established testicular cancer markers.

Free circulating nuclear and mitochondrial DNA content in serum and plasma were analyzed in 104 female patients who had epithelial ovarian cancer, benign epithelial ovarian tumors, and endometriosis [23]. Patients with epithelial ovarian cancer had significantly high nuclear and mtDNA fragments in their plasma compared to healthy control subjects and women with benign ovarian tumors. For patients with endometriosis, only cell-free mtDNA could significantly separate them from epithelial ovarian cancer subjects. These findings indicate a diagnostic potential for measurement of circulating nucleic acids in ovarian cancer diagnostics.

7.5.2 Tumor mtDNA in Tissue-Associated Body Fluids

Mitochondrial DNA alterations in cancer are also present in proximal fluids associated with the organ or tissue of cancer origin. These mtgenome changes from the cancer are usually more enriched in these fluids than in the peripheral circulation. Although the acquisition of some of these fluids is invasive (e.g., CSF), others such as urine and saliva are noninvasive and hence are attractive sample sources for disease management.

7.5.2.1 Saliva and Head and Neck Cancer

As an established risk factor for head and neck, esophageal, lung, and bladder cancers, the toxins in tobacco smoke can damage mtDNA and possibly contribute to the development of these cancers. Mitochondrial DNA content changes were measured in saliva of smokers and nonsmokers. Saliva from smokers, including those who had stopped smoking, contained significantly high levels of mtDNA compared to those from people who never smoked [24].

Mitochondrial DNA copy number or content alterations have been measured in head and neck cancers and matched saliva using quantitative polymerase chain reaction (qPCR). In an initial study of head and neck tumor tissues of various clinical grades reminiscence of disease progression, it was found that increased

mtDNA content correlated positively with histopathologic grade [25]. In a follow-up study, oral rinse samples were collected from 94 patients with head and neck squamous cell carcinoma (HNSCC) as well as 656 individuals without cancer for mtDNA content analysis [26]. Mitochondrial DNA levels were significantly higher in saliva from cancer patients than controls. Importantly, primary tumors also had higher levels of mtDNA than pretreatment saliva, suggesting the elevated levels of mtDNA in saliva of cancer patients were tumor derived. In another study, the effect of treatment on mtDNA levels was examined in 79 patients with HNSCC [27]. Analysis of pre- and posttreatment saliva revealed a significant decrease in mtDNA content in posttreatment samples. In addition, saliva from patients who had radiation treatment and those who never smoked contained lower levels of mitochondrial genomes than in saliva from nonirradiated and smoking counterparts. Thus, the harmful effects of smoking on mtDNA appear to be chronic and irreversible.

Mutations in the mitochondrial genome are present in early-stage head and neck cancers. Microarray-based resequencing of the entire mitochondrial genome was conducted on a panel of 83 HNSCC, and mutations were found in nearly half of the tumors [28]. Sequencing of dysplastic adjacent tissues of one tumor identified the same mutation in both tumor and margin samples, suggesting mtDNA mutations occurred early in this particular tumor and could be involved in disease progression. To demonstrate the possible utility of mtDNA mutations as a tool for early cancer detection in clinical samples, Fliss et al. used manual sequencing to study mtDNA mutations in primary head and neck cancers and matched available saliva [13]. Six of the thirteen solid tumors had mutations that were also detected in six of nine saliva samples. MitoChip resequencing of salivary rinses from patients with known tumor mtDNA mutations was studied by Mithani et al. for the possible presence of same mutations in noninvasively collected body fluids [29]. Using a proprietary algorithm, heteroplasmic mutations could be detected at even a dilution of 1:200. This sensitive approach enabled these investigators to uncover tumor mutations in saliva of 77 % of the patients. Thus, mtDNA content changes and mutations can be measured in saliva for diagnosis and management of HNSCC.

7.5.2.2 Bronchoalveolar Fluid, Sputum, and Lung Cancer

Lung cancer has a high morbidity and case fatality rate compared to other cancers. In addition to avoiding tobacco smoke, early detection of small tumors amenable to curative resection will help reduce the fatalities. Because smoking is a strong risk factor for lung cancer, and the genotoxins in tobacco smoke can cause damage to the mitochondrial genome, mtDNA damage and the levels of a 4977 bp common deletion (CD) were measured in BAL fluids from smokers and nonsmokers [30]. Not surprisingly, smokers had a highly significant damage in their mitochondrial genomes compared to nonsmokers. Although the levels of the CD were much higher in smokers than nonsmokers, this difference was insignificant. The study by Fliss et al. also included 14 lung cancer samples and matched BAL [13]. Six of the

fourteen tumors had mutations. Using a sensitive detection method, eight of the ten lung cancer mutations were easily detectable in matched BAL samples. Jakupciak et al. uncovered three somatic mutations in BAL from seven patients with lung cancer [31]. Sputum and/or BAL appear to be a useful source of exfoliated bronchial epithelial cells for mtDNA analysis for early cancer detection in high-risk population.

7.5.2.3 Nipple Aspirate and Ductal Lavage Fluids and Breast Cancer

Although death rates have declined partly due to early detection strategies, breast cancer is still the most commonly diagnosed female cancer and the second cause of all cancer deaths in women. Early detection is critical to achieving high cure rates. Ductal lavage and nipple aspirate fluids (NAFs) offer a practical possibility of early detection of breast cancer. As a proof of principle, free DNA from 26 ductal lavage and 6 NAF samples from 9 *BRCA1* mutation carriers and 5 noncarriers were studied for genomic alterations [32]. Somatic mutations at mtDNA D310 locus were found in three ductal lavage samples from three *BRCA1* mutation carriers and in none of the noncarriers. Four NAF samples were successfully analyzed with one having a mutation that was also present in the ductal lavage sample. Zhu and coworkers examined approximately 25 % of the mitochondrial genome in NAF samples from women with breast cancer and found four identical mutations in both NAF and cancer samples [33]. Importantly, NAF sequences from five tumors that did not have a mutation were also free of mutations. Thus, tumor-specific mtDNA mutations are detectable in ductal lavage and NAF from the corresponding breast. We explored the possibility of using NAF for early breast cancer detection with chip-based resequencing of mtDNA [34]. We were successful in sequencing the entire mitochondrial genome from 19 NAF samples with both microarray and capillary electrophoresis sequencing. Although after clinical investigation all our study participants had benign breast lesions, somatic mtDNA mutations were uncovered in four of the NAF samples. However, it is unknown whether these findings indicate disease progression toward malignancy. As a concept, mitochondrial genome sequencing using NAF is a relatively easy process that can have real clinical translation.

7.5.2.4 Urine and Urogenital Cancer

Urine is a suitable sample for detection and monitoring of urogenital cancers. Fliss et al. were the first to study mtDNA mutations in urine samples and matched bladder cancers [13]. Somatic mutations were observed in 64 % of the primary bladder cancers. Interestingly, the available three matched urine samples contained tumor-derived mutations as well. Using a sensitive and reliable microarray approach, the entire mitochondrial genome was interrogated in five bladder cancer samples and matched body fluids [35]. Tumor-derived heteroplasmic mutations

were easily assayed in four of five urine samples, confirming the ease with which bladder cancer mtDNA mutations can be detected in urine. Jakupciak et al. analyzed urine from three patients with bladder cancer and 12 with renal cancer using the MitoChip [31]. One and four somatic mutations were found in bladder and renal cancer urine samples, respectively. Thus, chip-based detection of cancer in body fluids has a real possibility of clinical translation.

Prostate cancer remains the most commonly diagnosed, and second cause of cancer deaths in American men, despite increased survival rates. The improved outcomes are partly due to early detection of organ-confined tumors from PSA and digital rectal examination screening programs. It is noteworthy that the PSA test lacks accuracy, hence the relentless search for sensitive and specific molecular targets. There is paucity of studies on mitochondrial genome analysis in body fluids of prostate cancer patients; however, the available evidence from studies of primary tumors points to early mtDNA damage in prostate cancer [19], suggesting targeting mitochondrial genome changes in body fluids can serve as a screening tool. Whereas manual sequencing of mtDNA in urine and saliva was feasible for mutation detection in bladder and head and neck cancer patients, respectively [13], this method of sequencing was not possible for detection of prostate cancer mutations in plasma and urine. However, employing the sensitive mismatch ligation assay, all three primary prostate cancer mutations were successfully detected in matched plasma and urine samples [19]. It is possible that enriching urine with prostate-derived nucleic acids by collecting urine after digital rectal examination, prostate massage fluids, or semen will enable application of even less sensitive, but reliable and clinically suitable, methods such as qPCR for prostate cancer early detection.

7.5.2.5 Cerebrospinal Fluid and Brain Tumors

Because cerebrospinal fluid (CSF) is obtained by invasive methods, it is unsuitable as a sample for diagnosis of brain tumors. However, for the purposes of disease monitoring, CSF can be a valuable sample to assay mtDNA changes. Wong et al. investigated the possible utility of mtDNA changes in CSF for prognostic evaluation of patients with medulloblastoma, a pediatric brain tumor [36]. They examined 15 primary tumors, 6 of which had somatic mutations. Seven matched CSF samples (including five with somatic mutations in tumors) were analyzed. A primary tumor mutation in one patient could be demonstrated in CSF at the end of therapy when conventional laboratory methods failed to identify any sign of persistent disease. This patient subsequently relapsed, suggesting the real possibility of using mtDNA changes to monitor disease progression. This finding is important because mtDNA mutations are present in other types of brain tumors [37, 38]. Several other mutations uncovered by Wong et al. especially at the D310 region were different between tumor and CSF, meaning these tumors could be of different clonal origins, consistent with the concept of field cancerization

[36, 39]. Alternatively, clonal evolution necessitated acquisition of novel mtDNA mutations.

7.5.2.6 Other Exfoliated Cells

There are other body fluids that can be obtained for mtDNA analysis, though these have not yet been examined. For example, gastric juice, pancreatic juice, biliary juice, peritoneal fluid, cervical secretions or swab, and menstrual fluid are all potential for development of mtDNA markers for early detection and management of organ-associated cancers. Mitochondrial genome analysis of pancreatic cancer mutations, for example, detected two of four mutations in pancreatic juices [35].

7.6 Summary

- The mitochondrion is a unique organelle. Being of prokaryotic origin, it has its own circular genome, the mitochondrial genome or DNA (mtgenome or mtDNA).
- The mtgenome encodes 2 rRNAs, 22 tRNAs for mitochondrial translation, and 13 proteins involved in the electron transport chain and ATP production.
- The mtgenome lacks protection by histone proteins, and yet it resides in an organelle that generates substantial amounts of ROS as by-products of energy production. Hence, it has a high mutation rate.
- Each cell has several mitochondria and hundreds of mtDNA. So, under mutational stress, some genomes are mutated while others are not. The presence of these two mixed populations of mtDNAs in a cell is referred to as heteroplasmy. For disease manifestation, the mutated mtgenomes must reach a certain percentage referred to as the threshold.
- Mitochondrial DNA mutations are important in cancer initiation, progression, and metastasis.
- Mtgenome mutations and content variations are detectable and quantifiable in blood and other body fluids as diagnostic biomarkers.
- Although still to be established as clinically actionable biomarkers, the high copy number of mtDNA suggests it should be analytically easy with targeting this genome in body fluids.

References

1. Nass S, Nass MM. Intramitochondrial fibers with DNA characteristics. Ii. Enzymatic and other hydrolytic treatments. J Cell Biol. 1963;19:613–29.

2. Nass MM, Nass S. Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. J Cell Biol. 1963;19:593–611.
3. Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457–65.
4. Clayton DA, Vinograd J. Complex mitochondrial DNA in leukemic and normal human myeloid cells. Proc Natl Acad Sci USA. 1969;62:1077–84.
5. Horton TM, Petros JA, Heddi A, et al. Novel mitochondrial DNA deletion found in a renal cell carcinoma. Genes Chromosomes Cancer. 1996;15:95–101.
6. Ishikawa K, Hashizume O, Koshikawa N, et al. Enhanced glycolysis induced by mtDNA mutations does not regulate metastasis. FEBS Lett. 2008;582:3525–30.
7. Kulawiec M, Owens KM, Singh KK. Cancer cell mitochondria confer apoptosis resistance and promote metastasis. Cancer Biol Ther. 2009;8:1378–85.
8. Goh J, Enns L, Fatemie S, et al. Mitochondrial targeted catalase suppresses invasive breast cancer in mice. BMC Cancer. 2011;11:191.
9. Sotgia F, Whitaker-Menezes D, Martinez-Outschoorn UE, et al. Mitochondrial metabolism in cancer metastasis: visualizing tumor cell mitochondria and the "reverse Warburg effect" in positive lymph node tissue. Cell Cycle. 2012;11:1445–54.
10. Witkiewicz AK, Whitaker-Menezes D, Dasgupta A, et al. Using the "reverse Warburg effect" to identify high-risk breast cancer patients: stromal MCT4 predicts poor clinical outcome in triple-negative breast cancers. Cell Cycle. 2012;11:1108–17.
11. Chen XH, Bao YY, Zhou SH, et al. Glucose transporter-1 expression in CD133+ laryngeal carcinoma Hep-2 cells. Mol Med Rep. 2013;8:1695–700.
12. Dakubo GD. Mitochondrial genetics and cancer. Berlin: Springer; 2010.
13. Fliss MS, Usadel H, Caballero OL, et al. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science. 2000;287:2017–9.
14. Polyak K, Li Y, Zhu H, et al. Somatic mutations of the mitochondrial genome in human colorectal tumours. Nat Genet. 1998;20:291–3.
15. Takeuchi H, Fujimoto A, Hoon DS. Detection of mitochondrial DNA alterations in plasma of malignant melanoma patients. Ann N Y Acad Sci. 2004;1022:50–4.
16. Hibi K, Nakayama H, Yamazaki T, et al. Detection of mitochondrial DNA alterations in primary tumors and corresponding serum of colorectal cancer patients. Int J Cancer. 2001;94:429–31.
17. Nomoto S, Yamashita K, Koshikawa K, et al. Mitochondrial D-loop mutations as clonal markers in multicentric hepatocellular carcinoma and plasma. Clin Cancer Res. 2002;8:481–7.
18. Okochi O, Hibi K, Uemura T, et al. Detection of mitochondrial DNA alterations in the serum of hepatocellular carcinoma patients. Clin Cancer Res. 2002;8:2875–8.
19. Jeronimo C, Nomoto S, Caballero OL, et al. Mitochondrial mutations in early stage prostate cancer and bodily fluids. Oncogene. 2001;20:5195–8.
20. Ellinger J, Muller SC, Wernert N, et al. Mitochondrial DNA in serum of patients with prostate cancer: a predictor of biochemical recurrence after prostatectomy. BJU Int. 2008;102:628–32.
21. Mehra N, Penning M, Maas J, et al. Circulating mitochondrial nucleic acids have prognostic value for survival in patients with advanced prostate cancer. Clin Cancer Res. 2007;13:421–6.
22. Ellinger J, Albers P, Perabo FG, et al. CpG island hypermethylation of cell-free circulating serum DNA in patients with testicular cancer. J Urol. 2009;182:324–9.
23. Zachariah RR, Schmid S, Buerki N, et al. Levels of circulating cell-free nuclear and mitochondrial DNA in benign and malignant ovarian tumors. Obstet Gynecol. 2008;112:843–50.
24. Masayesva BG, Mambo E, Taylor RJ, et al. Mitochondrial DNA content increase in response to cigarette smoking. Cancer Epidemiol Biomarkers Prev. 2006;15:19–24.
25. Kim MM, Clinger JD, Masayesva BG, et al. Mitochondrial DNA quantity increases with histopathologic grade in premalignant and malignant head and neck lesions. Clin Cancer Res. 2004;10:8512–5.
26. Jiang WW, Masayesva B, Zahurak M, et al. Increased mitochondrial DNA content in saliva associated with head and neck cancer. Clin Cancer Res. 2005;11:2486–91.

27. Jiang WW, Rosenbaum E, Mambo E, et al. Decreased mitochondrial DNA content in posttreatment salivary rinses from head and neck cancer patients. Clin Cancer Res. 2006;12:1564–9.
28. Zhou S, Kachhap S, Sun W, et al. Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. Proc Natl Acad Sci USA. 2007;104:7540–5.
29. Mithani SK, Smith IM, Zhou S, et al. Mitochondrial resequencing arrays detect tumor-specific mutations in salivary rinses of patients with head and neck cancer. Clin Cancer Res. 2007;13:7335–40.
30. Ballinger SW, Bouder TG, Davis GS, et al. Mitochondrial genome damage associated with cigarette smoking. Cancer Res. 1996;56:5692–7.
31. Jakupciak JP, Maragh S, Markowitz ME, et al. Performance of mitochondrial DNA mutations detecting early stage cancer. BMC Cancer. 2008;8:285.
32. Isaacs C, Cavalli LR, Cohen Y, et al. Detection of LOH and mitochondrial DNA alterations in ductal lavage and nipple aspirate fluids from high-risk patients. Breast Cancer Res Treat. 2004;84:99–105.
33. Zhu W, Qin W, Bradley P, et al. Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid. Carcinogenesis. 2005;26:145–52.
34. Jakupciak JP, Maggrah A, Maragh S, et al. Facile whole mitochondrial genome resequencing from nipple aspirate fluid using MitoChip v2.0. BMC Cancer. 2008;8:95.
35. Maitra A, Cohen Y, Gillespie SE, et al. The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. Genome Res. 2004;14:812–9.
36. Wong LJ, Lueth M, Li XN, et al. Detection of mitochondrial DNA mutations in the tumor and cerebrospinal fluid of medulloblastoma patients. Cancer Res. 2003;63:3866–71.
37. Kirches E, Krause G, Weis S, et al. Comparison between mitochondrial DNA sequences in low grade astrocytomas and corresponding blood samples. Mol Pathol. 2002;55:204–6.
38. Kirches E, Krause G, Warich-Kirches M, et al. High frequency of mitochondrial DNA mutations in glioblastoma multiforme identified by direct sequence comparison to blood samples. Int J Cancer. 2001;93:534–8.
39. Dakubo GD. Field cancerization: basic science and clinical applications. Hauppauge, NY: Nova Science; 2011.

Chapter 8
Circulating Tumor Cells and Cancer Stem Cells

Key Topics

- The discovery of circulating tumor cells (CTCs) by Thomas Ashworth
- The metastatic cascade and CTCs
- Site selection by CTCs
- CTC enrichment, isolation, and characterization
- Clinical applications of CTCs

Key Points

- Cancer metastasis involves the detachment and spread of cancer cells in circulation as CTCs, or circulating cancer stem cells (CCSCs). These cells are responsible for the establishment of metastatic deposits and thus a major cause of cancer mortality.
- Because the prognostic relevance of CTCs/CCSCs has been established, there has been a surge in the number of commercial enterprises developing different CTC enrichment and isolation technologies.
- The molecular characterization of CTCs is clinically very informative as well. While CTC analysis does not overcome tumor heterogeneity, the ability to track molecular cancer evolution, especially with acquisition of new drug-resistant mutations, informs therapy decision-making in real time.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_8

8.1 Introduction

In situ carcinomas and benign tumors are potentially curable by surgery and other primary prevention strategies, because these abnormal cells have not breached the basement membrane and have close phenotypic resemblance to their normal counterparts. The word "cancer" indicates the abnormal cell has broken through the basement membrane into the underlying tissue stroma. Thus, by definition, cancer cells are those that have the hallmarks of "detachment," "deformity," and "motility." The "detachment-deformity-motility" phenotype means the cells can actively change their cytoskeletal structures and maneuver in between endothelial cells and into the stroma and eventually the circulation. With the additional hallmarks of proliferation and induction of new blood vessel formation, the cells can also be passively pushed through the porous new blood vessel wall into the circulation. These cells in circulation are referred to as circulating tumor cells (CTCs), and their capture constitutes an aspect of the "liquid biopsy" concept.

Circulating tumor cells tend to home into distant preferential organs where they may remain dormant until they receive the right signals to begin regrowth. These dormant cancer cells, which are found in other organs, but mostly in the bone marrow, are referred to as disseminated tumor cells (DTCs). Though the mechanisms controlling clinical cancer dormancy are largely unknown, both microenvironmental cues and tumor evolution may account for the temporal dormancy, which can last for over a decade. Reversal into active proliferative state accounts in part for disease relapse and associated mortality.

To attain the metastatic phenotype, cancer cells have to undergo partial or complete epithelial-to-mesenchymal transition (EMT), a program or process used by embryonic cells for patterning and establishment of the various parts of the metazoan. Thus, metastatic cancer cells tend to be heterogeneous in regard to the extent of EMT phenotype attained. Conceivably, the process of EMT generates cells expressing epithelial, mesenchymal, and also stem cell markers. The population of cells expressing stem cell markers is referred to as circulating cancer stem cells (CCSCs). These CCSCs are resilient to chemotherapy and radiotherapy and may be the critical cells efficient at establishing metastatic growths. But the rarity of these cells makes it a formidable task to capture and study them.

The hematogenous spread of solid tumors constitutes the leukemic phase of the disease, and this is associated with diverse adverse clinical outcomes. The chances of surgical treatment failure, as well as resistance to other therapies, increase as a function of distance spread of cancer cells, especially with the formation of established metastatic tumors. Disseminated tumor cells, for example, are established prognostic biomarkers predictive of adverse clinical outcome and serve as surrogate markers for the presence of minimal residual disease following curative-intent surgery. Because DTC sampling is invasive, CTC capture and characterization has been explored for similar prognostic and other clinical applications. Several technologies have been developed by both industry and academia for CTC isolation and characterization. The CELLSEARCH® system is the industry

standard, being the first US FDA-approved CTC device for clinical use. However, other commercial devices and technologies are available for CTC capture and analysis.

Admittedly, molecular characterization of CTCs is unlikely to capture entire tumor alterations due to the heterogeneity of cancer cells. Other liquid biopsy approaches such as ctDNA and extracellular vesicular cargo represent better samples for such applications. However, CTC characterization has proven very useful in a number of applications, including study of tumor evolution, mechanisms of drug resistance and relapse, acquisition of metastatic potential, prognostication and detection of minimal residual disease.

8.2 Thomas Ashworth and Circulating Tumor Cells

The Australian physician, Thomas R. Ashworth, discovered CTCs in 1869, in his microscopic inquiry of a cancer patient's blood sample. On examination of blood from a patient who died from metastatic cancer, Ashworth observed cells that closely resembled those of the primary tumor and reasonably noted the implication of this in multiple cancers in the same patient. His observation that "*cells identical with those of the cancer itself being seen in the blood may tend to throw some light upon the mode of origin of multiple tumors existing in the same person*" is very telling today. Ashworth was able to conjecture based on this observation that these cancer cells must have detached and circulated over long distances to where he sampled the blood. Thus, he noted, "*one thing is certain, that if they came from an existing cancer structure, they must have passed through the greater part of the circulatory system to have arrived at the internal saphena vein of the sound leg*". As a deserving tribute, the Ingham Institute in 2014 organized the inaugural Thomas Ashworth CTC Symposium in his honor.

8.3 Metastasis and Formation of Circulating Tumor Cells

Circulating tumor cells are metastatic cancer cells that travel long distances in the vascular or lymphatic circulation to colonize distant tissues and organs. Their formation requires the completion of a complex series of activities referred to as the metastatic cascade (Fig. 8.1). These events are controlled by a balance between the activities of a "metastatic effector" and "metastatic suppressor" genetic events. Separate and overlapping molecular genetic cascades regulate tumor formation, progression, and metastasis. It is known, for example, that site-specific metastasis requires different gene expression patterns from primary prognostic expression profile.

The number of tumor cells and emboli in the circulation is a function of tumor size (larger tumors shed more cells than smaller tumors), and this may have

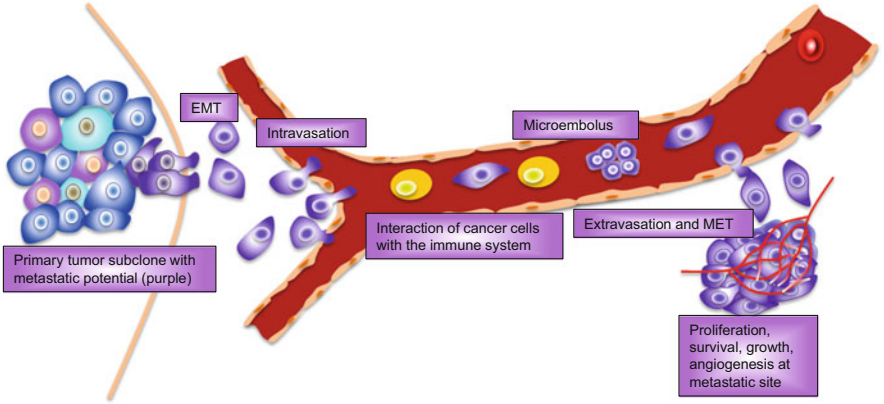

Fig. 8.1 The metastatic cascade. Loss of E-cadherin-dependent intercellular adhesions is an early event in metastasis of epithelial cancers. EMT follows this process, whereby cells acquire migratory phenotypes. Various molecules are then deployed for basement membrane disruption and eventual invasion of the surrounding stromal matrix. Intravasation then occurs, which could be augmented by tumor angiogenesis. Intravasation could also occur via lymphatics with eventual emptying into the bloodstream. Cells in circulation must overcome all the hazardous environmental cues including immunosurveillance and chemotherapeutic agents. Some cells migrate as microemboli, which may help prevent anoikis. Extravasation and homing occur at the target organ with eventual regrowth and formation of metastatic deposits. Extravasated tumor cells may also remain in a dormant state for possibly several years before regrowth

prognostic value that has been explored in CTC clinical applications. In spite of this, the inefficiency of metastatic deposit formation results in just a few cells being capable of forming metastatic tumors. The successful rate of metastatic deposit formation is estimated at 0.01 % of CTCs. Circulating tumor cells may extravasate after only 3 days in circulation. However, only ~1:40 extravasated cells will grow at the site and may be able to form micrometastasis, and even much fewer of these (1:100) can actually maintain growth to larger sizes to form tumors. Several reasons account for this inefficiency of metastatic deposit formation:

- Metastasis involves cellular detachment, evasion of anoikis, breach of the basement membrane, stromal interaction, vascular invasion, migration, extravasation, and reestablishment of tumor deposits. These processes require coordinated changes in expression of cell adhesion molecules, proteolytic enzymes, and cytoskeletal organization involving the RHO family members (RHO, RAC, CDC42). Evasion of anoikis, for instance, requires activation of survival pathways such as tropomyosin-related kinase B to inhibit caspase activation [1].
- Many CTCs are destroyed by one of three mechanisms: physical forces due to blood flow, elimination by effective antitumor immunity, and/or elimination by therapeutic agents. Expression of CD47 by colorectal cancer CTCs, for instance, enables escape from recognition by dendritic cells and macrophages [2].
- There is a tight spatiotemporal control of metastatic deposit formation.

- Microenvironmental conduciveness at the site of metastasis is important for successful tumor cell proliferation and regrowth.
- Many cells shed into the circulation may be executing apoptosis or other cell death processes and hence unable to establish tumor foci at distant sites.
- Molecular and gene expressional changes that control cell-cell communication, as well as growth factor receptor signaling, play important roles in successful colonization of the metastatic site.
- Metastasis requires reactivation of embryonic morphogenetic pathways needed to complete epithelial-to-mesenchymal transition (EMT) and vice versa (MET). EMT is dedifferentiation of epithelial cells into fibroblastoid cells that are migratory with altered mesenchymal gene expression. This process involves loss of E-cadherin functions and increased vimentin expression and is coordinated by several signaling pathways including the RAF/MEK/MAPK, PI3K/AKT, NF-kB, TGFβ, WNT, Notch, Hedgehog, and even hypoxia.

8.3.1 Basement Membrane Disruption

Normal epithelial cells usually require signals from the extracellular matrix (ECM) including the basement membrane to maintain their polarity. Cells can control the assembly and disassembly of the ECM. For cancer cells, the anchorage-independent growth frees them from normal epithelial-ECM control.

The beginning of locoregional or distant spread of the neoplastic cell is basement membrane derangement. This feature is a hallmark of cancer and differentiates benign tumors and premalignant lesions from invasive cancers. Once the basement membrane is breached, the number of cancer cells that have escaped cannot be estimated, or can it be ascertained that the tumor is organ confined, even if no gross or histopathologic evidence of local invasion is visible. Benign tumors can be large with highly disorganized epithelial structures, yet they consistently maintain intact basement membranes around them. Similarly, although in situ neoplasias have cellular atypia, they maintain intact basement membranes, and it is only when this membrane is disrupted that cancer can be diagnosed and warranting aggressive treatment. Indeed, what determines whether or not an in situ neoplastic cell breaches the basement membrane is unknown, hence the disagreement of establishing them as precursor lesions to the associated cancer. Compromised basement membrane is the definite initial process of tumor spread.

Migratory cancer cells require alterations in cell-cell and cell-ECM adhesions. This involves several cell surface receptors, but the type employed for such events depends on the type of tumor. Almost all invasive cancers express some receptors for the major adhesion molecules including integrins, cadherin, immunoglobulin superfamily members, and CD44. Decrease cell and matrix adhesion is required for escape, and contrary, at arrest and extravasation, increased cell and matrix adhesion is needed.

8.3.2 Invasiveness

Invasive tumor cells need to acquire special properties and complete an ordered cascade of events to be successful in establishing distant metastatic deposits. These properties and events include:

- Detachment from other cells without executing anoikis.
- Acquiring the ability to move and migrate (EMT).
- Breaching of the basement membrane.
- Acquiring the ability to intravasate (move between endothelial cells) into existing blood vessels or new blood vessels. This process involves cytoskeletal dynamics.
- Escaping from cell death signals, immune surveillance, and therapeutic agents.
- Ability to circulate as single cells or as cell clusters (microemboli or collective migration).
- Recognition and responding to distant site signals.
- Acquiring the ability to extravasate at a distant site.
- Encountering a microenvironment conducive for implantation, proliferation, and regrowth.

These actions require epigenetic and genetic changes that enable transition from epithelial-to-mesenchymal phenotype at initiation of invasion and the reverse from mesenchymal-to-epithelial phenotype at distant destination.

8.3.3 Molecular Genetic Control of Metastasis

Metastatic competence and behavior of cancer cells involve a complex array of events, orchestrated by several molecular signaling pathways. Molecules such as RHO GTPases, HGF, MMPs, IGF2, VEGFs, and autotaxin, among many others, play different roles in cancer cell metastasis. However, some established markers include changes in cell adhesion molecules (CAMs – cadherin, integrin, selectin, IgSF), CD44, FAK, BMP7, MMPs, and epithelial protein lost in neoplasia (EPLIN).

The loss of cell adhesion is an initial requirement for the cell to begin its egress. In the normal epithelium, cell junctions, namely, desmosomes and tight junctions, cohesively hold cells together. The formation and maintenance of these junctions are mediated by calcium-dependent interactions controlled by biomolecules such as cadherins. The cadherin (or calcium-dependent adhesion molecules) gene family of adhesion molecules is single-pass membrane glycoproteins with cytosolic, membranous, and extracellular segments (Fig. 8.2). The over 100 family members are subcategorized into five subfamilies, namely, classic types I and II cadherins, desmosomal cadherins, protocadherins, and cadherin-related proteins. Homotypic epithelial cell adhesions are regulated by epithelial cell classic cadherins or E-cadherin. The cytoplasmic or cytosolic component of E-cadherin physically

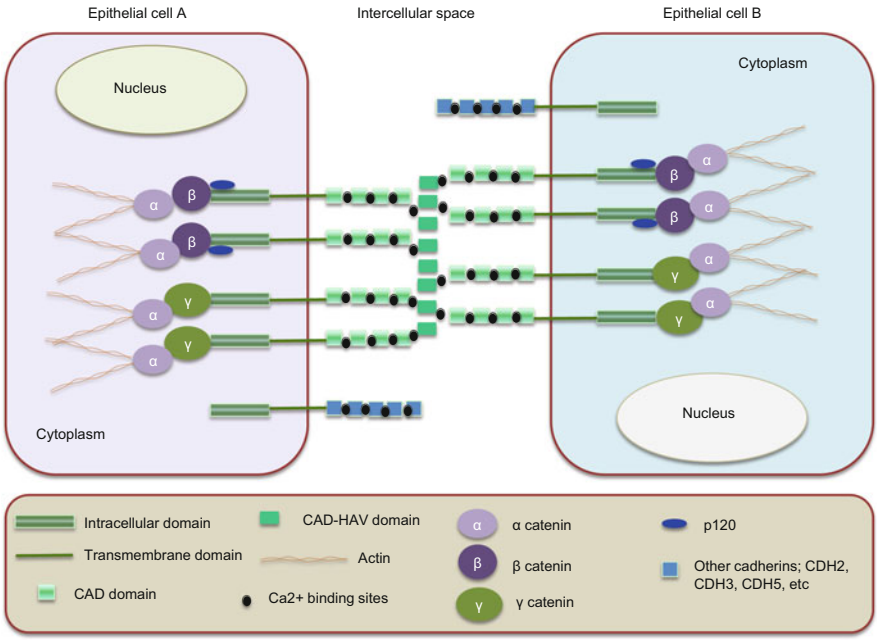

Fig. 8.2 E-cadherin adhesions. E-cadherins possess three domains, intracellular, transmembrane, and extracellular domains. Dimeric E-cadherins form hemophilic interactions with nearby cells through the extracellular domains. This interaction is stabilized and strengthened by calcium. The intracellular domain interacts directly with p210, β-catenin, and γ-catenin, which are anchored to the cellular actin cytoskeleton by α-catenin

interacts with the actin cytoskeleton through catenins. The extracellular segment of cadherin has five domains that mediate the epithelial cell-to-cell homotypic adhesions.

The catenin family of proteins forms complexes with cadherins. This family includes α, β, γ, and δ catenins. In addition to its cell adhesion functions, β-catenin is also a member of the canonical WNT signaling pathway. Under physiologic conditions, β-catenin is sequestered in an adhesion complex formed by α-catenin, p120 CAS, and the intracellular domain of E-cadherin (Fig. 8.2). This macromolecular complex is what tatters E-cadherin to the cytosolic actin cytoskeleton, thus maintaining normal cell-cell adhesion. The loss of E-cadherin-catenin complex, and hence loss of epithelial cell adhesion, results in the release of β-catenin from the complex. This loss of cell adhesion obviously frees in situ cancer cells to begin their migration. The released β-catenin also augments the neoplastic phenotype and may enhance migratory activity of the cancer cell. The released β-catenin accumulates in the cytosol and is usually kept in check by ubiquitin-mediated proteasomal degradation. The free β-catenin binds to APC and is phosphorylated by GSK3β for degradation. The absence or loss of APC functions, as occurs through mutations (e.g., in CRC), or loss of GSK3β functions possibly via WNT proto-oncogene

signaling, renders intracellular buildup of β-catenin. Intranuclear entrance of cytosolic β-catenin activates TCF/LEF1 family of transcription factors that induce expression of various oncogenes including *cMYC* and *CCND1*.

Changes in expression or function of E-cadherin or β-catenin through mutations, loss of heterozygosity, promoter hypermethylation, or proteolytic modifications can lead to loss of function and cell adhesion. Proteolytic modification of E-cadherin is demonstrated in mammary epithelial cells. Stromelysin-1, a member of the MMP gene family, can degrade the extracellular domain of E-cadherin resulting in loss of epithelial cell adhesion. Indeed, constitutive expression of stromelysin-1 in mammary epithelial cells causes EMT phenotype characterized by downregulation of cytokeratin expression, increased vimentin and MMP9 expression, and loss of catenin from adhesion complexes.

In addition to E-cadherin, the other family members also play roles in EMT and cancer invasiveness. The cadherin "switch" phenomenon underlies EMT. The downregulation of E-cadherin in cancer cells is often associated with increased CDH2/N-cadherin or neuronal cadherin. CDH2 binds to and activates the FGFR signaling pathway and also enhances β-catenin nuclear induction of EMT genes. Moreover, CDH1 expression, when coincident with expression of CDH3, causes CDH3-mediated IGF1R signaling and p120 phosphorylation to mediate cancer cell migration and invasiveness. Vascular endothelial cadherin or CDH5/VE-cadherin interacts with CDH2 to promote tumor growth, vasculogenesis, and intravasation. Truncated cadherin or CDH13/T-cadherin expression leads to EGFR signaling pathway activation. Thus, a multitude of cadherin actions, which may be due to tumor-specific induction (e.g., CDH17/liver-intestine or LI-cadherin and gastrointestinal cancer metastasis), underlie cancer metastasis and formation of CTCs.

Another molecule with established roles in cancer metastasis is CD44. CD44 is a transmembrane glycoprotein receptor for a ubiquitous component of the ECM, hyaluronan or hyaluronic acid (HA). The *CD44*, located on chromosome 11p13, has 20 exons that undergo alternative splicing to produce several isoforms referred to as variants (v). CD44 interacts with many extracellular components including collagens, osteopontin, and MMPs, but it is a major receptor for hyaluronic acid. Hyaluronic acid is upregulated during angiogenesis and is intrinsically an endothelial cell mitogen. Similarly, CD44 is overexpressed by many neoplastic cells and promotes cellular growth and invasion. The interaction of CD44 and HA facilitates binding and adhesion of cancer cells to the ECM, which promotes cancer invasion and metastasis. CD44 is also a cancer stem cell (CSC) marker and is involved in CSC invasive abilities. In colorectal cancer, CD44v6 expression in CSCs is involved in mediating the metastasis of these cells. $CD44^+$ $CD24^{-/low}$ Lin^- breast cancer stem cells can intravasate into the circulation and metastasize. Additionally, $CD44^+$ cells can efficiently establish tumors, indicating their important role in metastasis and establishment of new metastatic deposits.

A role for the immunoglobulin superfamily (IgSF) in cancer metastasis is well known. The IgSF is a very diverse group of proteins that include PECAM-1, ICAM-1, MCAM, NCAM, L1CAM, ALCAM, TCR, viral receptors, and MHC molecules. Molecularly, they are related by the shared immunoglobulin homology unit, which

is comprised of 70–110 amino acids that are organized into seven to nine β-pleated sheets. Structurally, many members of the group have N-terminal extracellular, a single transmembrane, and C-terminal cytoplasmic domains. The extracellular domain interacts with other members of the group (homophilic interactions) or with integrins or other carbohydrate moieties (heterophilic interactions), while the cytoplasmic unit interacts with the cytoskeleton and adaptor proteins that could mediate intracellular signaling. Functionally, the IgSF members are involved in cell-cell interactions. Cancer cells overexpress many members of the IgSF, which are implicated in cancer progression and metastasis. For instance, melanoma cell adhesion molecule (MCAM) mediates progression of melanoma and breast and prostate cancer cells. Similarly, the progression of several solid tumors is associated with altered expression of IgSF members.

8.3.4 Epithelial-to-Mesenchymal Transition

Epithelial-to-mesenchymal transition (EMT) mediates invasive behavior of most solid tumors. The majority of solid tumors are of epithelial origin, i.e., they are carcinomas. Epithelial cells are polarized in nature with apical and basal ends. The basal side interacts with the basement membrane through hemidesmosomes. In the epithelium, the cells are connected together by tight junctions toward their epical regions. While the epithelial cells can move within the epithelium, they never breach the basement membrane and enter into the ECM.

Greenburg and Hay [3] first observed the transition of epithelial cells into mesenchymal migratory phenotype and hence named this process EMT. Cultured adult anterior lens and embryonic cells in 3D collagen gels revealed changing morphology from polar epithelial to elongated mesenchymal cells that detached and migrated through the 3D gel as single cells. This process requires loss of cell-cell adhesion and acquisition of motile migratory ability in association with tissue reorganization, all of which recapitulate embryonic tissue patterning processes.

There are three types of EMT. Type I EMT operates during patterning and development of multicellular organisms. In the formation of some epithelial tissues, type I EMT cells revert to their epithelial phenotypes (epithelial plasticity). Type II EMT is described for the migration of cells during wound healing. Here the cells at the leading edges extend lamellipodia and migrate to close the wound. Thus, the leading keratinocytes (which have undergone partial EMT) drag the adjacent cells with them. The EMT-like process that occurs in cancer is referred to as type III EMT.

A functional hallmark of EMT is the change from stationary to mobile and migratory cell that can invade the ECM. However, other defining characteristics include altered morphology and differentiation marker expression. Morphological changes from apicobasal polarity to dispersed spindle-shaped cells with changes in cell-cell adhesion markers and cytokeratin intermediate filament to vimentin filament and fibronectin underlie EMT.

The defined molecular genetic control of EMT includes activation of several developmental signaling pathways and hypoxia that induce specific transcription factors to orchestrate the transition. An early event is loss of *E-cadherin*/*CDH1* expression. The expression of this gene is lost early during carcinogenesis, and germline mutations in this gene underlie familial gastric cancers. In addition to mutations and epigenetic gene silencing, signaling pathways also modulate E-cadherin functions. Tyrosine kinase receptor signaling, for example, negatively controls E-cadherin cell-cell adhesion functions. This pathway can phosphorylate β-catenin and the cytoplasmic domain of E-cadherin. Phosphorylated E-cadherin binds to Hakai E3 ligase that is internalized in endosomes and recycled to the cell surface or destroyed in lysosomes. Phosphorylated β-catenin is destroyed in proteasomes. These events initiate invasive behavior as cells lose adhesion and begin to migrate and spread.

Established transcription factors that mediate EMT include FOXC2, TWIST1, SNAI1, SLUG, GOOSECOID, SIP1 (ZEB2), and δEF1 (ZEB1). These markers are expressed by invasive tumors and facilitate invasive behaviors. In addition to other functions, SNAI1, SLUG, and SIP1 can repress E-cadherin expression. Upstream embryonic signaling pathways and hypoxia control the expression of these transcription factors as well. These pathways include TGFβ, WNT, and NOTCH signaling. For example, GSK3β phosphorylates SNAI1 for degradation. Hence, WNT signaling that inactivates GSK3β stabilizes SNAI1. Similarly, SNAI1 is activated by hypoxia, and TGFβ can induce expression of SIP1, SLUG, and SNAI1.

In the majority of invasive cancers, EMT is not a stable binary condition whereby cells are strictly epithelial or mesenchymal with reference to morphology and/or gene expressional changes (Table 8.1). It may be very transient and unnoticeable, and a complete mesenchymal phenotype may not be necessary for invasion. Additionally, some cancer cells may invade without undergoing EMT. Typical EMT is demonstrated in just a few cancers, especially diffuse-type gastric and invasive lobular breast cancers. The majority of these cancers lose E-cadherin expression and overexpress the EMT transcription factor TWIST1. Additionally, nuclear β-catenin accumulation in colorectal cancer cells with APC or β-catenin mutations is associated with increased expression of fibronectin and loss of E-cadherin expression. Novel approaches are required for the isolation of CTCs from these tumors, as the epithelial marker technology will miss them. Notwithstanding, the majority of CTCs, even invasive breast cancer, maintain epithelial morphology and express epithelial cell markers (EpCAM, CK) that have been successfully used for their isolation.

Distant metastatic deposits are epithelial in nature, often resembling the primary tumor of origin. Thus, EMT must be a reversible process. Even if tumors initially invade as mesenchymal cells, they must reverse at some point back to epithelial phenotypes before implantation and growth at distant sites [4].

Table 8.1 Epithelial, mesenchymal, and cancer stem cell markers

Cells	Markers
Epithelial	Cytokeratins 8, 18, and 19, E-cadherin, ESPR1, MUC1, HER2, EGFR
Mesenchymal	AKT2, SLUG, SNAI1, SERPINE1/PAI1, TWIST1, TG2, FOXC2, ZEB1, ZEB2, N-cadherin (cadherin 2), vimentin
Cancer stem cells	ALDH1, CD13, CD15, CD20, CD24, CD26, CD44, CD90, CD117, CD133, CD166, gangliosides 2 and 3 (GD2, GD3), Bmi1, integrin, nestin, LGR5

8.3.5 Circulating Tumor Cell Formation Independent of EMT

Cancer cells can also enter the circulation through other mechanisms including passive spread, as summarized:

- The work by Aceto et al. [5] suggests that CTCs can be released as 2–50 cell clusters held together by plakoglobin-dependent intercellular adhesions.
- The microtubule-organizing center (the centrosome) is amplified in some tumors, and those cancer cells are thought to enter the circulation in a non-EMT mechanism.
- Reduced intercellular adhesion is achieved via RAC1 signaling that induces Arp2/3-dependent actin cytoskeletal polymerization [6].
- Additionally, tumor angiogenesis generates leaky blood vessels. With tumor growth and expansion, cells or cluster of cells can be passively pushed or dislodged into these vessels and be carried away as CTCs.

These EMT-independent mechanisms should generate CTCs very suitable for epithelial marker-based assays. Similarly, migrating cell clusters should overcome anoikis.

8.4 Distant Site Selection: The Anatomic vs. Seed and Soil Hypotheses

Anatomic organization of tissues, their vasculature, and lymphatic supply provides facile explanation for target organ invasion by tumors. In this scenario, the organ initially encountered by the primary tumor cells in circulation serves as the target of choice. Thus, lymphatic and vascular drainage from the primary tumor site to regional organs or tissues accounts for such spread. A classical example is the drainage of the colon and colorectal cancer (CRC) hepatic spread. The hepatic capillaries are the first encountered by CRC cells in the blood (hepatic portal circulation), and hence the liver remains the most frequent initial site of CRC metastases. Such locoregional spread occurs in many cancers. However, such simplistic explanation cannot account for all cancer metastasis, especially to long

Table 8.2 Examples of "seeds" and "soils" of some cancers

Primary cancer (seed)	Distant metastatic sites (soil)
Melanoma	Lymph node, lung, liver, brain, gastrointestinal tract, bones
Breast cancer	Bone marrow, lung, liver, lymph nodes, brain
Lung cancer	Liver, bone, bone marrow, adrenals, brain, lymph nodes, contralateral lung, subcutaneous tissues, pancreas
Pleural mesothelioma	Retroperitoneal lymph nodes, contralateral lung and pleura, brain, spine, thyroid, prostate
Head and neck cancer	Lung, bone, liver
Salivary gland cancer	Lung
Thyroid cancer	Lung, bone
Esophageal cancer	Lymph nodes, bone
Gastric cancer	Liver, lungs, brain
Pancreatic cancer	Liver, peritoneal cavity, lungs
Colorectal cancer	Liver, lung
Hepatobiliary cancer	Liver, peritoneal cavity, lungs, pleura, brain, bone
Renal cell carcinoma	Bone, liver, lung, brain, distant lymph nodes
Bladder cancer	Retroperitoneal lymph nodes, lung, bone, liver
Prostate cancer	Bone, lung liver
Testicular cancer	Lung, liver, bone, other visceral sites
Ovarian cancer	Liver, lung, bone, supraclavicular and axillary lymph nodes
Endometrial cancer	Vagina, lung, intra-abdominal
Cervical	Mediastinum and supraclavicular lymph nodes, lung, peritoneal cavity, bone
Vaginal cancer	Aortic lymph nodes, lung, bone
Vulvar cancer	Iliac lymph nodes
Gestational trophoblastic tumors	Lung, kidney, GIT, spleen, liver, brain

distant organs. Tumors and site molecular characterization appear to offer some explanation for such target organ choices.

Stephen Paget conceived the "seed and soil" hypothesis in 1889. These early cancer researchers observed the predilection for some cancers to metastasize to the same organ. In the work of Paget, breast cancer spread to the liver, which is a frequent occurrence could not be accounted for by the easy accessibility of the liver by mammary blood supply, because other organs bathed by such circulating blood rarely developed metastases. Thus, Paget concluded that certain cancer cells (the seeds) could only successfully establish metastatic deposits in selective organs (the soil) that have suitable growth microenvironments (Table 8.2). Three important concepts characterize the "seed and soil" mechanism of cancer spread:

- Metastatic cells evolve as those that are fit to succeed in all phases of the entire metastatic process. Thus, a successful metastatic cell must complete the metastatic cascade.
- Primary tumors and their metastases consist of genetically diverse cells.

- Metastases generally develop in a site-specific manner. Because the microenvironments (the soil) of target organs are genetically and histologically different, specific cancer cells are only capable of colonizing one specific organ [7].

Two important complementary events are required for successful seed and soil interaction and establishment of metastases. The tumor cells must produce signaling factors responsive to the stroma and tumor microenvironment, but similarly the microenvironment must respond appropriately to these signals to permit habitation and growth of the new arrivals. Successful metastasis to an organ is precluded if this interaction fails, explaining why not all organs, even if the tumor cells encounter them, will be successful homes or soils for the seeds. Regulated activation of specific chemokines, cytokines, and proteases may guide metastasis to a specific site or organ. For example, lung metastasis may be predicted by the expression of inhibitor of differentiation-1 (ID-1), VCAM-1, MMP-1, and CXCL-12. Thus, breast cancer cells that express CXCR4 can selectively home into lungs because of the high expression of its only ligand CXCL12 [8].

Molecular characterization of tumors is providing insights into how "seeds" chose their "soils." Gene expression analyses of metastatic tumors have identified specific genetic signatures of some tumors for specific target organs. For instance, genes that mediate the metastasis of breast cancer to bone are different from those that mediate metastasis to the lung. Thus, different sets of genes create the right environment for tumor cells to specifically interact with the stromal cells of specific target organs [9].

8.5 Enrichment and Isolation of Circulating Tumor Cells

Circulating tumor cells have been enriched, isolated, and characterized using different methods and techniques (Fig. 8.3). About 80 % of all solid tumors are of epithelial origin. Hence, many CTCs are from epithelial cancers. In view of this, and the differences between epithelial cells and blood cells in regard to size and other physical properties, many earlier CTC separation techniques have involved exploitation of such physical differences. These physical property-based separation approaches include gradient centrifugation, filtration, and use of microelectrochemical systems. Because of the differences in molecular marker expression between epithelial cells and blood formed elements, newer methods have utilized tumor cell membrane unique epithelial surface markers such as EpCAM, for CTC isolation. These methods include immunomagnetic bead-based approaches and variations of this including the use of microfluidic platforms and nanotechnology approaches. There are currently >40 methods and techniques developed for CTC enrichment and isolation, and many more are being developed and optimized.

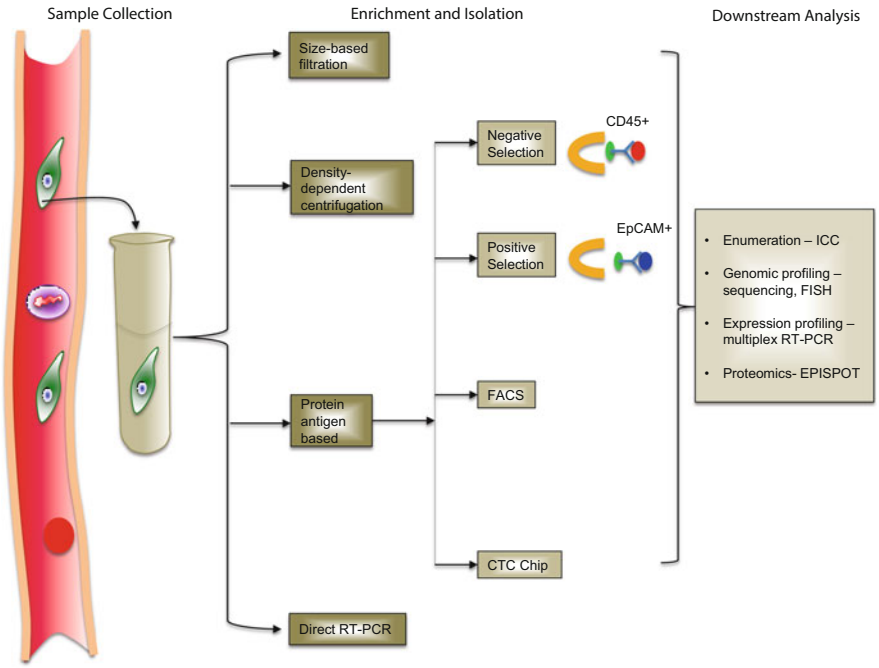

Fig. 8.3 Circulating tumor cell (CTC) enrichment, isolation, and characterization. CTCs can be analyzed using the various indicated methods of enrichment and characterization

8.5.1 The Filtration Principle

This procedure enables direct enrichment of CTCs based on size. Many of the normal cellular elements (RBCs, WBCs, and platelets) of blood are 8–10 μm in diameter, with epithelial cells being much larger with mean diameter of ~10 μm. In general, this separation method incorporates a polycarbonate membrane filter with pores of ~8 μm to separate most peripheral blood cells from epithelial cell components, which are mostly CTCs in advanced cancer patients. The method has some advantages over immunologic-based cytometry such as CELLSEARCH® and other CTC-chips. Because heterogeneity and EMT can result in loss of some tumor marker expression, this non-immunologic approach captures additional CTCs not amenable to antigen-antibody interaction technologies. Putatively, it thus captures more CTCs. Also, size-based methods and the herringbone (HB) chip are among the reliable methods for capturing circulating tumor microemboli (CTM). Described below are commercially developed CTC isolation techniques and devices based of the filtration principle.

8.5.1.1 ISET®

Prof. Patrizia Paterlini-Brechot, the founder of Rarecells, developed the isolation by size of epithelial tumor (ISET) technology for CTC isolation. The technology is based on CTC isolation base on the sizes of epithelial tumor cells. The procedure involves first the dilution of blood in RBC lysis buffer before being applied to the filtration membrane on the ISET device. The membrane has about ten wells with pore sizes of 8 μm. Following filtration, the membranes are washed and removed for staining and other analyses. The sensitivity threshold of this system is one CTC per 10 ml of blood treated within 4 h of collection.

8.5.1.2 ScreenCell®

Originally described by Vona et al. [10] and developed by ScreenCell, the ScreenCell® device enables CTC isolation by size using microporous membrane filters. Different filters are used to separate fixed cells for cytometry (ScreenCell® Cyto) and live cells for molecular biologic applications (ScreenCell® MB) or cell culture (ScreenCell® CC). The 18 μm-thick filter has a smooth hydrophilic surface with pore sizes of 7.5 ± 0.36 μm for isolating fixed cells and 6.5 ± 0.33 μm for capturing live cells. Similar to ISET, red blood cell lysis precedes filtration and the membranes are washed in PBS prior to use in downstream applications.

8.5.1.3 CellSieve™

Creatv MicroTech has developed an identical isolation method called CellSieve™. This uses biocompatible polymer filters of 13 mm diameters containing 7 μm pores for CTC capture. Pore density can be as high as 160,000 per 9 mm diameter.

8.5.1.4 Parsortix

The Parsortix system, developed by Angle plc (UK), uses a simple patented microfluidic technology to isolate CTCs based on size and compressibility (CTCs are less deformable than blood cells). A disposable cassette, which is of the size of a microscope slide, is placed in a clamp, and as blood flows through it, the Parsortix system automatically separates CTCs, which can then be stained and enumerated or harvested for other applications.

8.5.1.5 JETTA

DeNovo Sciences developed the JETTA technology for CTC isolation. It uses microfluidic slides with 56,320 capture chambers to isolate CTCs based on size. There are customized reagents and cartridges for breast, prostate, and colorectal cancers. This device has >85 % cell capture efficiency. Downstream applications of captured cells include immunohistochemistry, FISH, mRNA analysis, and PCR. Slides from JETTA can be transferred to the Vanguard Imaging System that prepares them for evaluation. The JETTA100 is a manual system; however, the new JETTA400 is a fully automated system that handles samples from input to result delivery.

8.5.1.6 ClearCell®

Clearbridge BioMedics has developed the ClearCell® system for CTC isolation. This label-free automated CTC enrichment system is based on a patented microfluidic biochip. This microfluidic biochip (CTChip®FR1) isolates CTCs based on size and deformity using Dean Flow Fractionation (DFF). Cells are focused within the microfluidic channels using the DFF process such that larger cells flow along the inner wall and smaller cells away from it. The entire process is complete in under an hour, and isolated viable cells can be used for multiple downstream applications.

8.5.1.7 ApoStream®

Developed by ApoCell, the ApoStream® uses differences in the dielectric properties of CTCs and blood cells to isolate CTCs. Dielectric properties of a cell depend on their size, volume, membrane area, density, and conductivity, which differ between cells. By using dielectrophoresis (DEP) field assist, cells are attracted to or repelled from a charged electrode. DEP forces pull CTCs toward electrodes on the chamber floor, repelling blood cells away into the eluent. The attracted tumor cells are then collected at the floor of the chamber through a port. The other blood cells are carried away in the eluent.

8.5.2 Centrifugation Principle

Density gradient centrifugation isolates both CTCs and mononuclear cells from blood. It involves centrifugation in density gradient (e.g., Ficoll) to create layers of cells based on their sizes and densities in relationship to the density of the suspension medium. The layers generated after the spin are from top to bottom, plasma,

CTCs and mononuclear cells (densities <1.077 g/mL), density gradient media (e.g., Ficoll), and then PMN leukocytes and RBCs (density >1.077 g/mL). Available kits for such cell isolation include Lymphoprep™ (Axis-Shield) and OncoQuick (Greiner Bio-One GmbH).

8.5.2.1 OncoQuick®

OncoQuick® (Greiner Bio-One GmbH) is a research use only product for the enrichment of CTCs from 15 to 30 mls of uncoagulated whole blood based on density gradient centrifugation. The 50 mL OncoQuick® centrifugation tube has two compartments separated by a porous barrier. The lower compartment has the blue-colored separation medium, while the upper chamber is where whole blood is loaded for separation. The sample is span at 1600 g at 4 °C for 20 min, resulting in column stratification of cells based on buoyant densities. The denser RBCs and PMN cells migrate through the barrier and displace the separation medium into the upper compartment. The less dense CTCs and other mononuclear cells are trapped between the separation medium and plasma. Targeted CTCs are harvested, washed, and used for downstream applications.

8.5.2.2 Lymphoprep™

Lymphoprep™ (Axis-Shield) is a ready-to-use media composed of reagents with composition conducive for the isolation of mononuclear cells in blood (monocytes and lymphocytes). Reported in 1968 by Boyum, this method relies on density disparity between mononuclear cells and other circulating cells [11]. Mononuclear cells have a much lower buoyant density than RBCs and PMN cells. Most mononuclear cell densities are <1.077 g/mL. Hence, centrifugation of blood in iso-osmotic medium of density 1.077 g/mL enables sedimentation of RBCs and PMNs, while entrapping mononuclear cells at the interface between the media and plasma. Optimal isolation requires mixing blood with saline at a 1:1 dilution and spinning at 600 g for 20 min at 20 °C. While this kit is not specified for CTC isolation, the identical buoyant densities enable the isolation of both cell types.

8.5.3 Immunomagnetic Bead Principle

This is a common method for CTC enrichment. It can be applied to whole blood or to cellular components obtained by centrifugation or filtration principles. Isolated cells can be analyzed by various molecular methods (RT-PCR) or enumerated after staining. Also FISH analysis can be performed on cell isolates.

The principle of cell isolation is simple. This could include only positive selection or both positive and negative selections (Fig. 8.3). Positive selection of

CTCs involves the use of antibodies attached to magnetic beads. These antibodies target epithelial or tumor-associated cell surface antigens such as human epithelial antigens, cytokeratins, epithelial cell adhesion molecules, oncofetal antigens (e.g., CEA), PSA, HER2, and several other markers. The inclusion of antibodies that target leukocytes (CD45, CD61 expressing cells) enables negative selection that can then be followed by positive selection. The method enables isolation and detection of living cells, quantification, and/or direct visualization of CTCs. These epithelial marker immunomagnetic enrichment methods include the CELLSEARCH®, AdnaTest, and various microfluidic devices.

8.5.3.1 CELLSEARCH®

CELLSEARCH® (Veridex) is an automated CTC isolation and enumeration assay approved by the FDA for management of metastatic cancer. It has been useful for enumerating CTCs for the prediction of overall survival and progression-free survival of metastatic breast, colorectal, and prostate cancer patients, but it is also being applied to the study of other epithelial tumors as well. The CELLSEARCH® CTC kit is used for selection, identification, and enumeration of epithelial CTCs in blood (requires 5–15 mls of blood). It uses an electromagnetic bead approach whereby ferrofluids loaded with EpCAM antibodies capture CTCs that are then visualized or detected with a cocktail of antibodies against positive (CTCs) and negative (other cells) cells.

The assay includes the CellSave preservative tubes that contain stabilizers for CTCs for up to 96 h to enable shipment at room temperature from distant locations. Additionally, optimal performance is monitored by inclusion of a control kit that contains low and high concentrations of fixed breast cancer cell line. The entire process from whole blood intake, aspiration of anti-EpCAM ferrofluid and buffer, magnetic incubation, cell separation, staining cartridge analysis, and presentation of fluorescent images for classification of CTCs is highly automated.

8.5.3.2 AdnaTest

AdnaTest (AdnaGen) is an epithelial cell marker-based CTC isolation process. It uses a proprietary blood collection kit (AdnaCollect blood) to obtain samples from patients for CTC isolation based on epithelial (EpCAM) marker expression. This procedure involves two main processes. The first is the AdnaTest Cancer Select that positively enriches CTCs using cancer-specific Select Beads applied to 5 mls of blood. The isolated CTCs are then subjected to the AdnaTest Cancer Detect analysis, which is an RT-PCR analysis of the sample. Cancer gene-specific primer pairs are employed for CTC detection. For example, the AdnaTest Breast Cancer Detect kit uses primers that target *MUC-1, HER2*, and *GA733-2*, with *ACTB* as the housekeeper control gene.

8.5.3.3 Microfluidic Technologies

The application of microfluidic technology has enhanced CTC isolation from whole unprocessed blood. The prototype, developed at Harvard by Mehmet Toner's group for CTC isolation, enables more cells to be captured per milliliter of blood than the industry standard, CELLSEARCH® [12–14]. This novel approach was able to capture ~50 CTCs per milliliter of blood compared to CELLSEARCH® sensitivity of 1 CTC/mL. Additionally, this microfluidic system enhances the purity of CTCs for subsequent enumeration and downstream molecular analysis. The version 1 chip, dubbed the CTC-chip, comprised of a silicon chamber with 78,000 EpCAM-coated microposts used for epithelial cell capture. Controlled laminar flow of 2–4 mls of unprocessed whole blood through the chamber facilitates CTC capture while reducing shear stress to cells. Viable captured CTCs can be fixed, permeabilized, and stained with DAPI, pan-cytokeratin, and CD45 FITC antibodies for imaging and enumeration. Alternatively, molecular analysis can be performed on biomolecules obtained from lysed CTCs on the chip.

While this proof of principle had enhanced CTC isolation, it had its own shortcomings. First, laminar flow of blood in a uniaxial streamline precludes some cells from making contacts with the microposts; hence, they escape from being captured. Second, the complex geometric design of microposts precludes high-throughput manufacturing that is necessary for large-scale clinical applications. To overcome these limitations, an improved version, called the herringbone chip (HB-Chip named after its configuration that resembles bones of a herring), was developed [15]. The HB design includes microvortex mixing of blood that disrupts the streamline flow and thus increases the interaction of CTCs with EpCAM-coated walls. It uses calibrated microfluidic flow patterns to drive cells into contact with the surface ridges. This simple geometry (without microposts) enables scale-up manufacturing for clinical evaluation. Additional advantage of the HB-chip is that the chamber is made of transparent materials that facilitate image resolution including possible use of electron microscopy in CTC analysis. In a proof of principle study of metastatic prostate cancer patients, the HB-chip demonstrated a median CTC capture of 63 per milliliter, which is a slight improvement over the CTC-chip.

8.5.3.4 On-Q-ity C5 Microfluidic Technology

On-Q-ity has developed proprietary technologies combining biomarkers and CTC analyses for personalized medicine including cancer treatment decisions and monitoring. The On-Q-ity C5 microfluidic chip uses dual capture technology to increase the sensitivity of CTC isolation from blood. This novel approach uses a combination of two microfluidic technologies into one, namely, separation based on antibody affinity method and size filtration technology. The blood flows in a streamline fashion through gradient gap sizes between posts that decrease from 40 μm to

12 μm. Additionally, the posts are coated with EpCAM antibodies. Because not all CTCs express EpCAM, this dual method enables CTC capture based on EpCAM expression and also enables non-EpCAM-expressing cells to be captured based on size. This dual capture method is superior to either technology used alone. Indeed, it can isolate between 1 and 30 CTCs in just 3 mls of blood, which is an improved sensitivity over other methods. Captured CTCs are detected in a similar fashion to CELLSEARCH®, which is staining with DAPI, CK, and CD45 antibodies.

8.5.3.5 CTC-iChip

The CTC-iChip is a novel CTC isolation device. In this process, whole blood treated with anti-CD45 and anti-CD66b magnetic beads (to capture WBCs) is applied to the CTC-iChip microfluidic device. The chip houses two separate microfluidic devices, but with three distinct microfluidic components that serve different purposes. The first part eliminates RBCs and platelets by separating nucleated cells (WBCs and CTCs) by size-based deflection referred to as deterministic lateral displacement. The second microfluidic component uses inertial focusing to line up the nucleated cells for subsequent separation. Finally, magnetic separation and magnetophoresis enable the separation of unlabeled CTCs from bead-coated WBCs [16].

8.5.3.6 IsoFlux™ System

Fluxion biosciences developed and commercialized a fully automated microfluidic technology for isolating viable CTCs and other rare cancer and immune cells for downstream applications. This proprietary microfluidic technology platform uses immunomagnetic bead capture procedure and is flexible in its application because cocktails of antibodies can be used to target different cell types such as CTCs, cancer stem cells, cells undergoing EMT, or immune cells. In principle, 7–10 mls of whole blood is first devoid of RBS through Ficoll gradient centrifugation, and the remaining cells are coupled with IsoFlux beads carrying one or more antibodies (e.g., EpCAM, EGFR, HER2, N-cadherin, etc). The sample is then loaded onto the IsoFlux cartridge that contains microfluidic flow channels at the bottom. At the cell isolation zone of the flow channels, cells coated with beads are pulled up to the top surface of the cartridge by neodymium magnet, while uncoated WBCs run into a waste well. The isolated cells can be enumerated or detected using immunofluorescence, and fluorescence in situ hybridization, or used for molecular analysis. The rate of CTC detection (>5 CTCs) ranges from ~60 % for lung cancer to ~100 % for breast cancer.

8.5.3.7 GILUPI CellCollector™

The GILUPI CellCollector™ (GILUPI GmbH) method is the first in vivo CTC capture microfluidic device. It consists of an EpCAM-coated wire device that is placed through a cannula into a vein of the patient. It is then left in place for 30 min, during which time an estimated 1.5 L of blood passes over the device. This process increases the chance and number of captured CTCs. In clinical validation studies, the CellCollector™ was able to isolate significantly high numbers of CTCs even in early-stage cancer patients. This has potential for monitoring early micrometastasis and minimal residual disease. Its detection rate of CTCs in patients with various solid tumors is up to 70 %.

8.5.3.8 CytoTrack

CytoTrack CT11™ technology is a rare cell (such as CTCs) scanning and isolation technology on solid surface. This approach to CTC detection marries conventional CD/DVD and flow cytometry technologies. RBCs in the collected sample are lysed, and the sample then stained, evenly spread, and immobilized on a CytoTrack disk. The disk has a high capacity that enables even up to 30 mls of blood to be evenly spread on. FITC-stained CK and other markers enable fluorescent scanning of positive cells. The disk with the sample is then placed in the CytoTrack FM3 scanner. The disk rotates at a very high speed, while scanning at high resolution of up to 100 million cells per minute takes place. The CytoTrack data management system processes and displays the results. The exact position of each target CTC detected is captured on the disk. High-quality images of CTCs are produced and displayed in gallery format for review. Each target is revisited, reanalyzed, and verified. CTCs are often seen in clusters, but single-cell genotypic and phenotypic analysis is possible. After CTC enumeration, the CytoPicker™ can lift off cells from the glass surface into Eppendorf tubes for other uses.

8.5.4 Other CTC Separation Methods

Vitatex technologies developed the cell adhesion matrix (CAM)-based CTC isolation technique. The CAM-based iCTC platform exploits the invasive properties of cancer cells to isolate metastatic-initiating circulating tumor cells (iCTCs). When patient blood samples are applied to the CAM-coated tubes (Vita-Cap™) or culture plates (Vita-Assay™), iCTCs preferentially adhere to CAM. The captured iCTCs are labeled due to their ingestion of FITC- or TRITC-labeled CAM. Isolated iCTCs can be used for cytometric or molecular analysis.

8.5.5 Detection and Characterization of CTCs

Following CTC enrichment, isolation, and capture, the cells have to be identified and characterized. Various cytometric and molecular biologic techniques enable bona fide CTC detection. Other downstream characterizations include single-cell genomics and next-generation sequencing.

Circulating tumor cells have been identified and characterized using PCR assays targeting tumor-associated transcripts. For example, in gastrointestinal cancer CTC analysis of blood samples from patients with esophageal cancer, density gradient enrichment of CTCs followed by RT-PCR for several mRNAs, including *CEA*, *BIRC5*, and *ERCC1*, has proved to be important in cancer management. Patients with persistently elevated CTCs as demonstrated by *CEA* mRNA levels are more likely to have a recurrent tumor. Similarly, survivin mRNA levels significantly fall following complete surgical tumor resection. Additionally, a significant decrease in survivin-expressing CTCs is found in patients with adenocarcinoma on neoadjuvant chemoradiation. Patients who respond poorly to neoadjuvant radiotherapy have significantly elevated CTC levels as assayed by *ERCC1* mRNA. Carcinoembryonic antigen RT-PCR after density gradient centrifugation of samples from patients with gastric cancer revealed absence in healthy controls and patients with benign conditions, but positive in 36.8 % of cancer patients. Carcinoembryonic antigen positivity correlated with depth of tumor invasion, and elevated levels were associated with the likelihood of systemic spread [17]. While qPCR has increased sensitivity for CTC detection, it has limitations such as primer specificity, heterogeneous gene expression by CTCs, and possible tumor marker gene expression by normal blood cells.

Circulating tumor cell enumeration and hence their absolute numbers have prognostic value in patients with different types of solid tumors. Various technologies use fluorescence immunocytochemistry to label, enumerate, and characterize CTCs following enrichment and/or isolation. Additionally, genetic analysis is also possible through fluorescence in situ hybridization to look for genetic and chromosomal alterations. For example, the CELLSEARCH® system stains for cell nuclear using DAPI, cytokeratins (8, 18, and/or 19) with anti-CK-PE antibodies, and leukocytes with anti-CD45-APC following enrichment. Cells that have dual positive and negative staining ($CD45^+/CK\text{-}PE^+$ or $CD45^-/CK\text{-}PE^-$) are excluded from enumeration. Thus, only $EpCAM^+$, $CK\text{-}PE^+$, and $CD45^-$ cells are scored as CTCs and used for patient management. Additional fluorescence channel is available for user-defined marker identification. For instance, this channel can be used for markers such as vimentin, E-cadherin, androgen receptor, PSA, HER2, or EGFR.

An additional armamentarium to CTC enumeration and characterization is the development of objective automated image analysis systems. The Ariol cellular imaging system is a fully automated method used to scan slides to identify CTCs using three fluorescent channels (for cytokeratins, CD45, and nuclear staining). Additionally, it enables bright-field imaging to delineate cell morphology. An

advantage of the system is the ability to quantify DNA ploidy, enabling genetic and chromosomal analysis of CTCs.

Viable CTCs have been detected and characterized using the EPISPOT (epithelial immunospot) assay. This assay enables real-time viable CTC detection based on secretion of specific cancer cell proteins. The enriched cells are cultured on membranes coated with specific antibodies that capture the secreted proteins by the CTCs. These proteins are then detected with fluorochrome-labeled secondary antibodies. This assay detected CK19-negative and MUC1-secreting circulating breast cancer cells even in patients with early-stage localized disease (54 % stage M0) [18].

8.6 Circulating Cancer Stem Cells

In many solid tumors, epithelial, mesenchymal, and CSC markers have been demonstrated in CTCs enriched from the same patient. In some patients negative for CTCs by standard definition, markers of EMT have been detected, indicating that these patients will be deemed CTC negative, and possibly classified as having favorable disease outcome, when indeed they harbor exclusively CTCs with mesenchymal phenotypes. In some instances, a single CTC can co-express both epithelial and mesenchymal markers. High expression of mesenchymal markers is associated with treatment resistance and worse prognosis. Thus, the development of technologies to capture these EMT cells and circulating CSCs such as the AdnaTest EMT-1/stem cell (with TWIST, AKT2, PI3K, and ALDH1) is commendable.

The process of cancer metastasis produces cells that acquire various phenotypes conducive to their functions. This process generates cells that express stem cell markers. These cells with self-renewal abilities are identified as cancer stem cells (CSCs) or cancer stemlike cells. In addition to their stemness, these cells express mesenchymal markers and are mobile and can intravasate into the circulation and become CCSCs. Circulating CSCs have been detected with epithelial markers, suggestive of transition from mesenchymal to epithelial cells in preparation for invasion and colonization of distant sites.

However, the accurate characterization of CCSC requires the use of just a few reliable markers. Markers with strong association to CCSCs include aldehyde dehydrogenase 1 (ALDH1), a putative breast cancer prognostic biomarker, CD44 cell surface glycoprotein involved in cell invasion and metastasis, as well as gangliosides 2 and 3 (GD2 and GD3). Another ganglioside, GD1a, is however a putative marker of CCSCs. The xenobiotic extrusion pump protein, ADCG2, is highly expressed by CSCs and may be involved in resistance mechanisms to chemotherapy. Thus, its expression in CTCs may indicate some components of stemness.

Conceivably, only a minute percentage (0.01 %) of all CTCs possess the ability to form metastatic deposits, and these are the rare cells that harbor stem cell-like features. Indeed, in breast cancer patients, many DTCs in bone marrow are $CD44^{+}$/

$CD24^{-/low}$, and this finding is associated with aggressive tumor behavior and successful establishment of metastatic tumors [19]. Circulating CSCs may be responsible for therapy resistance and hence treatment failures in patients with advanced-stage disease. Targeting these cells should be of utmost importance in cancer biology and pharmaceutical developments. To this end, numerous agents are in development and at various stages of clinical trials to target CCSCs and the mechanisms by which they are generated. Inhibitors of the Hedgehog, WNT, Notch, PI3K/AKT/mTOR, and other signaling pathways are developed and have shown promise at destroying CSCs. The Hedgehog pathway inhibitor, vismodegib, for instance, inhibits the growth of tamoxifen-resistant breast cancer cells. RKF118-310, an inhibitor of the WNT pathway, eradicates breast cancer stem cells, and γ-secretase inhibitors (GSIs) that block the Notch pathway are able to kill CSCs. The GSI, MK-0752 (RO4929097) has shown promise in phase I and II clinical trials in patients with metastatic breast cancer. The PI3K inhibitor, everolimus (RAD001), blocks CSC growth and has shown promise for treating breast cancer patients resistant to conventional therapy. This drug may be useful in trastuzumab-resistant patients and has undergone phase III clinical trials for this purpose (NCT00876395). The preliminary finding is that progression-free survival was insignificant between groups; however, a 7.2-month prolongation was observed with addition of everolimus in the hormone receptor-negative, HER2-positive patients [20].

8.7 Clinical Applications of CTCs

Traditionally, bone marrow aspiration for DTC detection has been used for prognostication of many solid tumors. Tumor micrometastasis preferentially homes into the bone marrow to establish dormancy from where they recirculate to preferred "soils" to establish metastatic deposits. The preferred site could be, or include, the primary site or organ of the tumor. Thus, the presence of DTCs indicates the possible future formation of both distant metastasis and local relapse.

Circulating tumor cell detection is not useful for early cancer detection, although it has an ancillary role in establishing the diagnosis of primary tumors. In clinically localized cancer (M0), CTC evaluation helps determine the presence of minimal residual disease. However, CTCs are established as prognostic biomarkers for several solid tumors. In some cases, CTC characterization has even proven superior to conventional imaging and other clinical metrics of treatment response prediction. In early-stage disease without clinical evidence of metastasis, CTC presence, albeit low, is of significant value in detecting minimal residual disease, and this correlates with prognostic factors such as disease-free survival, disease-specific survival, and overall survival. Expectedly, CTC detection rates are much higher in patients with metastatic disease than those with localized cancer. In patients with advanced-stage disease, baseline and posttreatment CTC numbers or presence predict worse outcome.

Molecular characterization of CTCs is clinically very informative in identifying therapeutic targets as well as uncovering the mechanisms of tumor evolution relevant to treatment resistance. The transition of the cancer cell from its initiation to metastatic phenotype involves acquiring additional specific hallmarks of cancer. The biologic mechanisms evoked to achieve the propensities to invade surrounding tissues and metastasize require overcoming selective pressures. To circumvent these pressures, some clones evolve demonstrating altered genomic expression and mutational status. In the advent of personalized oncology, targeted therapies are available for treating cancers with specific genetic signatures. Traditionally though, primary tumor tissue is interrogated for these changes to inform clinical decision-making. However, such therapies are targeted mostly at minimal residual disease and metastatic cells in circulation. It is well established that CTCs and CCSCs can have different genetic makeups from that of the primary tumor. For example, breast cancer patients with $HER2^+$ and/or ER^+ primary tumors are found to have $HER2^-$ and/or ER^- CTCs and vice versa. Similarly, *BRAF* and *KRAS* mutation status may differ between primary tumors and CTCs from the same tumor. The presence of *KRAS* mutant CTCs may explain treatment failures with anti-EGFR agents in CRC patients whose primary tumors have wild-type *KRAS* genotype. Thus, longitudinal sampling for CTCs informs appropriate clinical decision-making and also enables the study of various therapy-resistant mechanisms.

Finally, CTC detection and characterization can be useful in monitoring or determining the risk of disease recurrence after surgery or radiotherapy, as well as the origin of the tumor, if the primary is in doubt. Circulating tumor cells can also serve as surrogate endpoint biomarkers in clinical trials and in the development of companion diagnostics.

8.8 Summary

- Cancer invasion leads to the release of cancer cells and stem cells into the circulation as CTCs and CCSCs. These cells may circulate as single cells or as tumor microemboli.
- CTC formation involves specific gene expressional changes leading to phenotypic changes in the cell. An initial change from epithelial-to-mesenchymal phenotype helps their mobility and exits into the circulation. At the metastatic site, they reverse this phenotypic change and become epithelial to establish a metastatic deposit.
- CTCs may remain dormant in the bone marrow or in other tissues for years, and when given the right cues, they may reenter the circulation and subsequently form metastatic disease.
- Tumor metastasis and treatment failure are major contributing factors to cancer mortality. A reason for the failure in cancer treatment is the possible lack of response by a rare cell type, the CCSC, to treatment interventions.

- Another problem with cancer therapy is that targeted therapy decisions are often made based on the genotype of the primary tumor. Not only will tumor heterogeneity preclude a complete capture of the genetic alterations, but also metastasis is associated with tumor genetic evolution. Thus, effective cancer treatment should target CTCs and CCSCs in real time.
- CTCs can be enriched, isolated, and characterized using various technologies including centrifugation, filtration, immunologic capture, and molecular approaches.
- CTCs have prognostic value in cancer management. Not surprising, therefore, following the FDA clearance of CELLSEARCH® for clinical applications, a plethora of companies have emerged with novel technologies for CTC enrichment, isolation, and characterization.

References

1. Douma S, Van Laar T, Zevenhoven J, et al. Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. Nature. 2004;430:1034–9.
2. Steinert G, Scholch S, Niemietz T, et al. Immune escape and survival mechanisms in circulating tumor cells of colorectal cancer. Cancer Res. 2014;74:1694–704.
3. Greenburg G, Hay ED. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. J Cell Biol. 1982;95:333–9.
4. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer. 2002;2:442–54.
5. Aceto N, Bardia A, Miyamoto DT, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158:1110–22.
6. Godinho SA, Picone R, Burute M, et al. Oncogene-like induction of cellular invasion from centrosome amplification. Nature. 2014;510:167–71.
7. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer. 2003;3:453–8.
8. Bacac M, Stamenkovic I. Metastatic cancer cell. Annu Rev Pathol. 2008;3:221–47.
9. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;436:518–24.
10. Vona G, Sabile A, Louha M, et al. Isolation by size of epithelial tumor cells: a new method for the immunomorphological and molecular characterization of circulating tumor cells. Am J Pathol. 2000;156:57–63.
11. Boyum A. A one-stage procedure for isolation of granulocytes and lymphocytes from human blood. General sedimentation properties of white blood cells in a 1g gravity field. Scand J Clin Lab Invest Suppl. 1968;97:51–76.
12. Nagrath S, Sequist LV, Maheswaran S, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. 2007;450:1235–9.
13. Sequist LV, Nagrath S, Toner M, et al. The CTC-chip: an exciting new tool to detect circulating tumor cells in lung cancer patients. J Thorac Oncol. 2009;4:281–3.
14. Maheswaran S, Haber DA. Circulating tumor cells: a window into cancer biology and metastasis. Curr Opin Genet Dev. 2010;20:96–9.
15. Stott SL, Hsu CH, Tsukrov DI, et al. Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. Proc Natl Acad Sci USA. 2010;107:18392–7.
16. Karabacak NM, Spuhler PS, Fachin F, et al. Microfluidic, marker-free isolation of circulating tumor cells from blood samples. Nat Protoc. 2014;9:694–710.

17. Miyazono F, Natsugoe S, Takao S, et al. Surgical maneuvers enhance molecular detection of circulating tumor cells during gastric cancer surgery. Ann Surg. 2001;233:189–94.
18. Alix-Panabieres C, Schwarzenbach H, Pantel K. Circulating tumor cells and circulating tumor DNA. Annu Rev Med. 2012;63:199–215.
19. Balic M, Lin H, Young L, et al. Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. Clin Cancer Res. 2006;12:5615–21.
20. Hurvitz SA, Andre F, Jiang Z, et al. Combination of everolimus with trastuzumab plus paclitaxel as first-line treatment for patients with HER2-positive advanced breast cancer (BOLERO-1): a phase 3, randomised, double-blind, multicentre trial. Lancet Oncol. 2015;16:816–29.

Chapter 9
Extracellular Vesicles as Cancer Biomarkers in Circulation

Key Topics

- Types of extracellular vesicles (EVs)
- Exosome biogenesis and content
- Biogenesis of microvesicles
- Biologic functions of cancer-derived EVs
  - Immunosuppression
  - Horizontal transfer of tumor-promoting factors
  - Modulation of tumor microenvironment and metastatic niche
  - Promotion of cancer cell migration and invasion
  - Promotion of cancer biology
- Apoptotic bodies and oncosomes
- Quality issues in EV analysis
- EV isolation, characterization, and analysis

Key Points

- Healthy and cancer cells release membrane-enclosed structures into the extracellular milieu, from where they enter the circulation. These extracellular vesicles are variously called exosomes, microvesicles, microparticles, apoptotic bodies, oncosomes, or prostasomes based on their sizes, mechanism of release, marker composition, and the type of cell that releases them.
- Extracellular vesicles carry biologically active biomolecules used for intercellular communications. In cancer, their diverse functions include transfer of oncogenic molecules (e.g., drug-resistant factors and

(continued)

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_9

oncogenes), immunosuppressive effects, modulation of tumor microenvironment, and premetastatic niche formation.
- The clinical translational potential of extracellular vesicle biomarkers is being realized with the development of commercial products.

9.1 Introduction

Intercellular communication is central to homeostatic control in the metazoan organism. These communications may occur in a nearby paracrine fashion or at a distance in an endocrine mode. While various humoral factors such as hormones are commonly used, there is growing evidence of involvement of other bioactive cellular components such as receptors and nucleic acids in such communication networks. Thus, one mode of conveying these biomolecules to other cells is through membrane-enclosed structures.

It has therefore been established that both normal and abnormal or diseased cells constantly shed membrane-enclosed vesicles of various sizes and compositions into the extracellular milieu, circulation, and other body fluids. While there may be some overlaps, attempts are being made to subclassify these extracellular vesicles (EVs) based on their sizes, mode of production, unique membrane features, specific markers, and mode of isolation, among others. Well-characterized among these vesicles are exosomes that are small nano-size vesicles formed by inward budding of microvesicular bodies. A heterogeneous population of cellular-derived EVs exists with various physiologic and pathologic functions.

The process of EV formation results in having them packaged with bioactive molecules including DNA, RNA, proteins, and lipids that are used for intercellular communication in both paracrine and endocrine fashions. Their interactions with other cells usually lead to alterations in recipient cell functions. In pathologic conditions, such as in cancer patients, their levels are elevated in circulation and other body fluids and can serve as diagnostic or prognostic biomarkers. Moreover, the biomolecular compositions of EVs, especially miRNAs, proteins, and mutated cancer genes, are powerful biomarkers with great clinical utility. Commercial products based on extravesicular cargo are being advanced for clinical applications. While focused on the biology of EVs, this chapter also highlights the quality standards recommended by the international society for extracellular vesicles (ISEV) for profiling and analyzing EVs, as well as commercially available EV diagnostic assays.

9.2 Cancer-Derived Extracellular Vesicles

Friend et al. first reported the presence of cancer-derived extracellular (cdEVs) or microvesicles in 1978 [1]. They identified them in culture media of lymphoid nodules obtained from patients with Hodgkin's lymphoma. Their functional importance in cancer biology was revealed 2 years later by the work of Poste and Nicolson [2]. In their work on mouse melanoma cells, it was uncovered that metastatic phenotype could be horizontally transferred through extracellular vesicles. Highly metastatic F10-derived EVs could confer metastatic ability on a poorly metastatic melanoma cell, enabling their spread to the lung. These initial discoveries opened the door for further investigations into the importance of cdEVs in cancer biology. Our current knowledge has expanded the role of cdEVs considerably to include modulation of tumor microenvironment and premetastatic niche, immunosuppressive effects, and transfer and sharing of oncoproteins and resistant factors. They serve as vehicles and targets of cancer therapy and as biomarkers for cancer detection, prognosis, and monitoring.

9.3 Types of Extracellular Vesicles

Normal, cancer, and other abnormal cells release membrane-bound vesicles into the interstitial fluid, circulation, and other body fluids. These vesicles are given various names based on differences in their sizes, mode of release, contents, cellular origin, and some specific markers (Table 9.1). *Exosomes*, which are the best characterized of these vesicles, are fairly homogeneous, measuring between 30 and 150 nm. They are produced by inward membrane budding, and their membrane composition resembles those of endosomes more than the plasma membrane. The next well-studied extracellular vesicles measure between 100 and 1000 nm in size and are produced in an opposite fashion to exosomes, i.e., by outward budding. These larger vesicles (but could be of smaller sizes as well) are variously called *ectosomes*, *microparticles*, *microvesicles*, and *shedding vesicles*. In reference to cancer, these vesicles will be referred to as *cancer-derived extracellular vesicles* (cdEVs) to simplify nomenclature. Other extracellular vesicles released by dying cells are called *apoptotic bodies*. These are much larger structures (1000–5000 nm in size) that are produced by physical process of cellular contraction and hence random plasma membrane bulging and pinching off during cellular apoptosis. The membrane-associated molecules and intramembranous contents of EVs perform critical biologic functions.

Additional extracellular vesicles with different nomenclatures are *argosomes*, *prostasomes*, (also referred to as *epididymosomes* and *liposomes*), and *oncosomes*. Argosomes are identical to exosomes but have no known role in cancer biology. The basolateral membranes of the fly wing disc cells were first noted to release these exosome-like vesicles, which are now established to contain morphogens

Table 9.1 Features of some extracellular vesicles

Features	Exosomes	Microvesicles	Apoptotic bodies
Size	50–100 nm	100–1000 nm	400–1000 nm
Shape	Cup shaped	Irregular	Heterogeneous
Sedimentation	100,000× *g*	Variable depending on size (1200–100,000× *g*)	Variable depending on size (1200–100,000× *g*)
Lipid composition	Lipid rafts, cholesterol, ceramide, sphingomyelin, phosphatidylserine	Phosphatidylserine	NA
Markers	ALIX, TGS101, ESCRT, tetraspanins-CD63/CD9	Integrins, CD40 ligand, selectin	Histones
Mode of release	Multivesicular endosomes	Exocytosis	Condensed apoptotic fragments

needed for proper patterning of the multicellular organism. Prostasomes are 50–500 nm vesicles that were first described in 1977 [3, 4]. They are secretions from the prostate gland epithelial cells into seminal fluid. They carry cargos including tissue factor, decay accelerator factor, protectin, complement regulatory membrane cofactor protein, as well as high levels of calcium, zinc, and magnesium ions. These molecules play various physiologic roles in male reproduction, such as regulation of the timing of sperm cell capacitation, induction of acrosome reaction, stimulation of sperm motility, and sperm protection from destruction by immune cells in the female reproductive tract. However, cancer cell prostasomes are likely exosomes, because they are found in storage vesicles that resemble microvesicular bodies. Oncosomes are much larger (4000–10,000 nm) EVs generated from protrusions of cancer cells. They are commonly released as blebs from fast-migrating tumor cells. Their shedding can be induced by overexpression of MyrAkt1, caveolin-1, and HB-EGF and also by ERK-mediated silencing of diaphanous-related formin 3 (DIAPH3) [5]. They are enriched with tissue degradation enzymes such as metalloproteinase and scaffolding proteins such as caveolin-1. The cargo of oncosomes, which includes nucleic acids such as miRNA and proteins, is identical to other EVs, and hence their functions may be similar as well.

The problems with extracellular vesicular nomenclature include the absence of efficient means of differential extracellular vesicle isolation techniques. Methods currently used for purifications, mostly targeting exosomes, include laborious procedures such as differential centrifugation (increasing speeds to remove larger vesicles followed by ultracentrifugation (100,000× *g*) to isolate exosomes), ultracentrifugation, filtration with 220 nm filter to remove larger vesicles and debris followed by ultracentrifugation, sucrose gradient centrifugation, and the use of commercial technologies such as those developed by Exosome Diagnostics, Caris Life Sciences (immunomagnetic bead capture based on EpCAM expression), and System Biosciences (ExoQuick precipitation without centrifugation). All these methods are still faced with lack of purity in regard to source of EVs (exosomal or microvesicular), yield, and integrity and most are time consuming and costly.

9.4 Exosomes

Exosomes are membrane-bound nanovesicles (~50–100 nm in diameter) that almost all cells secrete into various body fluids including plasma, urine, saliva, nipple aspirate fluids, bronchial lavage fluid, pleural effusions, ascitic and other body fluids. Exosomes were originally discovered to be secreted by immature red blood cells (reticulocytes) and were functionally implicated in the removal of plasma membrane proteins and to contain heat shock protein 70. Structurally, exosomes are the smallest extracellular vesicles, are more homogeneous than the other extracellular vesicles, and have a density of 1.13–1.19 g/mL in sucrose. They are usually isolated by sucrose gradient centrifugation at 100,000$\times g$.

9.4.1 Exosome Biogenesis

Exosomes are formed and released from multivesicular bodies (MVBs) following the degradative endosomal pathway of endocytosed extracellular molecules (Fig. 9.1). Following endocytosis of extracellular content, an early endosome is formed, which is usually elongated and located just under the plasma membrane. Later, inward migration and interaction with the Golgi apparatus and endoplasmic reticulum leads to the formation of a more rounded late endosome. Endocytosis of endosomal membranes begins the formation of intraluminal vesicles (ILVs) and maturation of early endosomes into late endosomes that then give rise to MVBs. Microvesicular bodies are thus endosomal organelles whose membranes invaginate to create multiple ILVs, all enclosed by a single membrane. The process of MVB membrane invaginations causes the engulfment of cytosolic components, including proteins, lipids, and nucleic acids.

There are two pathways to MVB formation; endosomal complexes required for transport (ESCRT)-dependent and ESCRT-independent. In the first pathway, MVB formation requires four protein complexes (0, I, II, III) from the ~30 proteins of ESCRT. Associated proteins of ESCRT are VTA1, VPS4, and PDCD6IP/ALIX. ESCRT detects and hence signals the internalization of ubiquitinated proteins on MVB membranes. Specifically, ESCRT 0 sequesters ubiquitinated membrane proteins into the endosome/MVB, while complexes I and II mediate membrane budding and finally complex III pinches of the ILVs into MVBs [6]. The ceramide pathway mediates the alternate ESCRT-independent pathway of MVB formation.

The signals required for exosome formation are not yet well understood; however, tumor microenvironment appears to play a role, because growth factor signaling increases both early endosomal and MVB formation. Also activated p53 increases exosomal release through induction of TSAP6 and CHMP4C involved in exosomal production.

There may be different populations of MVBs with two different fates. In general MVBs are indeed pre-lysosomal structures involved in protein sorting, recycling,

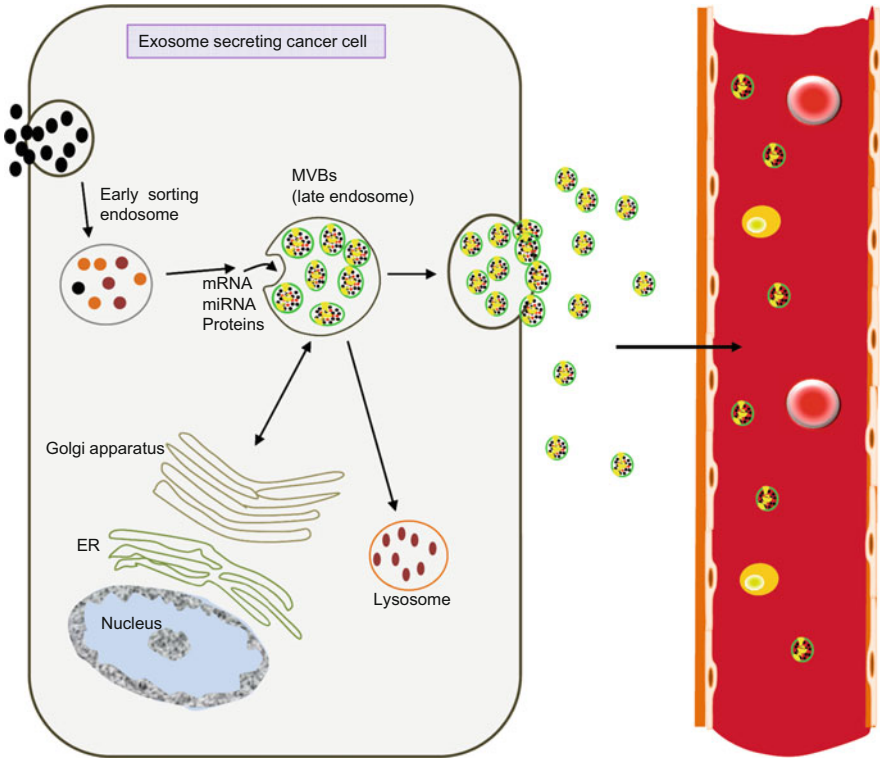

Fig. 9.1 Exosome biogenesis. Endocytosis of microenvironmental molecules leads to the formation of early endosome. Sorting and other modifications involving bidirectional vesicular exchange with the Golgi apparatus transforms early endosomes into late endosomes or microvesicular bodies (MVBs). Late endosome membrane invaginations and budding lead to intraluminal vesicle formation, which are released as exosomes when the MVB fuses with the plasma membrane. Alternatively they could fuse with lysosomes for degradation

transport, storage, and release. Thus, they usually will fuse with lysosomes leading to degradation. But also, MVBs that are positive for tetraspanin CD63, LAMP1, and MHC class II fuse with the plasma membrane. This second pathway is associated with the release of exosomes into the extracellular milieu.

Although they may be released constitutively, MVBs are transported toward the plasma membrane. This transport and their fusion and subsequent release also involve the activity of RAB27A and RAB27B [7] or the interactions of RAB11 and RAB35 with GTPase TBC1 domain family member 10A-C (TBC1D10A-C) [8].

9.4.2 Exosomal Content

Exosomes contain unique integral membrane proteins as well as internal cargo. Nucleic acids, proteins, metabolites, and all possible cytosolic contents can be encased in exosomes. However, heat shock proteins, major histocompatibility complexes, signal transduction ligands, and receptors among numerous other macromolecules have been identified as exosomal contents or anchored onto their membranes. Indeed exosomes contain over 4000 different proteins. They have universal or shared proteins, which include TSG101, ALIX involved in ESCRT, RAB11, RAB7, RAB2, and ANNEXINS used in intracellular transport and membrane fusion, as well as HSP70, HSP90, and the tetraspanins (CD9, CD63, CD81). In addition, exosomes possess cell-specific and tumor-associated antigens. For example, MET oncoproteins, mutated KRAS, tissue factor, p53, EGF, and FGF as well as numerous other proteins are incorporated in various cancer cell-derived exosomes. They also contain cytokines (e.g., TNF, TGFβ), cytokine receptors (e.g., transferrin receptor 2), and lysosomal proteins (e.g., CD68, CD61). Additionally, they contain nucleic acids, primarily mRNAs and miRNAs, but also DNA. Exosomes, thus, protect these nucleic acids from nuclease digestion. They are poor in rRNA, which is a major constituent (~80 %) of total cellular RNA that confounds RNA analyses in whole cell extracts. Therefore exosomal analyses should increase the sensitivity of tumor RNA detection. It is postulated that exosomal contents are not randomly packaged. Indeed exosomal contents of biomolecules often differ from those of the parental cell of origin, suggesting the selective packaging of their cargo. This cargo specificity is demonstrated in some cancers, where unique miRNAs have been profiled.

9.5 Cancer-Derived Microvesicles

Cancer-derived microvesicles (cdMVs) are distinct from exosomes (<100 nm) and apoptotic bodies (1000–3000 nm) by size, density, mode of formation, lipid composition, and shape (microvesicles tend to have irregular shapes), but carry similar cargo as exosomes. Cancer-derived microvesicles are submicron-sized (100–1000 nm) cellular vesicular fragments usually released by cancer cells. Microvesicles may be cancer-specific structures because normal parental cells appear not to release them. Besides, their levels mirror tumor progression being more abundant in circulation of patients with invasive cancers than those with non-metastatic cancers. Microvesicles are found in several body fluids including serum, plasma, urine, saliva, CSF, ascitic fluids, sputum, and bronchial fluid. Circulating cdMVs are elevated in patients with several types of cancer and the levels drop following surgical removal of tumors. Interestingly, circulating exosomes are rapidly cleared from the systemic circulation by various tissues including the lung, liver, kidneys, and spleen. Cancer cells shed MVs into body

fluids and some of these have specific molecular markers of cancer, such as HER2, EpCAM, VEGF, CCr6, and MMPs. Cancer cells communicate among themselves and to nearby and distant normal cells via microvesicular contents. Similar to exosomes, cdMVs contain nucleic acids (DNA and RNA), peptides/proteins, and lipids that they are able to transfer horizontally between nearby cells and to those at distant sites. Such information transfer could be involved in drug resistance and acquisition of aggressive and metastatic phenotypes and can even induce transformation of normal cells.

9.5.1 Microvesicular Biogenesis

Cancer-derived microvesicles are released by reverse shedding or budding and fission of plasma membrane. Many MVs express phosphatidylserine on their surfaces that are absent in exosomes, which tend to contain markers of endocytosis such as HSP73 and tetraspanins. The formation of microvesicles involves coordinated mechanisms of outward membrane budding and pinching off. In addition to their content of various signaling proteins, cdMVs also contain molecules involved in their biogenesis. An important molecule identified to play a role in MV biogenesis belongs to the family of RAS-related small GTP-binding protein, ARF6 [9]. ARF6 is more commonly associated with larger (~200 nm) cdMVs than those that resemble exosomes. ARF6-mediated endosomal recycling enables selective packaging of proteins into cdMVs. The complex functions of ARF6 include selective regulation of recycling of vesicles with targeted protein cargo to the cell surface for packaging into cdMVs. Activated ARF6 is implicated in endosomal recycling, actin remodeling at the cell periphery, and hence acquisition of metastatic potential by the cancer cell. Thus, ARF6 functions to direct cargo to cell periphery to be incorporated into MVs. In addition to proteins, nucleic acids are also present in cdMVs, and these may be trafficked to cdMVs in protein or lipid complexes. Other players of cdMV biogenesis are cell membrane content of cholesterol, because its depletion inhibits MV formation. Finally, fission of cdMVs requires actin-myosin-based contraction, which is also probably controlled by ARF6 activity.

9.6 Mechanisms of Extracellular Vesicle Uptake by Target Cell

Extracellular vesicles can act in an autocrine, paracrine, or endocrine fashion. Irrespective of how they act, EVs interact with the recipient cell in a number of ways. The mechanisms of exosomal uptake are best understood and will be discussed (Fig. 9.2). Targeting and fusion of exosomes to recipient cell can occur

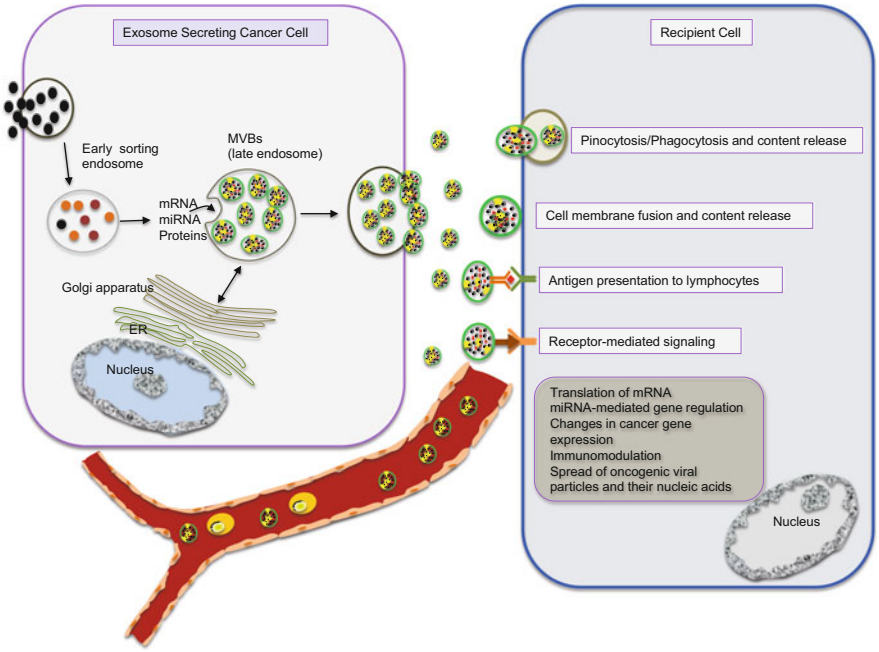

Fig. 9.2 Exosome uptake and functions in recipient cells. Exosomes and EVs are taken up by cell through several mechanisms including cytokine-mediated ICAM1 release, tetraspanin-integrin complex interaction, actin-cytoskeletal changes, PI3K signaling-mediated phagocytosis, macropinocytosis, and receptor-mediated endocytosis

in a number of ways. The pro-inflammatory tumor microenvironment can cause cytokine release that could act on recipient cells to express ICAM-1 to mediate exosome adhesion. Tetraspanin-integrin complex interaction can mediate target cell fusion as well. These target cell attachments can then lead to direct exosomal content release into the recipient cell. Additionally, exosomes can be internalized via actin-cytoskeletal mechanism and PI3K-dependent phagocytosis. Other mechanisms of exosomal uptake include macropinocytosis and receptor-mediated (clathrin, lipid raft, or caveolin-mediated) endocytosis. The internalization via lipid-dependent endocytosis can be enhanced by increased content of sphingomyelin/ganglioside GM3 on exosomal membrane. Putatively, the hypoxic tumor microenvironment (decreased pH) causes changes in the composition and integrity of exosomal membrane that enhances membrane fusion as well. The specific pathway adopted by a particular target cell may depend on the specific expressed surface molecules and microenvironmental conditions.

9.7 Biological Functions of Cancer-Derived Extracellular Vesicles

Exosomes and cdMVs can traffic nucleic acids such as mRNA, miRNA, and mutated DNA, as well as oncoproteins to nearby and distant cells. These horizontal molecular transfers can modulate cellular responses in transformed and even untransformed cells. Such intercellular communication serves numerous oncogenic functions within the tumor microenvironment and at distant sites. Within the tumor microenvironment and at the premetastatic site, exosomes promote extracellular matrix remodeling, angiogenesis, fibroblast activation, differentiation of fibroblast and mesenchymal cells into myofibroblasts, and the recruitment of bone marrow-derived cells to help premetastatic niche preparation and angiogenesis. Additionally, they can induce tumor cell proliferation and distribution or sharing of oncogenic proteins as well as chemoresistance factors. Cancer cells also use exosomal communication to stimulate and recruit cells needed for tumor invasion and dissemination to distant sites and sustain the functions of cancer stem and progenitor cells.

9.7.1 Immunosuppressive Effects of Extracellular Vesicles

The cancer cell has unique antigens that the normal immune system recognizes as foreign and hence mounts efficient immunity against (immune-surveillance). To maintain efficient tumorigenicity, cancer cells partly use exosomes to counter the antitumor immune mechanisms mounted by the body (Fig. 9.3). For instance, exosomes can enhance the proliferation of regulatory T (Treg) cells that counteract the antitumor activities of T cells. Exosomes from cancer cells can signal using Fas ligand to induce apoptosis in tumor-reactive CD8+T cells. Equally evidenced is the tumor-derived exosomal silencing of toll-like receptor (TLR) expression by microphages. Tumor-derived exosomes can activate myeloid-derived suppressive cells (MDSCs) from bone marrow to exert immunosuppressive effects on T-cell responses. Moreover, these exosomes induce IL-6 production by MDSCs through HSP72-mediated activation of TLR2 [10]. MDSCs then express receptors for IL-6 and hence undergo an autocrine signaling leading to phosphorylation of STAT3, which enhances their immunosuppressive activities [11]. In mouse model of breast cancer, the inhibition of exosomal release by blockage of Rab27a reduced neutrophil recruitment leading to decreased tumor growth and metastasis. Microvesicles from lymphoblastoma cells contain within them latent membrane protein 1 (LMP1) that inhibits leukocyte proliferation. Natural killer (NK) cells can mount efficient antitumor immunity. However, tumor-derived exosomes exert negative effects on NK cells. Tumor-derived exosomes reduce proliferation of NK cells and also block their activation by IL-2.

Surprisingly, exosomes from cancer cells can exert *antitumor immunity on cancer cells* as well. Exosomes from cancer cell lines can activate NK cells via

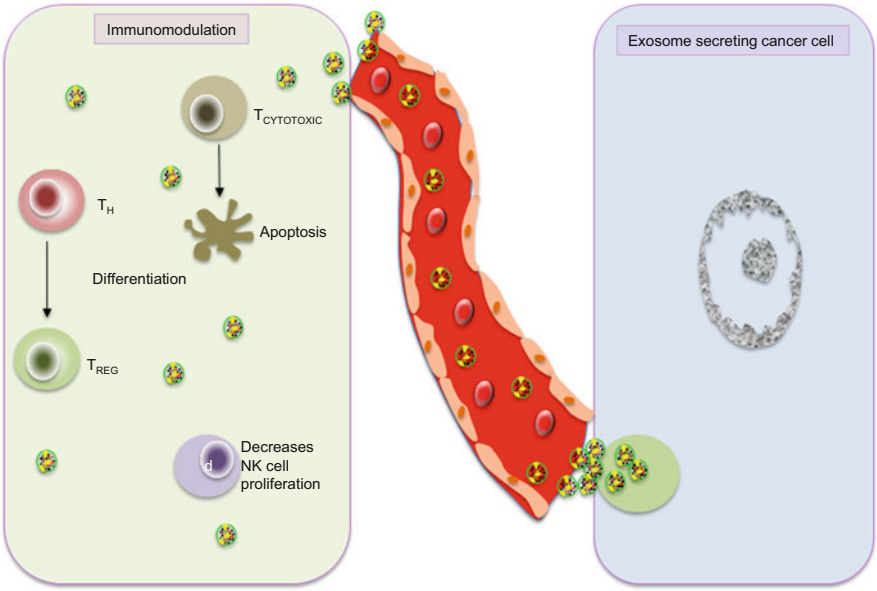

Fig. 9.3 Modulation of the immune system by exosomes. Exosomal content can induce immunosuppressive effects

HSP70, enabling them to kill cancer cells. In a mouse melanoma cell model, HSP70 activated NK cells, and this was associated with reduced primary tumor size and metastasis. Also exosomes from NK cells harbor perforins, CD56, and granzymes that are able to destroy cancer cells (similar to cytotoxicity by CD8+ T cells). Exosomes are able to incorporate cancer-specific antigens and deliver them to antigen-presenting cells. Because the antigens are delivered into the cytosol, they are processed via the MHC class I pathway and presented to CD8+ cytotoxic T cells. Activated CD8+ T cells can then target the cancer cells expressing those antigens and kill them.

9.7.2 Horizontal Transfer of Mutated Genes or Proteins to Other Cells

In order to perpetuate their malignant potential, cancer cells exchange oncoproteins and drug-resistant proteins via exosomes. Additionally, these mutated genes and their proteins are used to influence the functions of other cell types. Extracellular vesicles thus carry mutated oncogenes, including *KRAS*, *NRAS*, *MET*, *KIT*, *EGFR*, and *EGFRvIII*, and tumor suppressor genes such as *PTEN*, *TP53*, and *DKK4*, among several others. Exosomes carry on their membranes various signaling molecules and receptors and, in their internal milieu, a diverse array of biologically

active oncoproteins. For instance, colon cancer cells with mutant *KRAS* can secrete exosomes carrying this protein [12]. These exosomes can be internalized by nearby cancer cells harboring wild-type *KRAS*. The acquisition of this mutant protein enhances malignant behavior. The transfer of aggressive phenotype is illustrated in glioma cells. Aggressive glioma cells harboring mutant EGFRvIII are able to secrete exosomes carrying this receptor on their surface [13]. Transfer of EGFRvIII to cells lacking this mutant variant leads to induction of MAPK and AKT pathway activation, which then confer in them enhanced expression of survival genes and increased ability to grow and metastasize. Also demonstrated is the ability of cancer cells to share chemoresistance factors through exosomes. Exosomes from docetaxel-resistant prostate cancer cell lines are able to transfer MDR-1 drug transporter protein to drug sensitive cancer cells, thereby conferring in them docetaxel chemoresistance [14]. These examples illustrate the horizontal exchange of oncogenic molecules important in tumor progression and resistant mechanisms.

9.7.3 Modulation of Tumor Microenvironment and Premetastatic Niche

Premetastatic niche and tumor microenvironment must be primed and made conducive for spread and acceptance of incoming tumor cells. This process involves several microenvironmental changes that precede cancer cell docking and growth at the site (Fig. 9.4). The changes that occur include matrix remodeling, the establishment of a pro-inflammatory state, availability of bioactive oncogenes, neovascularization (vasculogenesis and angiogenesis), elevated redox state with increased ROS production, and hypoxia. Additionally, changes in various cell types including fibroblasts, endothelial cells, bone marrow-derived cells (BMDCs), and immune cells occur. The various proteins including signaling molecules, growth factors, proteases, glycoproteins, and nucleic acids, especially miRNA, mRNA, and retrotransposons in exosomes and cdMVs, are poised for such functions.

Cancer-derived microvesicles and exosomes contain abundant proteases that are released to degrade extracellular matrix (ECM) components. Metastasizing cells could use proteases in extracellular vesicles to degrade the ECM at the invading front producing a path of least resistance for their amoeboid movements to distant sites. Consistent with this, EVs from cancer cells contain various metalloproteinases and these vesicles easily rapture in the hypoxic and acidic cancer microenvironments.

New vessel formation is an important part of preparing a site for future colonization by cancer cells. To achieve this, cancer cells secrete exosomes that act in diverse ways to promote vasculogenesis and angiogenesis at their premetastatic niche. How the site is chosen is not so clear but could depend on the selective cargo in exosomes and hence their selective retention in a specific tissue. Thus, integrins in tumor-derived exosomes are capable of specifying metastatic organs

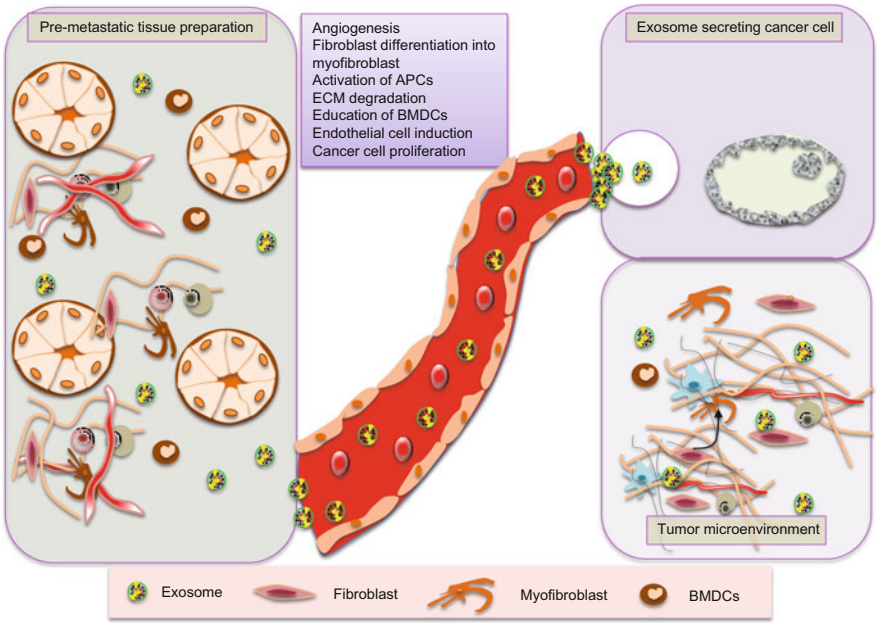

Fig. 9.4 Modulation of tumor microenvironment and premetastatic niche by exosomes

[15]. Neovascularization requires among other factors the recruitment and incorporation of bone marrow-derived VEGFR-2-expressing endothelial progenitors into the new vessel that are maintained by perivascular VEGFR-1 bone marrow-derived hematopoietic progenitors. Circulating chemokines that upregulate MMP, soluble KIT, and p27/p130 induce the expression of these molecules. Bone marrow-derived cells work in concert with tumor-associated stromal cells to establish new vessels and tumor growth.

Cytokines and growth factors mobilize BMDCs to establish the premetastatic niche. These BMDC-attractants include exosomes transported from the primary tumor. The exosomal content can educate BMDCs to assume a phenotype supportive of tumor invasion. For example, BMDCs can assume vasculogenic phenotypes induced by exosome-mediated MET receptor signaling. Melanoma cell-derived exosomes are able to recruit BMDCs to initiate niche preparation as well as increase endothelial cell membrane permeability in the lung, which is associated with increased lung metastasis [16].

The cell membrane receptor EGFR on cancer cells can be donated to endothelial cells via exosomes to activate VEGF/VEGFR-2 signaling pathway required for angiogenesis. Prostate cancer and mesothelioma cell line-derived exosomes contain biologically active TGFβ1, which can induce fibroblast differentiation into myofibroblast. Similarly exosomes from breast cancer cells can reprogram the differentiation of adipose tissue-derived mesenchymal stem cells into myofibroblast through SMAD-mediated signaling pathway. The specialized

myofibroblast is important in matrix remodeling and angiogenesis. Exosomes derived from CD105+ renal cancer cells can influence premetastatic niche formation by increasing the expression of MMP2, MMP9, and VEGFR. Additionally, these exosomes activate endothelial cells to reorganize themselves into a capillary-like format [17].

Other biomolecules involved in cancer pathophysiology are procoagulant factors. The coagulopathy observed in cancer patients is partly mediated by the release of cdMV and exosomal procoagulants that activate the clotting cascade. The exosomal disruption of host immune responses also facilitate premetastatic niche establishment. Thus, exosomes are critical factors in tumor metastasis. Therapeutic targeting of exosomes therefore holds promise in metastasis curtailment.

9.7.4 Extracellular Vesicles as Circulating Cancer Biomarkers

Extracellular vesicles have great potential as circulating cancer biomarkers for a variety of reasons. First, their levels are elevated in primary tumor tissue and in circulation and other body fluids of cancer patients compared to non-diseased controls. Second and importantly, their contents are selectively generated to enrich for biomolecules reflective of the specific cancers. Third, some biomolecules (e.g., RNA) that will otherwise be degraded by nucleases and proteases are protected within these vesicles, thus increasing their half-life and quality. Fourth, they are easily recovered from various body fluids, even following cryopreservation. Finally, their contents are more adapt to multiplex assays, which is an important requirement to overcome tumor heterogeneity in cancer diagnosis. All these features make cancer-derived extracellular vesicles excellent biomarker repository for the noninvasive detection and management of cancer.

The potential diagnostic and prognostic utilities of exosomes and cdMVs have been demonstrated (Table 9.2). MicroRNAs have well-established potential as circulating cancer biomarkers. While being protected by other factors, the further protection offered by their incorporation into exosomes and cdMVs make them even more attractive as circulating cancer biomarkers. Several different types of cancer have demonstrable differential exosomal miRNA signatures. For example, a number of circulating exosomal miRNAs have been associated with glioblastoma (miR-320, miR-574-3p, miR-RNU6-1) and cancers of the lung (miR-30a-3p, miR-100, miR-151a-5p, miR-154-3p, miR-200b-5p, miR-629), colorectum (let-7a, miR-21, miR-23a, miR-150, miR-223, miR-1229, miR-1246), and ovaries (miR-21, miR-141, miR-200a miR-200b miR-200c, miR-203, miR-205, miR-215) [18–21]. Importantly, exosomal levels are unaffected by collection strategies and storage times. Similarly, circulating exosomes and cdMVs from CRC patients harbor a set of miRNAs that achieved a sensitivity of 90 % in CRC detection compared to 16 % and 30.7 % for CA19.9 and CEA, respectively [19]. In

Table 9.2 Some cancer biomarkers (excluding miRNA) in circulating extracellular vesicles explored for cancer diagnosis, prognosis, and personalized care

Cancer	Biomarkers
Melanoma	Caveolin-1, MET, TYRP2, HSP70, HSC70, VLA-4, $BRAF^{V600E}$
Breast cancer	CD24
Lung cancer	EGFR, $BRAF^{V600E}$, CD91
CRC	CD9, CD147
Pancreatic cancer	KRAS, Glypican-1(GPC1)
Bladder cancer	EDIL-3/Del-1 (urinary exosomes)
Prostate cancer	PSA, survivin, TMPRSS:ERG, PCA3, PTEN
Glioblastoma	EGFRvIII
Ovarian cancer	TGFβ1, MAGE-3, MAGE-6
AML	CD34

glioblastoma patients, circulating exosomes harbor cancer-specific EGFR transcripts. Moreover, prostate cancer-associated biomarkers including PSA, PSMA, TPRSS:ERG, and PCA3 are found in prostate cancer-derived exosomes and cdMVs. A variety of exosomal proteins are cancer associated as well. Caveolin-1 and CD63 in exosomes and cdMVs were elevated in melanoma patients with diagnostic sensitivity of 69 % and specificity of 96.3 % [22]. Circulating melanoma-derived vesicles also demonstrate abnormal expressions of VLA-4, HSP70, HSP90, and TYRP2 [16]. The functions of the cargo of some EVs are associated with prognostic variables in cancer patients. In pancreatic cancer patients, increased plasma microparticles carrying tissue factor (an inhibitor of blood coagulation) are associated with aggressive tumor features such as poor differentiation, vascular invasion, and metastasis, and these findings are linked to thrombosis and increased mortality [23, 24].

9.7.5 *Extracellular Vesicles as* Vehicles *and* Targets *for Cancer Treatment*

Exosomes and cdMVs should serve the field of nanotechnology in oncology very well, as delivery vehicles of anticancer agents in view of their numerous advantages. These desirable features include their ability to use surface molecules to signal to, the ease of being internalized by, and the ability to discharge their cargo directly into other cancer cells. Exosomes can cross the blood-brain barrier, enabling targeted delivery of neuropharmacological agents. Engineering designs permit the loading of exosomes with miRNAs, siRNAs, and small reactive biomolecules, peptides, and ligands, among others to target delivery to cancer cells. They are also amenable to be used in gene therapy, whereby desired genes can be delivered to target cells without immune induction.

Tumor immunology is another area that will benefit from the use of tumor-derived extracellular nanovesicles. For instance, exosomes from human NK cells can kill tumor cells. In this field, the seminal works of Zitvogel et al. [25, 26] deserves attention. This group demonstrated that exosomes from dendritic cells express functional costimulatory and MHC class I and II molecules. Dendritic cell-derived exosomes (dexosome, Dex) pulsed with tumor antigens enabled these nanovesicles to induce antitumor cytotoxic T-cell responses, leading to suppression or eradication of established murine tumors. This finding sparked a phase I clinical trials in which Dex loaded MAGE-A3 peptides were used to vaccinate stage III and IV melanoma patients. After four intradermal injections carried out weekly apart, the delivery feasibility, safety, and antitumor activity of this Dex vaccine was demonstrated [27]. However, because of the weak antitumor immunity of these first generation Dex vaccines, a more potent second-generation Dex was developed [28]. The new Dex vaccines carry interferon-γ, which is critically needed for the expression of CD40, CD80, CD86, and CD54 by Dex, which enables potent CD8+ cell activation. This new Dex (interferon-γ-Dex) is being tested in a phase II trial as maintenance immunotherapy in stage IIIB and IV NSCLC patients on therapy. Autologously isolated and engineered interferon-γ-Dex is loaded with MAGE-3 and NY-ESO-1, MAGE-1, or MART-1 peptides for vaccination. This study, launched in 2009 at the Gustave Roussy and Curie Institute, has yet to have published data. These studies are being extended to include other cancers as well. For example, ascites-derived exosomes (Aex) that have maintained their antitumor immunomodulatory activities and carry tumor-associated CEA antigens are used in combination with GM-CSF to treat CRC patients. This agent is safe and well tolerated and has shown tumor-specific cytotoxic T lymphocyte responses.

Personalized medicine and oncology will equally benefit from the features and advantages offered by exosomes. Exosome harvesting and interrogation will complement the current "liquid biopsy" targets of CTCs and ctDNA, because exosomes much more represent tumor heterogeneity. These targets also offer real-time monitoring of tumor behavior to treatment, such as acquisition of novel mutations, lack of response, or even toxicity, all of which are cardinal features of personalized medicine. To illustrate this, the following companion diagnostic applications of exosomes and cdMVs should suffice. Melanoma patients eligible for targeted treatment with the *BRAF* kinase inhibitor, vemurafenib, require harboring the $BRAF^{V600E}$ mutation [29]. This mutation is shuttled in exosomes and cdMVs that enables treatment decision-making and longitudinal monitoring of vemurafenib efficacy. Blood-based companion diagnostic test for cetuximab (Erbilux) treatment is also notable, because cancer-derived nanovesicles carry both *EGFR* (drug target) and *KRAS* mutated genomes that indicate poor response. A subset of gastric cancers overexpress *HER2,* and these patients could benefit from targeted therapy. To this end, *HER2* is detectable in circulating gastric cancer-derived microvesicles. These findings can be extended to other cancers where genetic alterations needed for personalized medicine are available in tumor-derived extracellular vesicles.

9.7.6 Extracellular Vesicles as Mediators of Cancer Treatment Resistance

Cancer cells use several different mechanisms to develop resistance to chemotherapy. However, the contribution of exosomes to drug resistance includes drug expulsion from the cancer cell, drug sequestration within the cancer cell or outside of the cell, as well as the transfer of resistant factors to nonresistant cancer cells. Exosomes from cancer cell lines and primary tumors are loaded with various chemotherapeutic agents including cisplatin and doxorubicin that tumors easily develop resistance to. Men with castrate-resistant prostate cancer are often treated with docetaxel. Prostate cancer cells share resistance to docetaxel through exosomal transfer. Exosomes also can transfer resistance factors to chemoresponsive cells. Cancer cells share the MDR P-glycoprotein, an ATP-binding cassette transporter through exosomes. Some exosomes carry cancer surface molecules (receptors), so that their release by the cancer cell leads to drug sequestration away from the actual cancer cells being targeted. Microparticles have been shown to passively accumulate doxorubicin and daunorubicin into their nucleic acids and phospholipid cargo [30]. Additionally, the P-gp involved in drug efflux from cancer cells is oriented in an opposite direction (inside out) on microparticles such that they rather actively pump drugs into these microvesicles.

9.7.7 Extracellular Vesicles Promote Cancer Cell Migration and Invasion

Extracellular vesicles, especially exosomal cargo, can induce cancer cell migration and invasiveness as part of the metastatic cascade. Hendrix et al. demonstrated that secretory Rab27b from exosomes released by the breast cancer cell line MCF-7 enhanced tumor growth, matrix degradation, and cell migration [31]. EGF signaling is important in cancer biology. The full-length functional EGFR ligands are found in human breast and colorectal cancer cell-derived exosomes [32], and the concentration of exosomes containing this EGFR ligand, amphiregulin (AREG), correlated with the invasive propensities of CRC cells. Indeed, CRC cell-derived exosomes with mutant KRAS had high levels of AREG and exhibited aggressive invasive behaviors. One mode of enhancing tumor cell invasiveness is via modulation of the extracellular matrix (ECM). To this end, exosomes from Ras-transformed MDCK cell, 21D1, are enriched with MMP1, MMP14, and MMP19, as well as ADAM10 and ADAMTS1 [33]. The MMPs degrade ECM components including collagen, fibronectin, and gelatin to enhance tumor invasiveness. ADAM10 and ADAMTS1 proteolytically cleave cell adhesion molecules to enhance cancer cell invasiveness. Exosomes from melanoma and fibrosarcoma cells harbor and secrete MT1-MMP, which can degrade collagen [34]. HSP90, which can activate MMP2, is found in cancer-derived exosomes, and exosomal HSP70 promotes cell migration

[35]. KIT-containing exosomes from gastrointestinal stromal tumor cells can induce the release of MMP1 [36]. These KIT-containing exosomes could change recipient cell morphology, transform progenitor smooth muscle cells into tumor-promoting cells, and enhance tumor invasiveness. Thus, cancer-derived exosomes can modify the ECM, alter signaling pathways to promote cancer cell migration and invasion, and cause changes in recipient cell to release biomolecules that promote epithelial-mesenchymal transition.

9.7.8 Extracellular Vesicles in Cancer Stem Cell Biology

The importance of cancer stem cells and tumor-initiating cells in cancer progression from localized to invasive cancer cannot be overemphasized. One mode of achieving tumor progressive phenotype is through intercommunication networks between cancer-associated cells. These cancer-associated cells are found in the immediate and distant metastatic microenvironments and include cancer, stem, progenitor, endothelial, immune, fibroblast lineage, and smooth muscle cells as well as other stromal cells. Extracellular vesicles are involved in the intercommunication networks between the various cell types. In cancer biology, communications between cancer stem cells and the other cell types have been mostly studied. The effects of stem cell-derived extracellular vesicles on tumor microenvironment include metastatic niche preparation such as:

- Endothelial cell proliferation, differentiation, and angiogenesis.
- Myofibroblast and cancer-associated fibroblast differentiation involved in stromal changes.
- Cross-talk between stromal elements and cells.
- Information sharing between stem cell hierarchies in tumor microenvironment.
- Exchange of molecules involved in epithelial-to-mesenchymal transition.

Specific stem cell exosomal effects are demonstrated in a number of cancers. Exosomes are recognized to carry stem cell markers, including WNT3, Nanog, HOXB4, REX1, β-catenin, prominin1, Hedgehog, OCT4, CD44, and CD133 [37, 38], as well as signaling molecules involved in stem cell biology such as WNT ligands [39, 40]. Breast cancer stem cells use exosomes to channel hypoxic signals to stromal cells [41]. Bone marrow stem cell exosomes promote dormancy in breast cancer cells [42]. Renal cancer stem cells release EVs that interact with endothelial cells to initiate angiogenesis [17]. Vesicular transfer of transcription factors and proteins including WNT3 and OCT4 from embryonic stem cells to hematopoietic progenitors reprogrammed them into more primitive cells [43].

This communication is bidirectional. While less studied, exosomal signals from other cells influence cancer stem cell functions. Exosomes from stromal fibroblasts activated notch signaling in breast cancer cells that increased stem cell marker, aldehyde dehydrogenase expression. Also WNT signals from stromal fibroblasts could increase breast cancer cell metastasis to the lung [44]. Leukemic and

myeloma cells interact with stromal mesenchymal stem cells via microvesicular cargos such as receptor tyrosine kinase [45]. In addition to miRNAs, the stem cell communications mainly involve WNT, HH, and TGFβ signaling. WNT signaling is implicated in the establishment of "stemness" in cancer cells, and TGFβ in exosomes mediate stem cell differentiation.

9.7.9 Blockage of Extracellular Vesicle Release in Cancer Management

The emerging field of MV research has potential for cancer management. Undoubtedly, the contribution of these cancer-associated molecules to cancer progression, invasion, and metastasis is quite well established. The ability to therefore prevent the biosynthesis and/or release of various EVs can potentially thwart the deleterious effects of these cancer-derived vesicles. A number of agents have shown potential in this regard. Microparticle inhibitors, for instance, can prevent MP biosynthesis and/or release. In a number of cell types, ROCK inhibitor (Y-27632), pantethine, and calcium channel blockers prevent MP biosynthesis and release. Additionally, the calpain inhibitor II (ALLM) has shown significant blockage of MP production in both resting and cells activated with calcium ionophore (A23187) [46]. In this analyses, the ROCK inhibitor (Y-27632) prevented MP biosynthesis in only activated cells. These strategies as complementary cancer therapeutics hold potential.

9.8 Apoptotic Bodies

Apoptotic bodies are small, membrane-enclosed vesicles that are the final remnants of cellular apoptosis. Biologically, these vesicles function to prevent the release of toxic or immunogenic (e.g., viral or other microbial particles) contents that could otherwise induce immunologic cellular damage. Often, in a receptor-mediated process, phagocytes engulf these apoptotic bodies and subsequently may release their contents into body fluids. The contents of apoptotic bodies have been known for sometime and their functional associations speculated. The packaging of cellular content including intact organelles, chromatin with fragmented DNA and other nucleic acids, and their subsequent uptake and distribution by phagocytes is well known. DNA from apoptotic bodies, for instance, can be reused by nearby cells, and should these contain viral genomes, this will present an easy horizontal transfer to such cells without the need for specific receptors.

9.9 Oncosomes

Oncosomes are emerging as a potential clinically important class of cancer-derived-EVs. First coined by Al-Nedawi et al. in reference to EGFRvIII-containing glioma-derived EVs, this term is now gaining a broader meaning in oncology, as other cancer cells potentially release them as well [13]. Thus, Di Vizio and coworkers have also uncovered a related class of vesicles released by prostate cancer cells [5]. In view of their sizes (1–10 μm in diameter) and content, these have been named "large oncosomes." Unlike exosomes, oncosomes appear to be released primarily by aggressive cancer cells [47, 48].

The mechanisms mediating oncosome biogenesis are not fully elaborated; however, they are likely to be different from those involved in exosome biogenesis and may share some common features with the mechanism of MV release. Thus, identical to MVs, oncosomes are produced by pinching off of plasma membrane blebs from cancer cells. Non-apoptotic membrane blebs of migratory aggressive tumors produce oncosomes. Unlike the commonly understood single-cell mesenchymal movements of invasive cancers, some cancer cells migrate by amoeboid movement, involving lamellipodia. These membranous extensions on the leading fronts of such migratory cancer cells can be pinched off as oncosomes. While activated GTPase, RhoA, and its effector ROCK can induce the production of these membrane blebs, it is also possible that loss of diaphanous-related formin-3 (DIAPH3), a cytoskeletal regulator, plays a role as well. This is because cellular loss of DIAPH3 causes mesenchymal-to-amoeboid transition [49, 50]. In reference to large oncosomes, EGF and AKT signaling may be involved in their formation. Stimulation of prostate cancer cells with EGF and overexpression of AKT both promoted large oncosome production [5]. The mechanism involved in pinching off large oncosomes may resemble microvesicular production, because oncosomes also demonstrate increase expression of ARF6, which is involved in microvesicular biogenesis.

Although the biomolecular cargo of exosomes and oncosomes may differ, they appear to have some overlap. For instance, both contain miRNAs, and oncosomal EGFRvIII has also been associated with exosomes. Large oncosomes also carry other bioactive molecules such as hnRNP-K, pyruvate kinase M2 (PKM2), poly (A)-binding protein 1 (PABPC1), programmed cell death 6 interacting protein (PCD6IP or AlLIX) as well as metalloproteinases, and caveolin-1 [5]. Caveolin-1 is associated with metastatic prostate cancer, and circulating caveolin-1 containing oncosomes can discriminate between metastatic and organ-confined prostate cancer [48]. Large oncosomes are also demonstrable in tissue sections, and their presence is associated with advanced-stage metastatic diseases.

9.10 Quality Standards in Analysis of Extracellular Vesicles

Analytical standardization is a key requirement in all aspects of biomarker development. In the field of extracellular vesicle (EV) research, similar issues hamper the authenticity and comparison of various findings. In view of this, following the first ISEV meeting in 2012 in New York, a 2-day workshop involving participants from both academia and industry was organized to address the issue of standardization of EV isolation and downstream analysis [51]. The consortium addressed numerous issues including the choice of normal controls, physiologic variables that affect EV release, EV isolation and storage conditions, the choice of anticoagulants, sample draw, platelet removal, and technical issues pertaining to EV analyses using other body fluids. Salient issues in EV isolation and analysis using blood are summarized below. Pre-analytical precautions of other body fluids will be provided in their respective chapters and sections in other volumes of this series.

- Factors to be considered in selecting control subjects include age, gender, BMI, medication use, pregnancy status, fasting or postprandial state, illness, hospitalization, and physical activities, because all these can influence the levels of EV in circulation. Similarly, circadian rhythm should be noted and controlled for because it can affect platelet activation and release of EVs.
- Plasma is preferred to serum for RNA analysis. This is because the clotting mechanism causes platelet release of EVs that can account for over 50 % of the circulating EVs. However, serum will be preferable if the research question requires platelet-derived EVs.
- The choice of anticoagulant should be based on intended downstream applications/analyses of EVs. However, heparin should be avoided because it inhibits PCR and can bind to EVs and prevent their uptake by cells. Heparinase treatment may overcome these effects, and in studies of nucleic acids encased in EVs, extensive washes prior to nucleic acid extraction may be helpful.
- Platelets can be activated to release EVs due to pressure and other physical forces associated with the process of blood draw. Hence, a large bore needle (21 gauge) is recommended, and the procedure should be as gentle as practically possible.
- Collected blood should be handled gently as well. For example, blood should be inverted several times to mix anticoagulant rather than shaking. Preferably, blood should be processed within 30 min after collection.
- Centrifugation should be done quickly at room temperature using the same rotor and speed for all samples.
- Blood viscosity affects EV pelleting and hence in situations whereby blood viscosity may be increased (diseases associated with elevated ESR), dilution of blood before centrifugation should be considered.

- Hemolysis should be avoided because it can alter the concentration of nucleic acids. For example, miR-15b, miR-16, and miR-451 levels are elevated due to hemolysis and not just by disease state.
- It is recommended to proceed immediately with EV isolation following sample collection. However, experimental design and/or sample availability may hinder immediate processing, requiring storage for a period of time. There currently is no consensus as to which is the optimum storage temperature, but many investigators have reported on fluid storage at 4 °C, −20 °C, −80 °C, and −160 °C. The diverse storage conditions may also depend on the research question under investigation.
- Isolated EV should be stored in siliconized vessels to prevent loss of EV due to adherence to vessel surfaces. Samples should be resuspended in phosphate-buffered saline.
- Although freeze-thaw cycles may have minimal effects on EVs, it is prudent to minimize this. Additionally, freezing and thawing should be done rapidly to preserve EV morphology and functions.

9.11 Methods for EV Isolation

The methods adopted for EV enrichment and isolation are very identical to those employed for CTC work. These approaches equally exploit differences in the physical properties as well as uniquely expressed surface molecules. Thus, they can broadly be classified under filtration based on particle size, centrifugation based on differential densities, and capture based on different membrane-associated antigens.

9.11.1 Centrifugation

Centrifugation has been extensively used for EV enrichment and isolation. The application of ultracentrifugation or normal centrifugation with precipitation reagents enables enrichment of EVs in body fluids. Traditionally, differential centrifugation begins with spinning at low speeds of 300–500× *g* to remove cells, and progressively increasing speeds to 2000–20,000× *g* to remove cellular debris, and finally to ultra-speeds of 100,000–200,000× *g* to pellet EVs. Although a commonly used procedure, density-based centrifugation methods have some drawbacks. They are labor intensive, time consuming (takes ~5 h to complete), and yet associated with poor EV recovery. Sucrose gradient differential centrifugation improves EV yield and purity, but is also time consuming. Thus a number of commercial reagents (e.g., ExoQuick, ExoSpin, PureExo) are available for precipitation of EVs applying regular centrifugation speeds. While in general they provide higher EV recovery than simple ultracentrifugation, purity is low.

9.11.2 Filtration

The filtration through membranes of defined pore sizes aims at isolating EVs based on size. This method has been developed mostly for exosome isolation because of their defined sizes (30–150 nm) and probably the proven clinical relevance of their contents. Filtration through membranes with defined pores enables removal of cells and debris and thus enriching for particles with smaller than the defined pores. Various membrane materials including polycarbonate and polyvinylidene fluoride with pore sizes of 200 nm, 220 nm, and 450 nm have been used for EV isolation. Used in combination with ultracentrifugation produces higher quality and quantity of EVs than each used alone. Microfluidic approaches are also applied for EV isolation based on size. For example, Wang et al. developed a ciliated micropillar microfluidic device with porous silicon nanowires on the side walls for removal of cellular debris and proteins while trapping exosome-like lipid vesicles [52]. The trapped vesicles can then be removed by soaking the porous nanowires in phosphate-buffered saline.

9.11.3 Immunoaffinity

The uniquely expressed molecules on exosomal membranes including CD63, CD81, CD45b, CD9, CA125, EpCAM, E-cadherin, and IGF1Rα have been exploited as targets for immunoaffinity isolation. Various technological approaches including immunomagnetic bead capture and photosensitizer bead capture, as well as microfluidic devices and dielectrophoresis, have been deployed for EV isolation based on antigen expression. As examples, anti-EpCAM- and anti-CA125-coated magnetic beads were used to isolate EVs from just 30 μL of plasma for integrated protein analysis by chemifluorescence on this chip [53]. Another microfluidic chip developed by Shao et al. enabled the capture and analysis of EV RNA [54]. Microbeads coated with anti-EGFR could successfully (>93 % capture efficiency) capture exosomes from 100 μL of serum, and the entire process could be completed in 2 h. Tumor-derived exosomes could be captured and detected in real time using antibody microarray with surface plasmon resonance [55]. Various microfluidic technologies are in development for EV isolation and analysis for possible translation into point-of-care utility.

9.12 Analyses of EVs

Exosomes derived from diseased and healthy cells differ in regard to their structure, quantity, and contents. Indeed, the levels of circulating exosomes are higher in cancer patients than control healthy individuals. In addition to differences in the

types and amounts of exosomal contents between cancer and normal cells, communications between cancer cells and normal cells via exosomes mediates malignant progression. Exosomes have therefore been of interest in the search for cancer biomarkers and for the study of cancer biology. Therefore isolated exosomes are subjected to structural analysis and molecular profiling of their contents, while in vivo and in vitro molecular imaging is used to study the interactions of exosomes with their microenvironments.

Various technologies are employed for analysis of the sizes, size distribution, concentrations, and surface topology of exosomes. Both transmission and scanning electron microscopy and atomic force microscopy are used to study exosomal sizes, distribution, morphology, and surface markers. However, the most commonly applied method for measurement of exosomal sizes and concentration is nanoparticle-tracking analysis. This procedure involves exosomal isolation and purification (to avoid interference by impurities) followed by laser beam passage through the sample. The mean velocities achieved by each particle as a consequence of beam scattering are used to estimate their sizes, distribution, and concentration. The speed attained as a consequence of size (Brownian motion) is used for size determination. Another laser light scattering technique for measurement of exosomal size distribution also based on Brownian motion is dynamic light scattering. In this process, light scattering by particles at specific angles are detected, and the changes in light intensities enable size and distribution of exosomes to be determined. Dynamic light scattering provides information on average sizes of particles with similar light scattering properties.

Surface markers including MHC, tetraspanins, and HSPs can be studied with various techniques such as flow cytometry and ELISA. Fluorescent proteins tagged to exosomal surface markers such as tetraspanin CD63 enable in vivo and in vitro molecular imaging of exosomal migratory behavior and intercellular communication, while exosomal contents including nucleic acids, proteins, and other biomolecules are studied using conventional molecular biologic approaches.

It should be noted, however, that there are several other novel advanced technologies being developed and optimized for the study of exosomes. These include Raman microspectroscopy, fluorescence fluctuation spectroscopy, micronuclear magnetic resonance, surface plasmon resonance, nano-plasmonics, small angle X-ray scattering, nanoshearing, and several others.

9.13 Commercial Applications

There are currently two companies developing products for clinical applications based on microvesicles. These companies are Caris Life Sciences and Exosome Diagnostics.

Caris Life Sciences has developed the Carisome® Technology that uses a proprietary method to capture and analyze circulating microvesicles (cMVs). The method uses antibodies to isolate cMVs followed by their detection with

phycoerythrin-labeled anti-MV marker such as tetraspanin. For example, in prostate cancer, antibodies to PSMA, PCSA, and B7-H3 are used for the capture of prostate cancer-derived cMVs, which are then detected by anti-tetraspanin antibodies. Signal threshold indicative of the levels of circulating prostate cancer-derived MVs are then used to determine the risk of cancer. This assay has demonstrated superiority to PSA testing for prostate cancer detection, achieving a sensitivity and specificity of 85 % and 86 %, respectively. A better performance (sensitivity of 91 %, specificity of 85 %) was realized in men with elevated PSA (>4.0 ng/mL). The technology is being used to develop products for other cancers as well.

Exosome Diagnostics is championing the development of a pipeline of clinical products based on exosomal and microvesicular content. The company has intellectual property related to the isolation, extraction, and content analyses of exosomes and microvesicles. Developed and commercialized as laboratory developed test is the ExoDx™ *lung* (*ALK*), a qPCR-based test for the detection of five EML4-ALK fusion transcripts in exosomal RNA. This fusion gene is present in 3–4 % of NSCLC patients and is important in selecting patients who will benefit from treatment with an ALK kinase inhibitor such as crizotinib. Other plasma-based tests under development are ExoDx™ *lung* (*EGFR*) and ExoDx™ *lung* (*T790M*), both of which are important in personalized oncology.

9.14 Summary

- Cancer and healthy cells communicate partly through the release of membrane-enclosed structures, simply referred to as extracellular vesicles (EVs).
- Based on sizes, markers, mode of release, and even the releasing cell type, EVs are variously called exosomes, microvesicles, microparticles, apoptotic bodies, prostasomes, and oncosomes. Among them, exosomes are the best characterized.
- Exosomes are produced via the degradative pathway of endocytosed vesicles. Instead of late endosomal fusion with lysosomes, exosomal biogenesis causes invaginations of the endosomal membrane to form microvesicular bodies (MVBs). Expectedly, this process leads to selective and random packaging of cytosolic contents into the MVBs. Subsequent fusion with the plasma membrane leads to the release of exosomes.
- The contents of EVs determine their functions. In cancer, they are involved in several oncogenic functions including the horizontal transfer of oncogenic molecules (e.g., drug-resistant factors), immunosuppressive effects, modulation of tumor microenvironment, and premetastatic niche preparation.
- Apoptotic bodies are produced during cellular apoptosis. Because apoptosis is not a major cancer cell death process, apoptotic bodies are not a major contributor to circulating cancer biomarkers.

- Oncosomes are large blebs pinched off from the leading edges of metastatic cancer cells. Their functions are similar to exosomes.
- EV isolation and characterization are similar to those of CTCs. An important function of exosomes is selective packaging and protection of miRNA, which are emerging as powerful circulating cancer biomarkers.

References

1. Friend C, Marovitz W, Henie G, et al. Observations on cell lines derived from a patient with Hodgkin's disease. Cancer Res. 1978;38:2581–91.
2. Poste G, Nicolson GL. Arrest and metastasis of blood-borne tumor cells are modified by fusion of plasma membrane vesicles from highly metastatic cells. Proc Natl Acad Sci USA. 1980;77:399–403.
3. Ronquist G, Hedstrom M. Restoration of detergent-inactivated adenosine triphosphatase activity of human prostatic fluid with concanavalin A. Biochim Biophys Acta. 1977;483:483–6.
4. Ronquist G, Brody I, Gottfries A, Stegmayr B. An Mg^{2+} and Ca^{2+}-stimulated adenosine triphosphatase in human prostatic fluid: Part I. Andrologia. 1978;10:261–72.
5. Di Vizio D, Kim J, Hager MH, et al. Oncosome formation in prostate cancer: association with a region of frequent chromosomal deletion in metastatic disease. Cancer Res. 2009;69:5601–9.
6. Hanson PI, Cashikar A. Multivesicular body morphogenesis. Annu Rev Cell Dev Biol. 2012;28:337–62.
7. Ostrowski M, Carmo NB, Krumeich S, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. Nat Cell Biol. 2010;12:19–30. Sup 11–3.
8. Hsu C, Morohashi Y, Yoshimura S, et al. Regulation of exosome secretion by Rab35 and its GTPase-activating proteins TBC1D10A-C. J Cell Biol. 2010;189:223–32.
9. Muralidharan-Chari V, Clancy J, Plou C, et al. ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles. Curr Biol. 2009;19:1875–85.
10. Chalmin F, Ladoire S, Mignot G, et al. Membrane-associated Hsp72 from tumor-derived exosomes mediates STAT3-dependent immunosuppressive function of mouse and human myeloid-derived suppressor cells. J Clin Invest. 2010;120:457–71.
11. Wieckowski EU, Visus C, Szajnik M, et al. Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8+ T lymphocytes. J Immunol. 2009;183:3720–30.
12. Demory Beckler M, Higginbotham JN, Franklin JL, et al. Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. Mol Cell Proteomics. 2013;12:343–55.
13. Al-Nedawi K, Meehan B, Micallef J, et al. Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. Nat Cell Biol. 2008;10:619–24.
14. Corcoran C, Rani S, O'Brien K, et al. Docetaxel-resistance in prostate cancer: evaluating associated phenotypic changes and potential for resistance transfer via exosomes. PLoS One. 2012;7:e50999.
15. Hoshino A, Costa-Silva B, Shen TL, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015;527:329–35.
16. Peinado H, Aleckovic M, Lavotshkin S, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat Med. 2012;18:883–91.
17. Grange C, Tapparo M, Collino F, et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. Cancer Res. 2011;71:5346–56.

18. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecol Oncol. 2008;110:13–21.
19. Ogata-Kawata H, Izumiya M, Kurioka D, et al. Circulating exosomal microRNAs as biomarkers of colon cancer. PLoS One. 2014;9:e92921.
20. Cazzoli R, Buttitta F, Di Nicola M, et al. microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. J Thorac Oncol. 2013;8:1156–62.
21. Manterola L, Guruceaga E, Gallego Perez-Larraya J, et al. A small noncoding RNA signature found in exosomes of GBM patient serum as a diagnostic tool. Neuro Oncol. 2014;16:520–7.
22. Logozzi M, De Milito A, Lugini L, et al. High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. PLoS One. 2009;4:e5219.
23. Thaler J, Ay C, Mackman N, et al. Microparticle-associated tissue factor activity in patients with pancreatic cancer: correlation with clinicopathological features. Eur J Clin Invest. 2013;43:277–85.
24. Bharthuar A, Khorana AA, Hutson A, et al. Circulating microparticle tissue factor, thromboembolism and survival in pancreaticobiliary cancers. Thromb Res. 2013;132:180–4.
25. Zitvogel L, Regnault A, Lozier A, et al. Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. Nat Med. 1998;4:594–600.
26. Zitvogel L, Fernandez N, Lozier A, et al. Dendritic cells or their exosomes are effective biotherapies of cancer. Eur J Cancer. 1999;35 Suppl 3:S36–38.
27. Escudier B, Dorval T, Chaput N, et al. Vaccination of metastatic melanoma patients with autologous dendritic cell (DC) derived-exosomes: results of the first phase I clinical trial. J Transl Med. 2005;3:10.
28. Viaud S, Ploix S, Lapierre V, et al. Updated technology to produce highly immunogenic dendritic cell-derived exosomes of clinical grade: a critical role of interferon-gamma. J Immunother. 2011;34:65–75.
29. Thakur BK, Zhang H, Becker A, et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. Cell Res. 2014;24:766–9.
30. Gong J, Luk F, Jaiswal R, et al. Microparticle drug sequestration provides a parallel pathway in the acquisition of cancer drug resistance. Eur J Pharmacol. 2013;721:116–25.
31. Hendrix A, Westbroek W, Bracke M, De Wever O. An ex(o)citing machinery for invasive tumor growth. Cancer Res. 2010;70:9533–7.
32. Higginbotham JN, Demory Beckler M, Gephart JD, et al. Amphiregulin exosomes increase cancer cell invasion. Curr Biol. 2011;21:779–86.
33. Tauro BJ, Mathias RA, Greening DW, et al. Oncogenic H-ras reprograms Madin-Darby canine kidney (MDCK) cell-derived exosomal proteins following epithelial-mesenchymal transition. Mol Cell Proteomics. 2013;12:2148–59.
34. Hakulinen J, Sankkila L, Sugiyama N, et al. Secretion of active membrane type 1 matrix metalloproteinase (MMP-14) into extracellular space in microvesicular exosomes. J Cell Biochem. 2008;105:1211–8.
35. McCready J, Sims JD, Chan D, Jay DG. Secretion of extracellular hsp90alpha via exosomes increases cancer cell motility: a role for plasminogen activation. BMC Cancer. 2010;10:294.
36. Atay S, Banskota S, Crow J, et al. Oncogenic KIT-containing exosomes increase gastrointestinal stromal tumor cell invasion. Proc Natl Acad Sci USA. 2014;111:711–6.
37. Marzesco AM. Prominin-1-containing membrane vesicles: origins, formation, and utility. Adv Exp Med Biol. 2013;777:41–54.
38. Garnier D, Magnus N, Meehan B, et al. Qualitative changes in the proteome of extracellular vesicles accompanying cancer cell transition to mesenchymal state. Exp Cell Res. 2013;319:2747–57.
39. Koch R, Demant M, Aung T, et al. Populational equilibrium through exosome-mediated Wnt signaling in tumor progression of diffuse large B-cell lymphoma. Blood. 2014;123:2189–98.
40. Gross JC, Chaudhary V, Bartscherer K, Boutros M. Active Wnt proteins are secreted on exosomes. Nat Cell Biol. 2012;14:1036–45.

41. Papi A, De Carolis S, Bertoni S, et al. PPAR gamma and RXR ligands disrupt the inflammatory cross-talk in the hypoxic breast cancer stem cells niche. J Cell Physiol. 2014;229:1595–606.
42. Ono M, Kosaka N, Tominaga N, et al. Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. Sci Signal. 2014;7:ra63.
43. Ratajczak J, Miekus K, Kucia M, et al. Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery. Leukemia. 2006;20:847–56.
44. Luga V, Zhang L, Viloria-Petit AM, et al. Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. Cell. 2012;151:1542–56.
45. Roccaro AM, Sacco A, Maiso P, et al. BM mesenchymal stromal cell-derived exosomes facilitate multiple myeloma progression. J Clin Invest. 2013;123:1542–55.
46. Roseblade A, Luk F, Ung A, Bebawy M. Targeting microparticle biogenesis: a novel approach to the circumvention of cancer multidrug resistance. Curr Cancer Drug Targets. 2015;15:205–14.
47. Di Vizio D, Morello M, Dudley AC, et al. Large oncosomes in human prostate cancer tissues and in the circulation of mice with metastatic disease. Am J Pathol. 2012;181:1573–84.
48. Morello M, Minciacchi VR, de Candia P, et al. Large oncosomes mediate intercellular transfer of functional microRNA. Cell Cycle. 2013;12:3526–36.
49. Yilmaz M, Christofori G. Mechanisms of motility in metastasizing cells. Mol Cancer Res. 2010;8:629–42.
50. Hager MH, Morley S, Bielenberg DR, et al. DIAPH3 governs the cellular transition to the amoeboid tumour phenotype. EMBO Mol Med. 2012;4:743–60.
51. Witwer KW, Buzas EI, Bemis LT, et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. J Extracell Vesicles. 2013;2.
52. Wang Z, Wu HJ, Fine D, et al. Ciliated micropillars for the microfluidic-based isolation of nanoscale lipid vesicles. Lab Chip. 2013;13:2879–82.
53. He M, Crow J, Roth M, et al. Integrated immunoisolation and protein analysis of circulating exosomes using microfluidic technology. Lab Chip. 2014;14:3773–80.
54. Shao H, Chung J, Lee K, et al. Chip-based analysis of exosomal mRNA mediating drug resistance in glioblastoma. Nat Commun. 2015;6:6999.
55. Zhu L, Wang K, Cui J, et al. Label-free quantitative detection of tumor-derived exosomes through surface plasmon resonance imaging. Anal Chem. 2014;86:8857–64.

Chapter 10
The Critical Path to Successful Biomarker Development

Key Topics

- Challenges with Biomarker development
- Biomarker methods development and validation
- Biomarkers and surrogate markers
- Role of the Food and Drug Administration (FDA) in biomarker development
- The FDA's Critical Path Initiative
- The FDA's Drug Development Tools
- Evaluating clinical utility of biomarkers
- Biomarker adoption into clinical practice

Key Points

- The paucity of cancer biomarker products in the clinic is partly attributable to the inherent challenges in biomarker development. These challenges include tumor biologic variables, pre-analytical and analytical variables, study design issues, statistical and bioinformatics issues, and stakeholder factors. All these factors often lead to discovery of putative biomarkers that fail large-scale validation studies and hence unqualified for regulatory approval.
- To be clinically valuable, a biomarker should possess strong analytical method validation and meet the rigorous qualification standards for safety and effectiveness.
- Regulatory bodies, including the FDA, have developed programs to facilitate accelerated product development. These collaborative efforts among the FDA, industry, academia, stakeholders, and patient advocacy groups

(continued)

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_10

include the biomarker qualification program and the qualification process for drug development tools.

10.1 Introduction

Cancer biomarkers are important integral part of cancer research and oncologic practice. First, they can serve as laboratory tools for making clinical decisions such as cancer risk predictions, diagnosis, prognosis, treatment predictions, and monitoring for therapy response and disease relapse. However, to be even of much value, a clinical biomarker should be applied to diseases that have treatment interventional agents. Alternatively, a useful biomarker should have biologic implications in disease pathogenesis such that, in the absence of a therapeutic agent, targets can be identified for new drug development. Second, biomarkers are useful in pharmaceutical drug development. Indeed, 30–50 % of all biomarkers are coupled to drug development. Many drug development programs rely on model systems such as animal and cell lines, as well as clinical judgments. While these tools and metrics are useful in the drug development process, the adoption and use of biomarkers has resulted in cost reductions and accelerated development timelines. Specifically, prognostic, predictive, and pharmacodynamic biomarkers play unique roles that not only facilitate the development process but also can identify unique patient groups for targeted drug development. A prognostic biomarker may be used to select a group of patients who are more likely to have aggressive disease for a specific drug development program. In this situation, the ability to demonstrate safety and reductions in case fatalities among this cohort may enhance the drug approval process. A predictive biomarker may also be used to enrich for patients who are more likely to respond to a drug in a specific manner, to be included in a drug development program. Unique gene mutations, for instance, may confer sensitivity or resistance to a specific interventional agent. Finally, pharmacodynamic biomarkers are adept to early identification of drug toxicities prior to clinical evidence of such events. These pharmacodynamic biomarkers are therefore useful in phase II clinical trials to help establish safety and dose standards for phase III studies. Thus, the entire drug development process benefits from valid biomarkers.

To be useful, however, cancer biomarkers must be validated and qualified to meet safety and effectiveness standards within the context of use. Unfortunately there is a disconnection between the plethoric number of biomarkers discovered and those that end up as clinical products, or in use for drug development. This paucity of biomarker product development is partly due to the real challenges faced by the industry. Multiple factors impede the biomarker development process. Various issues including study design, pre-analytical and analytical issues (e.g., sample types and adopted technologies), as well as inherent issues with tumor biology, such as heterogeneous molecular cellular types in the primary cancer,

and clonal diversity due to selection pressures from the metastatic cascade and/or therapy. To enhance the chances for success therefore, biomarker development ought to be a collaborative process. Patients, their clinicians, biomarker scientists, biostatisticians and bioinformaticians, industry sponsors, stakeholders, and members from regulatory agencies should all be part of the biomarker development team from start to finish.

Biomedical devices and other products intended for public consumption must meet the claimed safety and effectiveness standards of the US Food and Drug Administration (FDA), because part of the mandate of the FDA is to ensure public safety. However, the organization is also tasked with improving public health, which includes accelerating the production and delivery of safe, healthy, and cost-effective products to the public. Hence, when the FDA recognized a decline in biomedical or biologic product submissions, and biologic license applications (BLAs), the organization performed a 10-year (1993–2003) retrospective review of biomedical research funding (from the NIH budget and pharmaceutical companies research and development investment) and new molecular entities submissions and BLAs. It uncovered that the increased research expenditure did not match product development. This observation prompted the FDA to establish the Critical Path Initiative (CPI) to help reverse this trend. Additionally, the FDA developed guidelines for drug development tools (DDTs) intended for drug development programs. These guidelines provide a forum for interactions between drug developers and the FDA's Center for Drug Evaluation and Research (CDER), aimed at bridging the previous barrier between the FDA and product developers. This initiative serves to accelerate the process of regulatory clearance of products for safe clinical applications. Additionally, this collaborative effort ensures confidence in CDER reviewers that the developed DDT meets the specific context of use (COU) and hence has potential utility.

Another challenge in product success is efficient ethical marketing. Product acceptance by the public is dependent on a number of factors. Product quality (safety and effectiveness) and approval by the FDA ensures public trust and comforts insurance companies and healthcare providers. It behooves every product developer to therefore heed the call by the FDA in collaborative product development.

10.2 Biomarker Development Challenges

Established are the enormous challenges that biomarker developers encounter. These difficulties translate into the virtual lack of clinical translation of the numerous biomarkers being discovered. In addition to the concerns discussed in Sect. 3.3.4, biomarker development also faces other issues such as tumor heterogeneity, applicable biomarker clinical utility, logistics, regulatory oversight, commercialization process, and eventual adoption (Table 10.1). Thus, the development and approval of biomarkers for clinical applications encounters comparable hurdles to

Table 10.1 Challenges in biomarker development

Variable	Challenging issues
Pre-analytical	Study design Biospecimens Patient variables – physiologic/pathologic
Analytical	Methods Techniques and platforms Statistical and bioinformatics Cutoff values
Biologic	Tumor heterogeneity Clonal evolution Acquisition of novel genome changes Predictive unpredictability
Clinical	Patient homogeneity for trials Attrition rate Predictive problems – combined agents from different companies
Logistics	Needed resources for biospecimen management Optimal facilities for study execution
Adoption	Failure to meet regulatory requirements Failure to meet claimed performance when on the market Slow or lack of adoption Reimbursement issues

those associated with the developmental process of a new drug. Here are some challenges and pertinent issues worthy of note in the biomarker development process:

- Biomarkers may be easy to develop if they give a positive measurable signal. Even then, very few such biomarkers demonstrate uniquely elevated levels in any particular cancer type. Unfortunately, many cancer-associated gene expressional changes involve molecular pathways also used for normal cell functions, such as proliferation, growth, differentiation, apoptosis, survival, and immune functions. Except for a few biomarkers such as *PCA3* and *TMPSS2-ERG*, finding a cancer-specific biomarker is almost impossible.
- Even when a biomarker appears to be associated with a particular cancer, it tends to be deregulated in several other different types of cancer, making its association with cancer from any specific tissue type unlikely. In many of these scenarios, the biomarkers often lack screening or diagnostic utility and are often reserved for other clinical uses following cancer diagnosis. These are issues faced by several traditional cancer biomarkers in clinical practice (e.g., CA125).
- Expressed transcripts do not necessarily correlate with protein levels, which are the preferred disease biomarkers. Posttranscriptional and posttranslational modifications account in part for this failure in correlation.
- A screening or diagnostic biomarker is considered relevant when therapy for the condition is already in place, and/or its role is established in the biologic process of the cancer. The functional role of the biomarker could then help with the

understanding of tumor biologic mechanisms with possible development of targeted therapies.

- Secreted or cell surface-bound biomarkers are more adept to clinical translation. Thus, biomarkers that are not released into body fluids or attached to cell surface may have limited applications and hence less attractive for developers.
- Genome-wide association studies (GWAS) have been very informative in uncovering risk alleles for diseases. However, the predictions of increased cancer risk based on such biomarkers are often not too relevant to the general population. SNPs in familial cancer syndromes such as *VHL*, *TP53*, and *BRCA* have proven very useful, but the associated cancers afflict only a small percentage of the population.
- Genomic alterations are emerging as useful players in cancer biology and hence of biomarker potential. However, even cancers of the same histology demonstrate heterogeneous genomic changes. For cancers whereby a single gene has been identified as relevant, very few show mutation "hotspots." Mostly, these mutations are scattered across the gene, which necessitates the development of cost-effective technologies for mutation scanning. Moreover, no single gene is altered in all (100 %) of any cancer type, suggesting the need to develop panel assays to cover the vast majority.
- To overcome the issues of overdiagnosis and overtreatment and hence reduce morbidity and healthcare costs, an effective screening biomarker should have prognostic predictions as well. Being able to differentiate at the very beginning between indolent and aggressive cancers will be the "Holy Grail" cancer biomarker. But, because of the intricate and adaptive pathways employed for malignant progression, it is almost an utopian endeavor finding such a biomarker. Tumor progression may also depend on cancer-cell interaction with the surrounding stroma, including immune cells. While heavily marketed, such issues are identified for the PSA test. A European randomized study of screening for prostate cancer, for instance, found that 48 PSA screened detection of men have to undergo surgery to save one man's life 9 years later [1].
- The presence of cancer stem cells poses a unique challenge to biomarker development. Many biomarkers are developed for the bulk of the cancer tissue or cells. Any signal from the negligible cancer stem cell niche is overshadowed. Thus, biomarkers that miss the stem cell population, which are important for cancer metastasis, cannot therefore accurately predict disease course.
- Many tumors grow slowly such that it takes several years from initiation to metastasis. High-throughput sequencing data suggest it may take 15–20 years for pancreatic cancer to initiate and metastasize [2]. Moreover, at the time of metastasis, cancers tend to be large, and such high tumor burden is often associated with worse outcome. To reduce mortalities, cancer must be diagnosed early to enable curative interventions. Circulating biomarkers of early cancer are often diluted, such that sensitive and accurate technologies are needed for clinical application.
- The lack of high-grade reagents also hampers the development of some biomarker technologies. For example, some immunoassays, such as been adopted in

microfluidics and nanodetection technologies, require very pure, well-characterized, and highly specific monoclonal antibodies, which are often lacking.

- Nucleic acid detection technologies are becoming common. However, many such assays (e.g., PCA3, Oncotype DX) require special sample handling, which limits their use to a few specialized facilities. Transferring these technologies onto a lab-on-a chip platform for point-of-care utility will enhance their global presence.
- Developing a biomarker test for clinical translation requires high performance characteristics in assay accuracy and reproducibility, to satisfy the requirements for clinical laboratory improvement amendment (CLIA) regulatory approval. The stringency required by the FDA is even much higher. Companies need to demonstrate assay safety, effectiveness, and performance that affect clinical decision-making in the realm of evidence-based medicine. Obtaining these metrics require large multicenter randomized clinical trials. Designing, conducting, and funding such studies (which often fail to demonstrate the desired high accuracy) are a formidable and costly undertaking.
- A predictive biomarker may fail as a valid biomarker due to flawed clinically design. For example, evaluating the predictive performance of the biomarker in a setting of combined use of two agents from different manufacturers may fail to reveal the biomarker's authentic ability to be predictive for one of the agents.
- Flawed biomarker discovery studies and prototype development affect subsequent validation efforts. The many possible factors in this situation include:
  - The use of inappropriate samples (e.g., using advanced tumor stage samples to discover early-stage biomarkers).
  - Not defining the clinical utility question prior to conducting studies.
  - Poor statistical power or analysis and often the use of nonstandard in-house proprietary algorithms that later fail simple statistical scrutiny.
  - Developing prognostic biomarkers with samples from patients who have already received some form of treatment.
  - Establishing prognostic biomarkers by comparing people with treatment failures or death to those without relapse years after therapy.
- Common confounding issues that need early consideration in biomarker development, but often neglected, are who is the intended target population? When should testing begin or be performed? How often should the test be performed? Does the test have the same performance metrics across different ethnicities and geographic areas? How should the test result be interpreted and used? How are the end users (clinicians and patients) going to be educated on the clinical utility of the test? Clinician education on the assay performance and meaning is important in ensuring accurate test interpretation. How is the test result to be communicated to patients, and how should patients subsequently be managed?
- Development of valid biomarkers requires teamwork. At the beginning, a multidisciplinary team needs to be assembled. At its minimum, this should comprise of patients, clinicians, epidemiologists, biomarker experts, engineers,

industry partners, possibly academia, and regulatory agencies such as members of the FDA biomarker qualification program.

10.3 Biomarker Method Development and Evaluation

Biomarkers are used in two broad applications: (1) as laboratory assays to help in clinical decision-making such as disease diagnosis and prognosis and (2) in drug co-development. The regulatory requirements for these two applications are identical but controlled by two bodies. The former utility falls under the requirements of the Clinical and Laboratory Standards Institute (CLSI) guidelines for CLIA certification. For use in co-development of a drug, the biomarker must meet the standards of the FDA. To help address the biomarker validation concerns, and the stringency required for safe applications in drug development, the American Association of Pharmaceutical Sciences (AAPS) and the Clinical Ligand Assay Society (CLAS) convened a joint meeting in 2003 to stratify and make recommendations for biomarker method validation. Additionally, the FDA developed guidelines for qualification process for drug development tools, as well as the biomarker qualification program, all aimed at helping sponsors, industry, academia, and other stakeholders to accelerate the pace of safe and effective drug development.

It should be noted that biomarker analytical method validation is different from biomarker qualification. The analytical method validation is the process of assessing the performance characteristics of a given biomarker assay. This requires stringent bioanalytical method development, taking into account all performance metrics. Clinical qualification of a biomarker is the evidence and statistical process that connect the biologic, pathologic, and clinical endpoints of the biomarker to drug effect.

10.3.1 Fit-for-Purpose Approach to Biomarker Method Validation

Biomarkers are integral part of the drug development process, from discovery, prototype, to product development for commercialization. Depending on the type of biomarker (prognostic, predictive, or pharmacodynamic), the intended or defined utility may be different. Biomarker utility may also change from the initial discovery phase to the more advanced validation and qualification stages. Thus, biomarkers may be applied early for selection of potential drug candidates for subsequent development. Such utility may be more economical than some model systems in regard to resources involved. Given the diverse uses of biomarkers in the drug development process, Lee et al. proposed the fit-for-purpose approach to biomarker methods development and validation, when used in co-development of

a drug [3]. This strategy of biomarker method validation relaxes the stringency at the initial exploratory phases of drug development and increases the rigor of method validation as the drug progresses to advanced stages and toward eventual approval. One advantage of this approach is economics, because it enables resource conservation at the exploratory stages of biomarker methods development. The intended use or purpose of the biomarker method validation data determines the stringency or rigor of the bioanalytical methods used for data generation, and this increases as critical decisions on drug development are being made.

The nature of the bioanalytical method of the assay used for validating the biomarker also needs consideration. Lee et al. recognized four categories of the assay methods, namely, definitive quantitative, relative quantitative, quasi-quantitative, and qualitative.

- Definitive quantitative assay provides absolute quantitative values using physicochemical or biochemical methods such as mass spectrometry.
- Relative quantitative assay relies on the use of response-concentration calibration function, such as in immunoassays (ELISA).
- Quasi-quantitative assay does not use calibration standards, but rather a continuous response, whereby assay results depend on a characteristic of the test sample. Examples are antibody titers and qRT-PCR.
- In qualitative assay, the data generated does not reflect on the mount in the analyte. Data output can be nominal (yes or no) or ordinal using discrete scoring system as in immunohistochemistry.

The fit-for purpose approach also identifies four interrelated levels of assay validation, namely, pre-analytical considerations, exploratory method validation, advanced method validation, and in-study validation with sample analysis acceptance criteria. Detailed recommendations are provided for consideration in the execution of each level of the biomarker methods development [3].

- Pre-analytical considerations: This is a planning stage whereby the study design in regard to the intended use of the biomarker must be considered. A study plan should be crafted to include defined objectives, intended target population, sample types (body fluids and tissues), sample collection, and bioanalytical methods to use for initial assay design.
- Exploratory method validation: This is the stage of initial quantitative method development and characterization of its basic performance in regard to a number of parameters including accuracy, precision, analytical sensitivity, biomarker normal levels (baseline range in healthy individuals), biomarker stability, assay dynamic range, as well as statistical consideration for advanced validation for its intended purpose.
- Advanced method validation: This level is an extension of the exploratory validation stage with increased levels of biomarker characterization. The analytical rigor is increased and tailored to decisions on intended purpose, while considering important issues of safety, efficacy, pharmacodynamics, and

surrogacy. The formal acceptance performance of the biomarker for its intended use in drug development is established.
- In-study validation with sample analysis acceptance criteria: While in use for its intended purpose in drug development, the biomarker should continue to perform optimally as established. This in-study validation therefore ensures continuous robust assay performance as expected or defined for each study run. These validation runs also help with the establishment of definite acceptance criteria for the assay throughout the drug development process.

10.4 Biomarkers and Surrogate Endpoints

In clinical trials, an outcome measure is needed for prediction of drug effect on the body in regard to safety and effectiveness. The clinical endpoint, which is a characteristic that reflects on how a patient feels and functions or how a disease progresses in terms of survival, death, or relapse, is the ideal measure of the effect of an intervention (e.g., drug treatment). However, relying on clinical endpoint such as death, disease relapse, or the appearance of a sign or symptom in a clinical trial may be unethical or less effective approach to the drug development process. Therefore a biomarker can be used as a substitute or stand-in for the expected clinical endpoint, and such biomarkers are referred to as surrogate endpoints or surrogate markers. Biomarkers used as surrogate endpoints undergo very stringent method validation and qualification processes and are usually well characterized and evaluated to possess the desired clinical relevance and as substitutes. Even then, such valid surrogate endpoints can be misleading or poor predictors of clinical outcomes and hence can adversely affect the drug development process. In view of these shortcomings, the FDA permits provisional approvals that are dependent on surrogate endpoint predictions of safety and effectiveness. However, this acceptance usually requires a follow-up phase IV study or data collection to show a correlation between the actual clinical and surrogate endpoints.

10.5 FDA and Biomarker Development

The FDA is a federal agency with a number of mandates including:

- To promote and protect public health. This mission includes ensuring the delivery of safe, secure, and effective treatments to the public.
- To be responsible for enhancing public health. This is achieved through leadership in ensuring the rapid development of innovations to accelerate the process of making medicines safe, effective, and affordable to the public.
- To help provide information on accurate and scientifically sound evidence on available medications to the public.

Over the past couple of decades, the FDA recognized a problematic trend concerning the development of prototypes, devices, and biologics and thus provides solutions to mitigate them. Analysis of a decade trend (1993–2003) revealed that while the spending on biomedical research has been on the rise, submissions of drugs and biologic products to the FDA had been on the decline. In 2004, the Critical Path Initiative (CPI) report was issued to address this eminent crisis, which not only could stall innovation but also could contribute to increase healthcare cost. The report centers on three important issues needed to accelerate basic science discoveries to market:

- It aimed at identifying the problem in terms of what contributed to this observed pattern, so as to offer actionable solutions.
- Provides some efforts made by the FDA to improve the critical path and offers opportunities for future progress in product development for approval.
- Importantly, it calls for collaborative efforts among academia, industry, and the agency (FDA) to scope out the problems and hence offer effective solution in terms of how to create the needed path forward for success.

10.6 The Problem with Successful Biomarker Development

The considerable funding of basic science research, including the expansive biomedical discoveries after sequencing of the human genome, has resulted in several discoveries of biologics (biomarkers) that putatively should translate into better disease prevention, early diagnosis, treatments, and possible cures for many serious afflictions, including cancer. The progress from "bench-to-bedside" has unfortunately been lagging behind this rapid discovery phase. Translating the biomedical discoveries into effective safe and affordable medical products for public use has been dismal. Continuing on this trajectory without actions will lead eventually to stagnation or decline in biomedical innovations with associated increases in healthcare cost. Compounding this problem will be the proclamation by many companies of having developed valid biomarkers for translation without any outside oversight. Overseeing all these issues will further burden the efforts of the FDA.

The FDA recognized the disconnection between basic science discoveries and the applied sciences needed for medical product development. The focus of research has heavily been on developing cutting-edge technologies for discovery research. Thus, the accelerating technologies for discoveries outpaced those required for product development.

A major issue with biomarker or device development is performance, which is a measure of product safety and effectiveness. Thus, to accelerate biomedical product development for commercialization, the FDA recognized the need for new tools to cost-effectively demonstrate product safety and effectiveness. The challenging issues with biomarker validation at the clinical trial phases are obvious. Many

prototype and biologics fail to reach the desired effectiveness of safety standards for public use. The FDA has therefore identified the need to develop "new product development toolkits," with the aim of accelerating efficient development of safe and effective medical products. The critical path to new medical product is a result of this effort.

10.7 The Critical Path Initiative

Considering all the identified issues, and the eminent catastrophe with inaction, the FDA developed the CPI to help product developers accelerate the process of getting safe and effective drugs, devices, and medical biologics to market. As soon as developers have identified prototype or discovery product, they are encouraged to consider entering into the program. Once they enter into the critical path, FDA experts, developers, and other stakeholders ensure a rigorous evaluation process toward product approval.

The disconnection between basic science discovery and commercialization has been recognized globally by a number of organizations as well, and efforts have been made to help bridge this gap. These endeavors aim at accelerating the clinical translation of biomarkers, including the NIH *roadmap*, the NCI SPORE program, the EORTC, the NIST biomarker program, and the National Translational Cancer Research Network from the British government.

However, by definition, translational research extends from prototype, device, or discovery stage through preclinical to clinical development. The gap from this stage to regulatory approval becomes murky at this stage. How these products pass the stringent approval process of safety and effectiveness is not addressed. To this end, the FDA recognized the need for "bridged" research efforts, which is referred to as *critical path research* (Fig. 10.1). This targeted or focused research is aimed at providing the critically needed path from translational research to successfully approved product development. A challenge for the critical research mandate is the high failure rate of getting discovery research through product development. It is recognized that a newly discovered medical biologic entering phase I studies has just 8 % chance of success, partly because of the failure in identifying candidate medicinal compounds or biologics likely to succeed. That is, from the discovery phase, being able to identify which biomarker, compound, or device has the highest probability of success is critical to downstream efforts at further development. In view of these difficulties, the goal of critical path research is to develop *tools* to guide efficient product development. These tools include standards, assays, biomarkers, clinical trial endpoints, and computer-modeling techniques aimed at identifying promising agents early in the development process, so as not to pursue products that are likely to fail. Obviously this approach should save thousands of dollars in pharmaceutical development processes.

The CPI report identifies three not mutually exclusive dimensions in development of biologics, drugs, and devices. These are product safety establishment,

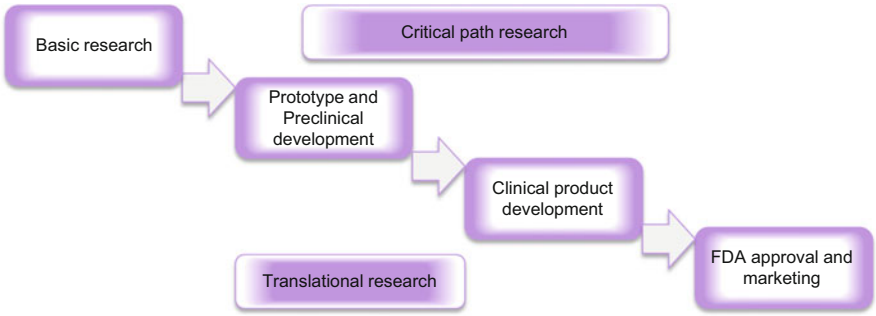

Fig. 10.1 The place for critical path research in the biomarker product development (adopted from the CPI report of the FDA)

demonstrating medical utility, and the industrialization process. These three critical dimensions must meet desired quality at all stages of product development.

- Is the product safe enough for human testing and eventual clinical distribution?
- Will the product be beneficial to the public? This requires demonstrating clinical effectiveness.
- How do we get the product from bench to market? This process involves developing a high-quality product for mass commercial production. This process may involve product characterization and specification, physical designs, manufacturing scale-up, and quality control for patient safety.

There are specific tools for each of these dimensions. These are tools for assessing safety, demonstrating medical utility and characterization and manufacturing.

10.7.1 Product Development Tools

In order to accelerate development of products, tools are needed to ensure product safety, effectiveness, characterization, and manufacturing throughout the process. While the FDA is providing leadership in this direction, opportunities exist for industry, academia, and other interested parties to help identify and develop some of these tools.

10.7.1.1 Safety Assessment Tools

Medical product safety for public use is a major mandate of the FDA. Therefore, part of the CPR is devoted to developing tools for safety product assessment. To reduce cost to industry, safety concerns should be identified and addressed early in the development of a biologic. The paucity of safety assessment methods can result

in advancing products to late stages before recognition of how unsafe they may be to the public or even result in product withdrawal after commercialization. Such issues have occurred in regard to a number of drugs. Methods for early detection of significant safety issues should be a major focus of product developers.

While any product may have safety concerns, much attention in this area has been on drug development. In regard to biomarkers for clinical applications, safety issues will relate mainly to product performance accuracy, such that the harms to consumers are minimized. Tools that can be developed to help achieve this will include developing standard operating procedures with efficient protocols for biomarker qualification programs.

The FDA's identified areas for safety assessment method development include product testing for possible contamination, toxicology testing, and human exposure. Animal models have commonly been used for drug toxicity testing; however, they can be labor intensive, costly, and yet unreliable. More reliable methods are therefore needed, and the FDA's leadership in this area is commendable and noteworthy:

- One example of successful safety too development is the recommendations by the FDA for the use of in vitro human cell lines to characterize drug metabolism. This leadership has helped with reductions in the issues of drug safety after development due to drug-drug interactions.
- Another success story from the FDA is the development and standardization of methods for documenting the clearance of retrovirus-like particles from tissue culture media. This feat has accelerated the development of safe monoclonal antibodies, some of which are currently been advanced rapidly as targeted therapies.

The FDA has identified several opportunities for the expansion of the safety toolkit, which needs to be reviewed by industry, academia, and other researchers for development.

10.7.1.2 Medical Utility Tools

To be of benefit to patients and the healthcare system, medical products must be effective at what they claim to accomplish. This area of method development has been challenging partly because of the lack of predictive value of animal models, and yet sources of variations during human trials are unpredictable. Hence, the FDA has earmarked a number of opportune areas to be explored. Additionally, the FDA has made some efforts at enhancing the ability for drug developers to demonstrate medical utility in a timely fashion. For example:

- Adopting CD4+ T-cell counts and measurement of viral loads as surrogate biomarkers of the effectiveness of anti-HIV drugs accelerated the drug development and approval process.
- Adopting the eradication of *H. pylori* as surrogate for effective treatment for duodenal ulcer disease enhanced drug development.

- Finally, by accepting the use of validated surrogates as evidence for vaccine effectiveness in establishing desired immunity has enhanced the vaccine approval process.

10.7.1.3 Characterization and Manufacturing Tools

Another issue with product development is the apparent dissociation of discovery efforts from the product manufacturing process. New technologies are surging for discovery research and possible prototype developments. These techniques tend to focus very much on analytical sensitivities, which are of importance in early disease biomarker discovery. The obvious issue is the often little attention given to the transition of such technologies to market. Thus, product failures may occur because of problems with the transition to industrialized production. The product development process can be expedited if technical standards are developed, especially when new technologies are used for initial research and development. Additionally, developers should consider what is needed for physical design, characterization, manufacturing, and scale-up in order to successfully bring products to market.

10.8 Biomarkers in the Drug Development Process

Biomarkers fall into two utility categories: (1) for clinical applications such as screening and (2) for drug development. These are not mutually exclusive applications of any specific biomarker. The very tenet that biomarkers are entities or factors that are objectively measured and, which modulate in physiologic states, with disease features or processes as well as treatment responses pivots their utility in the drug development process. For drug development purposes, biomarkers are classified simplistically as diagnostic, prognostic, predictive, and pharmacodynamic biomarkers.

- Diagnostic biomarkers are used to stratify people into diseased and healthy categories in regard to the physiologic or pathophysiologic characteristics of the disease.
- Prognostic biomarkers predict the natural course of a disease prior to any treatment interventions. They identify an individual's risk level of disease progression.
- Predictive biomarkers forecast treatment outcomes in relation to the likelihood of response to a specific therapeutic agent.
- Pharmacodynamic biomarkers measure therapeutic effects. They change to indicate that a biologic response to an agent being received by a patient has occurred.

These definitions do not mutually exclude any specific biomarker. A single biomarker may possess any combinations of the four designated utilities.

10.8.1 Utility of Biomarkers in Drug Development Programs

On the thesis of the above classification, biomarkers can be applied in a number of ways in the drug development process.

- Pharmacodynamic biomarkers may be used to monitor for early detection of drug toxicity before clinical evidence of a disease state. They are also useful in treatment regimen and dose establishment, usually in a phase II trial to plan for phase III studies.
- Biomarkers may be used for patient selection during a clinical trial enrolment. Predictive biomarkers may identify a group of patients likely to respond to a drug in a predictable manner. Such treatment outcome predictions can be very important in successful drug development. For example, the EGFR inhibitor, gefitinib (Iressa), failed as a lung cancer-targeted therapy in unselected lung cancer patients. However, by targeting only the ~5 % of adenocarcinoma subtype of NSCLC with EGFR mutations, the dramatic effectiveness of the drug was proven.
- Biomarkers may be used for patient stratification for randomized clinical trials. For example, a prognostic biomarker may identify a group of patients more likely to develop progressive metastatic disease for a specific drug development program. Being able to demonstrate a reduction in metastatic events can be evidentiary of drug effectiveness.
- Biomarkers can serve as surrogate endpoints in the drug development process. A stringent qualification is mandatory for such biomarkers to prevent failures to demonstrate desired benefits or outcomes at large clinical trials.
- Some biomarkers can identify a subgroup of patients with specific diseases for which there currently is no effective therapy available. These patients can be enrolled for new drug development programs.

10.8.2 Regulatory Requirements for Biomarker Use in Drug Development

Drug developers may use an existing qualified biomarker for drug development. Under such circumstances, there are two possible regulatory choices or paths of methods or assays to employ:

- Developers can use within their investigational new drug application (IND), new drug application (NDA), or biologic license application (BLA) the assays or methods that were reviewed and used for biomarker qualification. If within their IND/NDA/BLA, the proposed methodology to be used in the drug development program is the same as was used for biomarker qualification, it can be used without any further regulatory submissions.

- An alternate method or assay may also be used. Should this path be opted for, a submission of information to support equivalency of the new methods or assays to the one used for biomarker qualification has to be made to the review division via submission to the particular IND/NDA/BLA. A performance comparison of appropriate samples using the new assay can be carried out with the assay used for qualification of the biomarker. If successful, further required regulatory evidence on the new assay/methods will not be required.

An imaging biomarker may be chosen for drug development. This understandably may require the injection of an imaging drug. In this situation, the marketing regulatory requirements for the use of the imaging drug and its use as an imaging biomarker are distinct. An IND for the imaging drug may be required prior to use. However, if the FDA has already approved the imaging drug, and the biomarker utility falls within this approval, it can be used without the need for an IND.

10.8.3 Use of Diagnostic Devices for Qualified Biomarkers

Biomarkers used in drug development may already be cleared for clinical practice. The vast majority of biomarkers require measurements using diagnostic devices. Thus, the review or evaluation of data submissions for the drug under development also considers the analytical performances of the device used for biomarker measurement. Fortunately, many of these devices, especially those used in the clinics, will have been reviewed and cleared for marketing by the CDRH/CBER branch of the FDA. The qualification program for drug development is however not dependent on any specific diagnostic device. Thus, if many different devices were used to measure the biomarker, they should yield the same results. This should provide a level of comfort in terms of rigor in biomarker performance.

From a regulatory standpoint, the drug development qualification program is separate from the device review and market approval program. Therefore, an FDA-cleared device does not necessarily imply that what it has been used to measure constitutes its qualification for use in the drug development program. Conversely, the eventual qualification of a drug development tool does not imply that the device used is cleared for commercial use. It is imperative that the use of an FDA-cleared device has some advantages; however, the use of multiple devices to measure the biomarker, at least initially, may offer much comfort.

10.9 Biomarkers and Qualification Process for Drug Development Tools

The FDA defines drug development tools (DDT) as "methods, materials, or measures that aid drug development." The defined or identified DDTs by the FDA are not the only paths or elements needed for valid drug development, but are useful recommendations for any developer. Thus, the DDT qualification program is not required by a drug developer, but is useful and obviously accelerates the approval process because it is a collaborative effort that meets the requirements of both parties. It is a process that instills comfort in DDT developers as they get to understand the process by which submitted data will be reviewed by the FDA-CDER for specific context of use (COU).

10.9.1 Qualification

Qualification is when the DDT can be relied on to provide specific interpretation and application in drug development and regulatory review as stated in the COU. Thus, the COU guides the utility and purpose of the DDT and fully provides the circumstances under which the DDT is qualified. This path ensures that when the DDT is qualified for a specific COU in drug development, it will guide the production of analytically valid measurements reliable to have specific use and meaning. This is a critical path useful for drug development because in the specific qualified context, the DDT can be used for IND, NDA, and BLA submissions. The review of such submissions is relatively easy on the CDER reviewers because of their participation and hence prior knowledge of the DDT. Here are some advantages of DT qualification program:

- Ease of IND, NDA, and BLA review, thus accelerating the drug development process. This benefits the developer, patients, and the healthcare system.
- The qualified DDT can be used for additional studies to generate new data for new DDT submission to support expanding the original qualified COU. The new DDT can hence be used in IND programs for new COU.
- The presence of qualified DDT also helps the FDA. For example, an unqualified DDT can be used by multiple sponsors/developers, or a particular DDT can be used by one sponsor/developer in multiple different clinical setting. In these scenarios for the utility of such a DDT, the FDA-CDER staff need to review each DDT separately to justify its use on a case-by-case basis. However, with qualified DDT, review of data has to be conducted only once.

The FDA qualification programs are based on three types of DDTs: biomarkers, clinical outcome assessments, and animal models.

10.9.1.1 Biomarkers

Biomarkers are important in the qualification process of DDTs. Because of the objectivity in their measurements, they can accurately identify changes in people in an unbiased fashion. Valid biomarkers of a disease process that are measurable prior to treatment interventions may serve as useful tools for patient selection for a clinical trial. Biomarkers for treatment monitoring modulate with or change following treatment. These changes may be indicative of pharmacologic response or raise an alarm in regard to possible toxicity. The directions or levels of biomarker change may be useful in assessing pharmacokinetics and pharmacodynamics of the drug and help in establishing efficacious dose without unwanted effects.

Because of tumor heterogeneity, valid biomarkers for DDTs utility may not be a single, but multiple entities. Multimarker or panel biomarkers (referred to by the FDA as "composite biomarkers") are usually combined using some form of statistical algorithm to make a defined call.

10.9.1.2 Clinical Outcome Assessment

Clinical outcome assessment (COA) is an important component of the DDT qualification process. It is the measure of patient outcome in regard to symptoms, mental status, or effect of ill health on how the individual functions. This evaluation predicts drug performance in terms of benefits or harms.

Because COA needs to be an objective measure, and yet is a complex process, they are reported as a scoring system in conjunction with methods and instructions used in administering and assessing response. COA of treatment outcome may be from direct or indirect evidence. In the case of the latter, the evidence has to address how adequately the COA corresponds to how patients feel or function. The reported clinical outcome can be from three possible sources: directly by the patients, or indirectly by a physician, or another third-party observer. The patient-reported outcome (PRO) is a direct report on patients symptoms or functioning without any interference such as interpretations or otherwise by a clinician or anybody. The clinician-reported outcome (ClinRO) is an assessment (e.g., motor, sensory, or cognitive performance) based on clinical observations and interpretations by a qualified clinician. An observer-reported outcome (OBsRO) is an assessment made by non-clinician observer.

10.9.1.3 Animal Models

Animal models continue to be an integral part of biomarker research and development. For the purposes of DDTs, an animal model is considered by the FDA as the combination of the appropriate animal species, the agent used to establish disease in the animal (the challenge agent, referred to as etiologic agent in humans), and the

route of exposure to recapitulate the human disease process or pathology in the animal.

The animal model qualification (AMQ) program by the FDA refers to appropriate model animals used in efficacy studies to provide ample evidence of effectiveness for the drug. The developer can establish different animal models and use each for efficacy testing of multiple investigational drugs for the same-targeted disease. To be accepted in the qualification process, animal models must meet three important requirements to provide comfort that the efficacy or otherwise outcomes in the animal model closely represent expectations in humans. These three FDA requirements are:

- The animal model disease process or condition should correspond in multiple important aspects to the disease or condition in humans.
- The pathologic or toxic mechanisms should be similar in the model animals and humans.
- The challenge agent and etiologic agent should be the same.

These requirements ensure that the disease etiology, mechanisms, and processes in the animal model faithfully recapitulate the human condition.

10.9.2 The Qualification Process

The FDA has defined a streamlined process for the DDT program. The DDT qualification process is split into three stages: (1) initiation, (2) consultation and advice, and (3) full qualification package review. The center for drug evaluation and research (CDER) staff is supportive and works proactively to help DDT submitters through the development process.

Stage I or the initiation stage begins the DDT qualification process. After a developer has identified and defined a sound DDT concept, a request is made to the FDA-CDER for a DDT tracking number. This request initiates the generation of a tracking number, which is provided to the developer and also entered into the electronic database of the FDA. The next step is the writing and submission of a letter of intent (LOI) to the FDA. This letter should include consultation request with the FDA-CDER to discuss potential value of the DDT, a description of the DDT, the proposed context of use (COU), and planned activities to be conducted to support qualification of the DDT. Once received, the FDA-CDER reviews the LOI and decides whether or not to proceed with the request. This decision is based primarily on the proposed scientific merit of the DDT and its COU, as well as availability of resources to undertake the review and the process. Should the request be unacceptable, the developer may be required to provide additional information or otherwise explanations as to why the request was declined is provided, together with advice on other paths for the DDT development process.

Stage II or consultation and advice stage is entered into when the review of an LOI is successful. At this stage the developer agrees to have the qualified DDT

made public on the FDA's guidance web page under drug development tools. The primary purpose of stage II is coordinated expert advice to the developer to help with the acquisition of the relevant data for submission of a full qualification package (FQP) for internal review.

A team of experts including CDER staff, relevant centers, and disciplines form a qualifications review team (QRT). This team then requests the developer to submit an initial briefing package (IBP) to be reviewed at their initial meeting. The IBP includes biomarker qualification, clinical outcome assessment (COA) qualification, proposed plans for COA qualification, summaries of completed or planned studies, conduct of study to ensure data quality and integrity, limitations of the qualification, and any questions or clarifications from the FDA-CDER.

A meeting is held to review the IBP in detail such that appropriate advice can be provided to the DDT developer on needed evidence for the qualification. The developer then works to generate and gather any additional data required, in consultation with the QRT. Periodic meetings between both teams are conducted to review data, identify gaps, alignment of studies and DDT with COU, and any expert advice necessary to generate data for the qualification DDT review submission.

Stage III or review stage is entered when both teams are satisfied with the qualification data accumulated. The developer now prepares and submits a full qualifications package (FQP) to the QRT for internal review. In addition to detailed descriptions of studies and analysis, primary data from studies conducted may be included. With the agreement by the developer, summary information from the package is posted on the FDA-DDT qualification web page.

Internal meetings by the QRT ensue to review the FQP so as to make eventual recommendations. During this process, the QRT interacts with the developer should any clarifications or request for more information becomes necessary. Although the developer may make a specific COU for the qualified DDT, the final decision of COU is made by the QRT based on submitted supporting data. Should a successful recommendation be made to qualify the DDT, the FDA will post summary of the recommendation on its web page. The qualified DDT may also be expanded or withdrawn in the future as additional data become available.

10.10 Evaluation of Biomarker Clinical Utility

Biomarkers for clinical utility or drug development must go through rigorous evaluation process to be considered valid for clinical applications. To be useful in the clinical setting, appropriate use of biomarkers is mandated. The clinical utility of a new biomarker can be evaluated by adopting the tumor marker utility grading system (TMUGS) proposed by Hayes DF and colleagues [4]. TMUGS helps clinicians and developers determine the efficiency of biomarker clinical utility. The rubrics of the TMUGS consist of two parts: (1) definitions and specifications

and (2) clinical utility of tumor markers. The TUMGS rubrics are summarized herein:

- Biomarker needs to be clearly identified in regard to its molecular form or otherwise: genes, transcripts, proteins, metabolites, cells, vesicles, or even a process.
- The specific alterations being detected or measured ought to be clearly noted, e.g., gene amplification, mutations, and protein levels.
- The assay used to detect and/or measure the changes should be identified, e.g., RT-PCR, ELISA, and massively parallel sequencing.
- Identify and use specified reagents. Standardize reagent conditions and keep note of possible lot-to-lot variations.
- Specify how the signal is detected and biomarker quantified. Should there be a cutoff value for the biomarker, specify how this was established. The robustness of the statistics used to arrive at the concluded biomarker cutoff value should be clarified.
- Specify specimen type and source used for biomarker evaluation, as well as pre-analytical handling in regard to methods of collection, preparation, storage, and transportation, e.g., serum, EDTA, 4 °C, transported on ice. Any temporal delays, storage times, and freeze-thaw cycles should also be noted.
- Specify the disease for which the tumor marker is being evaluated for, e.g., colorectal cancer.

The clinical utility evaluation is very important as well. Determine the biomarker development process and the level of evidence for claims being made.

- Clinical applications of biomarkers should be determined, e.g., screening, prognostic, etc. Note that a single biomarker may have more than one utility – a prognostic biomarker may be predictive as well.
- The biologic relevance of the biomarker should be noted. A biomarker linked to a biologic process or pathway may have treatment reagents present that patients can benefit from. Alternatively, a biomarker with clear biologic attributes can inform discovery and development of therapeutic reagents targeting pathway components.
- To be useful, a biomarker clinical utility should help with favorable patient outcome in terms of survival benefits, reduced morbidity (increase quality of life), and decreased overall healthcare costs.

10.11 Biomarker Adoption into Clinical Practice

Another challenge often faced by biomarker developers is ensuring successful adoption of the new biomarker into routine clinical practice. A number of issues affect how slowly or quickly a particular biomarker gets accepted for making clinical decisions:

A valid biomarker is more likely to be adopted than one that makes unsubstantiated marketing claims on performance. There is a need to establish uniform standards for demonstrating biomarker validity before commercialization. To achieve this requires oversight from authorities such as the FDA, NIH, Centers for Medicare and Medicaid Services (CMS), and NIST. Collaborative development between these authorities, industry, and academia, as being offered by the FDA CPI, the biomarker qualification program as well as the qualification process for drug development tools, should enhance the development of biomarkers with clear clinical impact. There should be facile clinical adoption of such biomarkers.

To be successful, there ought to be reimbursements for the new biomarker. Healthcare payers demand evidence of biomarker value in regard to cost and clinical effectiveness in order to make reimbursement decisions. Because all such information may be lacking for a new biomarker (e.g., being able to provide cost-effectiveness data without marketing), slow adoption may require lower or appropriate pricing. Stakeholders and expert panels can determine these coverage and pricing issues. Controlled coverage may have to be adopted, with limited use in specified circumstances until such time that sufficient data has been accrued to demonstrate a value proposition for the biomarker.

An adopted test may not necessarily be able to sustain its initial claims once on the market. Post-market assessment is needed for the accuracy, quality, clinical and cost-effectiveness of the test, or biologic. Preferably, an independent unconflicted body should conduct this assessment.

10.12 Summary

- Many challenges are encountered during cancer biomarker development. These challenges include the inherent tumor biologic variables, pre-analytical and analytical variables, statistical and bioinformatics considerations, and stakeholder factors, all of which often lead to the generation of putative biomarkers that fail large validation studies and hence unqualified for regulatory approval.
- In view of these challenges, many discovered cancer biomarkers never proceed to product development.
- The post-human genome project era has been filled with excitement about novel genetic and genomic product developments for clinical oncology.
- On the contrary, the FDA recognized a rather decline in biologic submissions for review, despite increasing biomedical research funding.
- To help avert this trend, various organizations and institutions including the NIH and the FDA put into place actions and mechanisms to accelerate the development and commercialization of safe and effective biomedical products.
- One solution to the problem is rigorous biomarker analytical method development and validation.
- The FDA also produced guidance documents to help with collaborative efforts in this regard. The Critical Path Initiative, the biomarker qualification program, and

the qualification process for drug development tools are all aimed at assembling stakeholders and FDA scientists/staff to collaborate in accelerating product development.

References

1. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320–8.
2. Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;467:1114–7.
3. Lee JW, Devanarayan V, Barrett YC, et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharm Res. 2006;23:312–28.
4. Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst. 1996;88:1456–66.

Further Reading

Guidance for Industry and FDA Staff. Qualification process for drug development tools. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. January 2014. Procedural.

Innovation and Stagnation. Challenge and opportunity on the critical path to new medical products. US Department of Health and Human Services, Food and Drug Administration. March 2004

Chapter 11
Ethical Considerations in Biomarker Development

Key Topics

- Bioethics
- Research participants and the informed consent process
- The role of biobanks in cancer biomarker development
- Ethical issues in cancer screening and diagnosis
- Ethical issues in industry-academia collaborations

Key Points

- Biomarker product development is impossible without provision of precious samples linked to private information of cancer patients, their families, and other citizens. The collection, storage, and utilization of such samples and data raise critical ethical issues that all stakeholders in the biomarker development cycle must be knowledgeable about.
- Participant confidentiality and privacy must be protected. One aspect of ensuring this is through consenting. Crafting a consent letter to balance optimal participant protection and seamless biobank operation and research activities is almost impossible to achieve. The path forward is good governance, transparency, trust, and refraining from any form of manipulation, exploitation, or use of moral pressures.
- While industry-academia relationships are necessary for biomarker product development (e.g., clinical trials), there are financial conflicts of interest issues that may pose complex ethical challenges in regard to how these relationships are established and fostered.

G.D. Dakubo, *Cancer Biomarkers in Body Fluids*,
DOI 10.1007/978-3-319-01580-4_11

11.1 Introduction

There are ethical, legal, and social issues (ELSIs) connected with biomarker development, from discovery to its clinical application. The developmental process requires clinical samples from patients and volunteers, in addition to their phenotypes and clinical, demographic, and even biometric data. Collecting these samples and associated information raises privacy and confidentiality concerns, which are usually addressed through consenting. The development of biomarkers and therapeutic products always involve some collaborative relationships between industry and academic medical centers. This relationship is central to advancing biomedical product development and has been successful in bringing some novel cancer therapies to the clinic. However, there can be substantial financial conflict of interest (FCOI) issues, which may cast doubt on the scientific validity and judgment about the product.

There are biomarkers in use for cancer screening and diagnosis, and this utility is not without ethical problems. The issue of overdiagnosis, and potential overtreatment, is topical in molecular diagnosis. Knowing the causes of overdiagnosis enables the development of mechanisms to reduce their occurrence.

Finally, since the 1980s, there has been a surge in direct-to-consumer (DTC) advertising, especially in the United States and New Zealand. While DTC advertising is under regulatory control, and potentially empowers patients in making decisions about their care, DTC advertising appears to adversely affect healthcare delivery and the role of the physician as an unbiased care provider. In circumstances whereby the healthcare provider has FCOI in the advertised product, ethical dilemmas may emerge.

11.2 Bioethics

Ethical guidelines for care of the sick, for framing healthcare policies, and for the conduct of research involving humans have been established for over four decades. Importantly, ethical issues in all these domains involve everyone, including patients, doctors, healthcare team, administrators, scientists, politicians, advocacy groups, organized institutions, and societies. Bioethics is an evolving science that shapes the discussion as new discourses arise. The primary purpose of bioethics is to improve the standards of care and conduct of research. In its holistic form, the bioethical framework of Beauchamp and Childress, governed by the four principles of respect for autonomy, non-maleficence, beneficence, and justice, must be exercised at all times:

- Respect for autonomy involves acknowledging a person's right to hold views, make choices, and take actions based on their values and beliefs.
- Non-maleficence means abstinence from causing harm to others intentionally.

- Beneficence implies acting for the benefit of others. That involves providing benefits to people and contributing to their welfare.
- Justice means fair, equitable, and appropriate treatment of people, in light of what is due and owed to them. Also benefits and burdens should be distributed fairly.

In addition to these four principles, referred to as the Georgetown Paradigm, Beauchamp and Childress included four other rules to guide ethical practices of biomedical research involving humans. These are veracity (truth and objectivity), privacy, confidentiality, and fidelity. The conscious application of these principles and rules at all levels of patient care, or in a clinical research setting, will probably not eliminate but substantially reduce any possible cause of harm or exploitation of an individual or society. But while the principles need considerations in all cases, a balance or compensation may be necessary in special situations. For instance, the Georgetown Paradigm principles and rules appear to weigh disproportionately toward protecting the individual and less so in the interest of public health. While individual ethical issues must be valued, the need to enhance biomarker development for public health benefit must also be considered under certain situations.

11.3 The Cancer Patient and Research Participants

The entire human population is central to biomarker development. For cancer biomarker development, cancer patients, their families, patients with benign conditions, and healthy volunteers all contribute valuable samples for research and biomarker development. The biomarker development cycle (Fig. 11.1) involves several players, with sample donors or research participants central to its success and yet can be vulnerable in the process. Not only are these people not inconvenienced, but also the process of sample acquisition can be uncomfortable. Additionally, what happens to this precious sample, and the associated donor information, is unclear and can be of major concern because any shared genetic information may lead to discrimination and possible stagnation against the donor.

Ethically, therefore, the participant/donor has to consent to participate in the research. However, informed consent has its challenges, because often they are incomplete. For instance, many neglect to disclose the full nature of the research, such that the risks and benefits to participants and society are not explicit. How individual privacies are protected, especially with collection of "big genomic data and its storage in megadatabases," needs explicit understanding by participants. Expectedly, there are issues with informed consent that need some improvements.

Informed consent came to being in medical research as an ethical and legal regulatory requirement and was clearly emphasized in the 1960s with the establishment of bioethics. How the best-informed consent should look like in order to protect patients or research participants, citizens, researchers, and other stakeholders, while enabling seamless conduct of research, is still blurry. Authorities

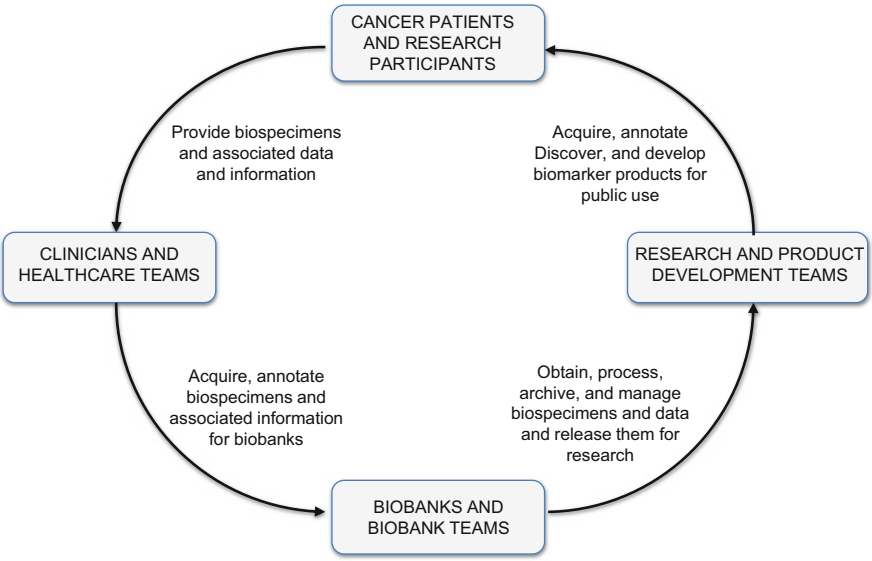

Fig. 11.1 The biomarker development cycle

have developed various models aimed at balancing this equation. However, informed consent can be broadly viewed as either offering optimal protection to the participant with several limitations to research (narrow consent model) or giving limited protection to participants while making it easy for biobanks and researchers to operate (broad consent model). Therefore, how the informed consent is structured places responsibilities more on the participant (narrow consent model) or the researchers and their institutional review boards (IRBs) and ethics committees (broad consent model). Because the narrow consent model requires specific and concrete information about the intended research, it is limited in its utility for different research using the same samples. Under circumstances whereby the same samples are needed for a different research project, the participant will have to be contacted again for a new consent. Issues raised have been logistics and problems with inability to trace participants (who may be dead).

There are issues with the type of informed consent participants/donors prefer. Central to this preference is "trust and lack of donor exploitation." A bigger issue with regard to consent in this genomic era is the establishment of biobanks (see Sect. 11.3). These biobanks collect and store large amounts of biological materials alongside with donor's information including clinical, demographic, phenotypic, lifestyle, behavioral, nutritional, exercise habits, and alcohol and tobacco use. All this private and confidential information about the donor requires tight privacy protection. Besides, the possible linkage of donor data to genetic information requires strict confidentiality guidance. The preference of the type of consent differs substantially depending on the use and sharing of material and information. In general, the opinion of donors on material use and data sharing is influenced by trust

and by the lack of exploitation by industry or any stakeholder. Obviously, donors will frown on data sharing without explicit consent. This is even worse if the consent form includes information about biobanks receiving funds from pharmaceutical companies. Under such conditions, donors are more likely to request for narrower consent, due simply to lack of trust on privacy and utility of data. On the contrary, when properly communicated, such that there is trust and satisfaction between the healthcare team and how biobanks operate, donors will be more open to accepting broad consents, thus making it easy for researchers to do their work. In this instance, donors trust that surrogate decision-making by IRBs and ethics committees will be in their best interest.

Informed consent, irrespective of the type, is often delivered under circumstances whereby participants fail to concentrate. These may be patients who are experiencing some form of depression from their illness, family members who are worried, or healthy donors who have little interest in the research and hence pay little attention to the content of the consent letter. A general issue therefore is a lack of recall on the part of participants. In several different analyses, it has been clear that participants lack knowledge on a wide variety of issues including:

- Not remembering to have signed a consent form.
- Not remembering to have donated biospecimens.
- Not remembering to have provided personal and medical information.
- Not remembering about issues with data and information usage and sharing.
- Not remembering the contents of the consent form.
- Not having any knowledge about biobanks.
- Not remembering anything about risks and benefits of the research.
- Not knowing that they have the choice to withdraw, if they so desire at any point in time during the research.

With the use of alternative strategies, this knowledge gap could be improved upon. Personal or small group discussions with the researcher or team members from biobanks on the purpose of the research, potential risks, and discomfort, as well as use of different forms of communication strategies such as audiovisual sessions, should help improve understanding and comfort in participants. In fact, adopting a participatory action research approach may help empower the public to have a positive attitude toward participation in biobanking. As perceptions of trust, risks, and potential benefits play important roles in participant decision-making, it is prudent to provide ample information and, possibly, different consent models. The risks of possible misuse of medical information should be explicit. Potential risk of losing information to employers or insurance companies, though very unlikely, which could be used for discriminatory purposes should be clear. Indeed, legislation on this issue (information sharing with employers and insurance companies) must also be explicit. To this end, also involving participants at some stages of specific research being conducted with their materials, by sharing grouped data, may offer some comfort as well.

In the era of genomic research and biobanking, there are numerous issues with how research should be conducted and importantly who owes the end results. In the

systematic review by Khan et al., the main bioethical issues that emerged in regard to genetic and genomic research were confidentiality of participants, documentation of informed consent, public attitudes, future use of participant samples and data, and disclosure of results [1].

11.4 Biobanks

The biobank industry is relatively new, dating back approximately three decades. In the era of genomic research, biomarker development, and personalized medicine, biobanks hold tremendous value in ascertaining the easy acquisition of high-quality and well-annotated biospecimens for research. An important component of this endeavor is reductions in pre-analytical variables in sample handling. Even then, establishing and managing a biobank is a complex and formidable process. But what exactly are biobanks?

Being number 8 of the 10 Ideas Changing the World Right Now published by *Time Magazine* in 2009, biobanks were simply referred to as safe houses for tissue samples, tumor cells, DNA, and blood for research into new treatments for diseases. However, the evolving roles of biobanks and their teams are much broader than this definition. National biobanks are sprouting in many countries including Britain, Canada, Norway, Sweden, and the United States. Thus, a better definition is needed and has been provided by various authorities, societies, institutions, and organizations. While the wordings may differ, they all convey the same central idea of what a biobank should look like and how it should operate.

- A pioneer and major player in the biobank industry, the pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) defines biobanks as *containing biological samples* and *associated information* that are *essential raw materials* for the advancement of biotechnology, human health, and *research and development* in life science.
- In 2008, Kauffmann and Cambon-Thomsen simply referred to biobanks as *organized collection* of human *biological material* and *associated information* stored for one or more *research purposes* [2].
- The European Commission referred to biobanks as *organized collections* consisting of *biological samples* and *associated data* of great significance for *research and personalized medicine* [3]. This definition was later expanded upon to include not just samples and data for research, but aspects of participant privacy and confidentiality, as well as governance of biobanks.
- The International Society for Biological and Environmental Repositories (ISBER) regards biobanks as *entities that receive, store, process, and/or disseminate specimen* as needed. This society also recognizes the physical location and full range of activities associated with the operation of a biobank.
- The Organisation for Economic Co-operation and Development (OECD) sees a biobank as a collection of *biological material* and *associated data and*

information stored in an *organized system*, for a population or large subset of a population. The OECD also provides guidance for the establishment, governance, management, operation, access, as well as use and discontinuation of human biobanks and genetic research databases.

- Artene et al. break down biobanks into two parts, (1) *biological material* that is collected, processed, and stored for long time and (2) the database, including *information about demographical and clinical data* for each sample and also associated with the bank inventory and the main activities: biospecimen collection, processing, storage, and distribution of biological material [4].

From all these definitions, a biobank can simply be considered as the:

- Physical structure containing the biospecimens (biorepositories).
- Associated data and information (demographic, phenotypic, clinical, lifestyle, among several other pertinent information) in a database.
- The system that grants access to biospecimens and linked information for research use, which is the governance and management of the infrastructure.
- The system that is responsible for ensuring ethical and legal issues pertaining to privacy, confidentiality, and security are adhered to, which is the governance and management of the infrastructure.

Human biobanks contain all sorts of samples, including tissue, cells (cell lines, stem cells), all kinds of body fluids (blood, plasma, serum, other blood components, saliva, urine, CSF), nucleic acids, proteins, and any pertinent biologic sample relevant to human research. Thus, biobanks are rich sources of biomaterials for biomarker and drug development.

In a historical context, "biobanks" have been in existence for centuries, probably since research on human tissues, cells, and body fluids began. This will have involved collection and storing samples for research use in any research facility. In the recent past, they have taken the form of academic repositories that store annotated samples for particular research projects. The investigator and a small research team usually manage these repositories. Alongside these single-user or research-goal-focused samples, various institutions and even governments establish centers that collect a large number and different types of samples for broader research uses. Additionally, various industries usually establish their own biorepositories where samples are managed for use in commercial research and product development. Modern biobanks, the ones poised to change the world of biomarker and other research in the future, are bigger and complex systems established through national and international efforts. These usually contain broader sample types from the entire population that enable various studies including genotype-to-phenotype correlations, case-control and cohort studies for biomarker discoveries, and drug development. Indeed, there are virtual biobanks, which are able to provide services on the physical locations of biospecimens and sample retrieval.

What are the different forms of biobanks? Similar to its definition, the classification of biobanks has also taken different approaches and hence terminologies.

They can be classified based on physical location and hence "ownership" or the types of samples it contains as:

- Hospital based.
- Academic institution based.
- Not for profit.
- Industry based.
- Government controlled.
- Network based.
- Disease based or population based.

From a research point of view, however, Gottweis and Zatloukal identify four types of biobanks [5]:

- Clinical based: these contain biologic samples with specific diseases and non-disease controls.
- Longitudinal population based: these contain longitudinal samples from a portion of the population that has been followed over a long time.
- Population isolate biobanks for study of genetic environment interactions: these biobanks have samples with homogeneous genetic and environmental setup of the population represented.
- Twin registries whereby samples from monozygotic and dizygotic twins are stored.

There are several other classifications; however, the BBMRI provides a simple and acceptable classification:

- Population based: these contain prospectively collected biospecimens focused on the study of the development of common and complex diseases which occur over time (e.g., diabetes, essential hypertension).
- Disease-oriented or clinical biobanks: these contain biospecimens and clinical data that are organized into disease categories.

11.4.1 Custodianship

Despite the established value of biobanks in biomedical research, there are issues on custody of biospecimens, data and information, as well as informed consent used for biospecimen collection and subsequent research use. A consequence of ethical and legal battles over biospecimen custodianship simply retards biomarker discovery and drug development and hence negatively impacts public health. The ethical approach to custodianship whereby participants organizations or stakeholders are governed by ethical principles is more attractive in advancing biomedical research than legal discourse. This ethical model enhances research while protecting research participants. The model requires that transparent, fair, and accountable principles be in place for managing biospecimens and data throughout the research

project(s), following termination of research. This custodian model requires some minimum management principles by biobanks:

- The biobank management should include defined principles that govern access to biospecimens and data. Decisions on biospecimens and data use should be based in part on the scientific merit of the project, ethical issues governing the study, and transparency to help ensure trust in participants.
- Biobanks must be governed by current legislature, federal or state, on sample and information collection, storage, use, sharing, and any other pertinent protections offered by law.
- Eminent or potential financial conflicts of interest must be managed properly. Any conflicts on the part of the biobanks, industry, researchers, institutions, or other stakeholders should be made explicit in the consent form.
- As the informed consent explicitly respects withdrawal of participants from studies, policy and logistics should be in place to manage discontinuation. When such situations occur, all aspects of the participant's samples, data, and information must be discontinued and removed from the system immediately.
- Biobanks need to be financially managed properly as they are entrusted with valuable resources. Plans should also be made to be able to transfer materials and information to other biobanks. This possibility should be in the consent form, such that participants are not surprised to know sometime that their information is with different biobanks.
- To help boost morale and trust, participants should be actively provided access to summaries of aggregate research data. Not only will they come to understand the low risk associated with their information use, but also the potential benefits of the research to society will enhance their future participation and probably participation by their family members and friends.
- Biobanks must develop operating procedures for sample acquisition, processing, annotation, storage, monitoring, release, distribution, and inventory management system for tracking.
- Biobanks should have policies for training of stakeholders including participants, researchers, funding agencies, biobank facility owners, and members of ethics committees and IRBs.
- National policy guidelines should be followed, and a sound economic model must take into account the governance, funding sources, user fees, and intellectual property issues.

11.4.2 Privacy Protection

Participant confidentiality and privacy protection is critical to biobanking. Various models are used by different biobanks.

- Some protection is offered by the procedure proposed by the International Conference on Harmonisation of Technical Requirements. This process involves:
  - The participant samples and data are labeled with personal identifiers such as names and identification numbers.
  - Coded samples and data are labeled with at least one specific code and are devoid of personal identifiers.
  - Anonymized samples and data are initially single or double coded, but do not have any linkage between participant identifiers and the unique codes, because it is deleted at this stage.
  - Thus, samples and data can no longer be traced to the participants using the unique codes.
- Another model is the tissue trustee or honest broker model. Here, the appointed trustee's role involves protecting research participants, while facilitating progress in research. This individual coordinates activities among participants, the healthcare system (where biospecimens are acquired), and researchers. The trustee ensures biospecimens and participant data are made available to researchers, while protecting participant privacy and confidentiality. There are other variations to this model, including the use of a third-party subcontracted person, computerized informatics system, medical archivist, and alphanumeric codes.

Obviously these models put the onus on biobank managers and decision-makers to enforce ethical principles in regard to participant protection.

11.5 Ethical Concerns of Cancer Screening

With genomic research advances, especially GWAS, risk alleles have been identified for a number of cancers including breast, ovarian, and prostate cancers. The finding of these alleles may lead to prophylactic treatment or in many cases stratify individuals for active surveillance screening for early detection. Ethical, legal, and social issues (ELSIs) will significantly influence the screening for common cancers based on genetic risks or susceptibility. The ELSIs pertaining to screening for cancer based on genetic risk may face similar issues as other biomarker studies and utility. The screening may be longitudinal, possibly over the lifetime of the individual. In this case, sample storage and management, data security, need to trace participants at a later date, participant attrition, continuous capacity for the individual to consent, surrogacy of consent, and how to convey incidental findings during testing, among many issues, will need ethical consideration. Many of these issues will be addressed in the consent letter. However, other considerations in these screening tests are:

- How to handle incidental findings. The consensus of genomic research is that unsolicited findings with established risks, whereby interventional strategies are available, should be communicated to the participant. Therefore, there should be clear indication in the consent form as to how such findings will be made available and to which individual (e.g., family members, healthcare providers) and at what time in their life.
- Though possibly of minor concern, how people's behavior will be impacted by acquired knowledge on their genetic susceptibility to disease is unclear and needs consideration to prevent possible harm.
- Genetic risk assessment data may be collected and used for public health purposes to benefit a sector of the population without causing any harm. Identified risk alleles in the population can be used to make informed decisions of possible disease onset at specific ages. In this case, genotyping all newborns could potentially generate a wealth of data for actionable interventions, but this raises considerable ethical issues.
- Genotype-based screening should also consider the religious and cultural beliefs of others. An inclusive discussion on sample types, mode of collection, handling, and participation of family members, for instance, must be considered.
- Screening programs should be fair and equitable and refrain from preferences given to a select group of individuals based on socioeconomic, educational, or other elite status in the community.
- How genetic information will be used to stratify individuals for screening needs to be addressed.
- How can fairness be granted to individuals at low risk for a particular condition? Usually, these individuals are not offered the same level of surveillance as those at high risk, and yet some could still develop the condition. A good communication of potential harms and benefits associated with screening should be offered (e.g., need to avoid overdiagnosis and possible overtreatment and its associated side effects).

11.6 Ethical Concerns of Cancer Diagnosis

With the current efforts at early cancer detection, and its obvious advantage of possible cure, there have been intense genotype risk-stratified, targeted, and untargeted population screenings for cancer. The increasing public awareness of cancer screening tests (partly through direct-to-consumer advertising), coupled with the development of sensitive technologies, has resulted in cancers being detected at very early stages. While this is commendable, it has its side effect, that of overdiagnosis. By definition, overdiagnosis is the detection of cancers (usually indolent tumors), which have no major adverse effects on the health of the patient. In these circumstances, death is more likely to result from causes (e.g., cardiovascular disease) other than the cancer being detected. Another problem associated with screening is that, following a positive test result, invasive

procedures (e.g., biopsy) may be required for the final diagnosis. Following a definite diagnosis, the question then is whether or not to treat the cancer and how aggressive should treatment be? On many ethical and legal grounds, including patient anxiety and physician fears, treatment ensues, although this offers no benefit to the patient. Thus, the second side effect of this diagnosis is overtreatment. Both the invasive diagnostic tests and overtreatments are not without harms. Ethically, therefore, physicians and the public should be knowledgeable about the causes of overdiagnosis and hence take active actions to reduce their occurrences. Some causes of overdiagnosis are:

- The limitations with screening tests, including lead-time and length-time biases. The ideal screening biomarker should be able to predict the natural course of the cancer at diagnosis. However, such biomarkers are virtually not in existence.
- Analytical sensitivities of diagnostic technologies. With improvements in technology, tumor signatures are being detected even at extremely low levels, usually associated with small and insignificant tumors.
- Increased public education and fears about cancer diagnosis.
- A strong push from diagnostic and pharmaceutical companies, who aim to make big profits.

Actions should therefore be implemented to prevent cancer overdiagnosis. Herein are some directions:

- Efforts are being made by societies in this direction to help curtail overdiagnosis. For example, efforts aimed at identifying tests that potentially can result in overdiagnosis are the "Choosing Wisely" and "Smarter Medicine" initiatives launched by the American Board of Internal Medicine and the Swiss Society of Internal Medicine, respectively.
- Other interventions to minimize overdiagnosis include implementation of targeted screening (of those at very high risk, e.g., smokers and lung cancer), reductions in screening frequencies, and avoidance of useless screening tests such as the PSA test for prostate cancer.
- A major cause of overdiagnosis is financial interests of industry. In making diagnostic guidelines, therefore, any expert member of the panel with heavy financial conflict of interest should be excluded.
- The public needs adequate knowledge on screening tests being offered in terms of potential benefits vs. harms, so that informed decisions can be made.

11.7 Industry-Academia Relationships

There are disturbing ethical concerns in the manner that cancer researchers and oncologists relate to industry. First, substantial financial conflict of interest (FCOI) exists between several academic cancer research centers, investigators, and industry, mostly the pharmaceutical industry. Second, the recent surge in direct-to-consumer

advertising poses ethical issues for industry but also for oncologists in the care of their patients. These relationships are complex and so are the ethical dilemmas.

There is a necessary relationship between industry and researchers/oncologists in biomedical product development. While in the past studies such as clinical trials were funded by the government, >40 % of all clinical trials are now sponsored fully or in part by industry. Thus, industry has a profound effect on cancer research and oncology practice. The FCOI relationships are often complex, ranging from research support, honoraria, fees, gifts, payments for travels, time, speeches, meals to academic centers, and even in kind, and not as explicit direct monetary payments. Some investigators own shares, royalties, licenses, and other products that they share with manufacturers. Indeed, the FCOI issue is escalating, as has been observed at meetings such as those of the American Society of Clinical Oncology. Not surprisingly, this escalated FCOI is now even attached to abstracts that receive better attention at these meetings, most probably because those involved are the influential people at these meetings and hence can easily feature their research at these prominent sessions. The relationship between cancer researchers/oncologists and industry is understandable and can be positive. This relationship is partly because the pharmaceutical industries have or develop the agents and the infrastructure to carry out clinical trials on the research output from the academic researcher. These relationships have driven the development of many products currently deployed for cancer care. But there are issues with some of these FCOI relationships. How are research integrity, research conduct, decisions, and scientific bias influenced by these FCOIs? Profit sharing by industry and researchers may drive poor or inaccurate science, all of which may lead to patient and public mistrust.

A number of policies have been formulated and implemented to curtail these potentially unhealthy relationships. There are rules mandating disclosure of any conflict of interest at meetings or in publications. While well intended to increase transparency and to gain public trust, it is unclear whether such declaration has any effect on the relationships between FCOI and research conduct, integrity, and oncologic practice. In view of this uncertainty, an alternative solution is the call for academic medical centers to desist from receiving gifts, meals, and payments for travel or participation in continued medical education (CME) by industry. In oncologic practice, attempts have been made to distance the physician from industry through the issuing of vouchers for the underprivileged instead of providing samples. Additional measures are the provision of research support to the academic institution for general use rather than to individual researchers, and the cessation of payments for CME.

In 2007, the US Congress introduced a bill (S.2029), to help curtail the potential unethical relationships between industry and physicians. Named the Sunshine Act, it puts in place some demands to deter researchers/physicians from such actions and to inform the public about specific FCOI of their providers. These include:

- The requirements for companies to submit information regarding their financial relationships with providers to the US Department of Health and Human Services (HHS).
- The provided information is made public via a searchable database, Open Payments.

Patients can search Open Payments to ensure their providers are not conflicted in the way they are being managed (e.g., what prescription they are given), and physicians can also search the database to ensure their information is accurate. In fact, Open Payments provides a detailed breakdown of the different types of potential FCOIs such as "consulting fees," "serving as a speaker," "honoraria," "royalty or license," and many more. While this may appear to offer some solutions, the picture is more complex than this. For instance, physicians feel stigmatized and hence shy away from engaging in even honest and healthy financial deals with industry. It clearly calls for moral reasoning such as evoked in duty-based theories and ethics of care on the parts of physicians and cancer researchers.

The other problem the industry has created that poses ethical challenges to physicians and involves biomarker utility is direct-to-consumer (DTC) advertising. DTC advertising began in the 1980s and has grown since then. Although disguised to have some benefits to the patient, the actual motive of industry is to increase sales and hence profit. The majority ($>80\,\%$) of DTC advertising involves oncology products, primarily drugs, genetic testing, and medical imaging. Ethically, the industry could claim its primary goal is to improve patient education, so as to enhance communications with their physicians for optimal care. One can even argue that a principle of the Georgetown Paradigm, respect for autonomy, is being satisfied in this circumstance. This is so because DTC advertising could empower patients, enabling them to make choices and informed decisions and take actions based on beliefs of what has been advertised. The real problem, however, is that patients are putting pressure on their care providers for products that may not offer any advantage over existing low-cost alternatives. Indeed, $>40\,\%$ of patients who request for DTC products actually receive them. A role of the physician is to be an unbiased individual with the patient's well-being at heart. However, under situations whereby the physician has FCOI on an advertised DTC product, that may blur or compromise such a role.

11.8 Summary

- Bioethics came into being as a necessity to help improve upon the standards of patient care and the conduct of research involving humans. Principlism or the Georgetown Paradigm offers unique principles and rules, which when followed should enable fair and moral treatments of everyone.
- Biomarkers cannot be developed in the absence of patients and other research participants. But given that private information is collected from research

participants, and usually linked to their "biomarker" of genetic data, a major concern emerges in regard to privacy protection, confidentiality, and security. A mechanism to help offset this problem is the issuing of a consent letter or form.

- Consent for research can be worded differently, resulting in several possible models. However, two extremes exist, narrow and broad consent models, in regard to the usage of biobank samples.
- The narrow consent model is usually worded to offer optimum protection for the research participant. This type usually focuses the consent on specific research projects, requiring that whenever samples are required for different research projects that the participant is recontacted for a different consent.
- On the contrary, consent can be worded such that decisions on sample usage are left to institutional review boards and ethics committees (broad consent model).
- Trust, satisfaction, and any lack of potential exploitations favor broad consent model, while any lack of trust, e.g., potential use of genetic information for discriminatory purposes, results in narrow consent models.
- Biobanks are critical to biomarker development, as they collect, annotate, process, store, and manage valuable high-quality biospecimens and data. Biobanks exist in various forms and at various levels of complexities.
- Cancer screening and diagnosis are associated with diverse ethical, legal, and social issues (ELSIs), worthy of note. The concept of overdiagnosis, and potential overtreatment, is a major concern in the field of molecular diagnosis.
- Biomarkers and novel therapeutic agents cannot be developed in the absence of industry and academic medical center collaborations. However, these relationships can result in significant financial conflicts of interest that can adversely impact the authenticity of a particular product.
- Direct-to-consumer advertising, although it may appear to empower the public, has problems with doctor-patient relationships.

References

1. Khan A, Capps BJ, Sum MY, et al. Informed consent for human genetic and genomic studies: a systematic review. Clin Genet. 2014;86:199–206.
2. Kauffmann F, Cambon-Thomsen A. Tracing biological collections: between books and clinical trials. JAMA. 2008;299:2316–8.
3. Zika E, Paci D, Braun A, et al. A European survey on biobanks: trends and issues. Public Health Genomics. 2011;14:96–103.
4. Artene SA, Ciurea ME, Purcaru SO, et al. Biobanking in a constantly developing medical world. ScientificWorldJournal. 2013;2013:343275.
5. Gottweis H, Zatloukal K. Biobank governance: trends and perspectives. Pathobiology. 2007;74:206–11.