Leena Rawal · Sher Ali *Editors*

Genome Analysis and Human Health

Springer

Genome Analysis and Human Health

Leena Rawal • Sher Ali
Editors

Genome Analysis and Human Health

Springer

Editors
Leena Rawal
Department of Cytogenetics
National Reference Laboratory
Dr Lal PathLabs
New Delhi
India

Sher Ali
Centre for Interdisciplinary
Research in Basic Sciences
Jamia Millia Islamia
New Delhi
India

ISBN 978-981-10-4297-3 ISBN 978-981-10-4298-0 (eBook)
DOI 10.1007/978-981-10-4298-0

Library of Congress Control Number: 2017941725

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Genome research has indeed pampered our optimism as it was construed that the same would enhance our understanding on the mechanisms that lead to genetic diseases. In the context of human health, genetics involves study on single gene and their regulation to improve public health and prevent diseases. Genetic research helps to identify diseases and health problems that are more likely to be influenced by genetic factors. Genetic tests enable the risk assessment and determine the predisposition of an individual to various diseases by uncovering the mutations or variations in the genome. Such information may be useful in managing an individual's lifestyle and healthcare system. In addition to testing for particular conditions, genetic research provides solutions to health problems caused by genetic abnormalities and mutations either by medications or genetic modification. Most genetic disorders cannot be cured; however, many people have restored their health and avoided potentially life-threatening diseases with the help of genetic research by taking due precautions coupled with advanced medicaments and changed lifestyle.

Continuous technological improvements in DNA sequencing have created an ambiance par excellence that a large number of disease-causing microbe and viral genomes are sequenced on regular basis. The availability and the integration of genetical information have been the driving forces toward our understanding of the normal and abnormal genomes.

We believe that newer and far more despicable diseases would continue to emerge so also the quest to fight these diseases. Conceptually, advances in genetical knowledge fueled by technology could be used to prevent diseases creating much healthier gene pool. Thus, genome analysis both for normal and diseased ones would continue to upgrade our knowledge ensuring hope and assuring a healthy world.

New Delhi, India

Leena Rawal
Sher Ali

Contents

1 **Genetic and Epigenetic Regulation of Autophagy in Cancer** 1
Anup S. Pathania, Ubaid S. Makhdoomi, and Fayaz A. Malik

2 **Cancer Genomics and Precision Medicine: A Way Toward Early Diagnosis and Effective Cancer Treatment** 31
Ritika Tiwari and Bushra Ateeq

3 **Genetics of Liver Diseases** 43
Nirupama Trehanpati, Dinesh Rawat, and Hamed Hemati

4 **Implication of Pre-replication Complex Proteins in Human Disease** 67
Abid Khan, Arindam Chakraborty, and Supriya G. Prasanth

5 **Non-muscle Myosin II Motor Proteins in Human Health and Diseases** 79
Venkaiah Betapudi

6 **Bioinformatics Databases: Implications in Human Health** 109
Leena Rawal, Deepak Panwar, and Sher Ali

7 **Genomics of the Human Y Chromosome: Applications and Implications** 133
Sanjay Premi, Jyoti Srivastava, and Sher Ali

8 **Human Microbiome: Implications on Health and Disease** 153
Princy Hira, Utkarsh Sood, Vipin Gupta, Namita Nayyar, Nitish Kumar Mahato, Yogendra Singh, Rup Lal, and Mallikarjun Shakarad

Introduction

A human genome contains approximately 3.3 billion DNA bases. By comparing an individual's sequence to a human genome reference sequence, DNA changes are detected at almost every DNA base position. Depending on the location, changes in the DNA bases may or may not alter the gene functions. Even if it does not affect a gene, the changes may still affect the genetic structure of an individual. The data gained from the genome sequence is subjected to various bioinformatics annotation tools and analyzed so as to decipher the core reasons behind these DNA changes or variations that may have an impact on individuals' health. Some of the DNA changes identified in the sequence are linked to genetic disorders that can be inherited within a family. These variations can affect the molecular pathways of the cell, leading to alterations in the physical trait, or can be linked to risk for common diseases.

With the advent of new high-throughput technologies, the conventional focus on genetics and single genes is drifting toward the study of the whole genome including the exome sequencing, the study of complex genes, gene–gene interactions, and the association between genes and environment (epigenetics). This evolution in genomics, genetics, and other related molecular biology technologies has created substantial avenues for the advanced understanding, prevention, treatment, and cure of human diseases.

This book is intended to provide basic information on genome analysis and its impact on human health. It focuses on different approaches that have been adopted to address one or the other issues related to human health including cancer. Additionally, it covers the domain that still must be explored in order to understand the signaling processes in the genome and gene–gene interactions encompassing a large number of still undefined and poorly understood interactomes that affect human health.

About the Editors

Sher Ali is an internationally respected scientist in the field of genomics and molecular genetics. He received his PhD from the University of Delhi in 1981 and subsequently worked as an Alexander von Humboldt fellow at the Max Planck Institute of Immunobiology in Freiburg, Germany. He retired from the National Institute of Immunology (NII), New Delhi, in 2015 after serving the institute for more than 25 years. At the NII, he was head of the Molecular Genetics Laboratory. Currently, he is working as a professor at the Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi.

His research interests include genome analysis, molecular human and animal genetics, germ line genetics, toxicogenomics, DNA diagnosis, and cancer genetics. His outstanding research has attracted attention from various pharmaceutical companies. Dr. Ali has published over 100 peer-reviewed papers in reputed international journals. He is a member of all three academies of sciences in India, namely, the Indian Academy of Sciences, National Academy of Sciences, and Indian National Science Academy, and has been a reviewer and editorial board member for various prestigious journals such as *Molecular and Cellular Probes*, *American Journal of Cancer Genetics*, *Scientific Reports*, etc. He is the recipient of several national and international awards, including most recently an Alexander von Humboldt-Stiftung fellowship in Bonn, Germany, and the N.K. Iyengar Memorial Gold Medal.

Leena Rawal is currently working as a senior scientist at the Department of Cytogenetics, Dr. Lal PathLabs Limited, New Delhi. She was awarded PhD degree in Molecular Genetics by the University of Delhi in 2015 and subsequently worked as a postdoctoral fellow at the National Institute of Immunology, New Delhi. She has since worked as a senior scientist and head of the Molecular Diagnostics and Research Division at a renowned pharmacogenomics-based organization. Dr. Rawal has published several original papers in international peer-reviewed journals like *PLOS ONE*, *BMC Genomics*, *DNA and Cell Biology*, *Gene*, *Journal of Cellular Biochemistry* and the *Journal of Biomolecular Structure and Dynamics*, etc. Her expertise lies in the area of human and animal genetics, proteomics, cytogenetics, and bioinformatics. She has presented her work at various national and international platforms.

Genetic and Epigenetic Regulation of Autophagy in Cancer

1

Anup S. Pathania, Ubaid S. Makhdoomi, and Fayaz A. Malik

1.1 Introduction

Cancer is a class of disease characterized by cells' abnormal growth and division. Cancer cells grow very fast in an uncontrolled manner as compared to normal cells and form lumps or tissue mass called tumor (except leukemia). Solid tumors are benign in nature as long as they are localized to their tissue of origin and become malignant when cells migrate to distant vital tissues of the body like, brain, bone, liver, lung, etc. through blood or lymphatic system. The transformation of normal cells into cancerous cells is a multistep process caused by mutations. The process of accumulating mutations normally takes many years, and several mutations are needed for a normal cell to acquire such oncogenic behavior. In all cancers, these mutations are mainly found in tumor suppressor and proto-oncogenes. Mutations in tumor suppressor genes render them with loss of functionality and inactivate their inhibitory properties on cell growth and division. Such mutations are also known as loss of function mutations and they are common in cancer cells. Some common examples of tumor suppressor genes bearing such mutations in cancer cells are retinoblastoma gene (RB), p53, BRCA (breast cancer genes), APC (adenomatous polyposis coli), PTEN (phosphatase and tensin homologue), p27, etc. (Lee and Muller 2010). In normal cells, regulated counterparts of oncogenes are known as proto-oncogenes that control cell division and proliferation. Mutations in proto-oncogenes deregulate their activities (also known as gain in function) leading their conversion into the cancer-forming

A.S. Pathania • U.S. Makhdoomi • F.A. Malik (✉)
Department of Cancer Pharmacology, CSIR-Indian Institute of Integrative Medicine, Canal Road Jammu, Jammu and Kashmir 180001, India

Academy of Scientific and Innovative Research (AcSIR), New Delhi 110001, India
e-mail: fmalik@iiim.ac.in

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_1

oncogenes like those of RAS, Myc, HER2, Cyclin D, Bcl-2, etc. (Lee and Muller 2010). Many such genetic changes that happened are like point mutations, insertions, deletions, gene amplifications, or chromosomal translocation of proto-oncogenes to another normal gene that dysregulate its expression. The factors behind these genetic changes are random and are not cell specific and are different in different types of cancers, e.g., lung cancer, 90% of cases are associated with cigarette smoking and risk of cancer increase with tobacco dose and is the reason for the mutations in lung cells (Sasco et al. 2004). Similarly, the use of alcohol, tobacco, and human papillomavirus or Epstein-Barr virus infection are important risk factors for head and neck carcinomas (Goldenberg et al. 2004; Leemans et al. 2011). The presence of carcinogens like heterocyclic amines (HCAs), N-nitroso compounds (NOCs), and heme in red meat damages the DNA of cells that line the digestive system (Alexander and Cushing 2011). However, it is still to be understood the reasons behind the acquisition of mutations in key genes of normal cells which become nonresponsive to cellular homeostasis. A nonsmoker can develop a lung cancer; women who take normal food and exercise regularly with no genetic history of breast cancer can have this type of tumor. As our understanding of cancer is continuously growing, it has been established that besides environmental factors, genetic predisposition also plays a major role in cancer development. Genetic predisposition means increase in the likelihood of developing a particular disease based on the genetic makeup of person he or she acquired from his parents or ancestors. Mammalian cells have two copies of genes, and as long as cells contain at least one functional copy, the gene remains fully functional. The examples of such genetic mutations in mammalian cells include the retinoblastoma gene (RB), p53, BRCA1 and BRCA2 in breast cancer, TERT in melanoma, APC in colon cancer, etc. However, it is to mention that more than 75% of cancers are sporadic, which means they occur by chance and have no familial history. Apart from genetic defects, influence of epigenetic changes is also equally responsible for cancer development and progression. Though epigenetic alterations do not include any changes in cellular DNA sequences, they are mitotically and meiotically inheritable. Epigenetic changes can switch on or off the genes by controlling their transcription. The common epigenetic events occur in cells include methylation of cytosine bases of DNA present within CpG dinucleotides that are found in 5′-end regulatory regions of many genes and almost in all housekeeping genes. Similarly, acetylation or deacetylation occurs on the lysine residues present within N-terminal tail of histone core of the nucleosome as a part of gene regulation. Small noncoding regulatory RNAs (siRNAs) also play a critical role in tumorigenesis. In this chapter, we will try to discuss the role of genetic and epigenetic changes associated with the process of autophagy and its implications in cancer. Autophagy is a catabolic process that uses cell lysosomal machinery to degrade unnecessary or dysfunctional cellular constituents. This process not only eliminates the damaged organelles but also provides raw materials to the cells under stress-related conditions to maintain homeostasis. Autophagy has been shown to play an important role in cancer progression (Choi 2012), angiogenesis

(Du et al. 2012), epithelial to mesenchymal transition (Li et al. 2013), metastasis (Peng et al. 2013), resistance (Peng et al. 2013), inflammation (Levine et al. 2011), infection (Deretic 2010), and neurological disorders (Nixon and Yang 2012). However, prolonged autophagy activation may also lead to cell death due to excessive catabolism, known as type II programmed cell death or autophagic cell death.

1.2 The Process of Autophagy

Autophagy starts with the formation of membranous structures or phagophore around the cytoplasmic sites known as pre-autophagosomal structure (PAS) discovered in yeast (Klionsky 2007). Phagophore membrane formation is mainly contributed by ER, Golgi, and endosomes under the control of various signaling events. Sequential activation of specific autophagy-related genes (ATGs) is involved in the formation of phagophore. ATGs were first identified in yeast through genetic screening, and many of their homologues have been subsequently found and characterized in higher eukaryotes. Autophagic process mainly involved five events, induction, vesicle nucleation, vesicle elongation, autophagosomes formation, and autophagosome-lysosome fusion. The main regulator of autophagy induction in cells is the mammalian target of rapamycin (mTOR) (Jung et al. 2009), a nutrient sensor of the cell that inhibits autophagy under nutrient-rich conditions. mTOR phosphorylates autophagy protein ATG13L at multiple serine residues and mammalian homologues of ATG1 protein ULK-1 or ULK-2 (UNC-51-like kinases). mTOR-mediated phosphorylation of ULK-1 and ULK-2 inhibits their activity, rendering them unable to phosphorylate and activate focal adhesion kinase family-interacting protein of 200 kD (FIP200). ULK-1, ULK-2, and ATG13 form a complex with FIP200 known as ATG13L-ULK-1/2-FIP200 complex which recruits other proteins for autophagosome formation. During nutrient starvation, cellular ATP levels decrease in cells; as a consequence of which, cAMP levels increase which further activates energy sensor protein, AMPK (AMP-activated protein kinase). AMPK inhibits mTOR activity and promotes hypophosphorylation of ATG13L which favors the formation of active ATG13L-ULK-1/2-FIP200 complex and hence the induction of autophagy (Jung et al. 2009). Vesicle nucleation starts with the recruitment of ATGs to the growing phagophore. Although the process is less understood, type III PI3 kinases play an important role in the recruitment of ATGs to the phagophore. Type III PI3 kinases consist of single catalytic subunit VPS34 (homologue of yeast vacuolar protein sorting-associated protein 34) and produce PtdIns3P (phosphatidylinositol 3-phosphate). VPS34 interacts with autophagic proteins Beclin and VPS15 and forms type III PI3 kinase complex, also known as Beclin-VPS34-VPS15 complex. This complex is recruited to PAS by ATG14 where it acts as a localization site for most ATG proteins that facilitate autophagosome formation (Mizushima et al. 2011). The third step, vesicle elongation, involves the expansion of

phagophore membranes around PAS for the formation of autophagosomes. Two ubiquitin-like conjugation systems, ATG5-ATG12 and ATG8 are involved in this process (Nakatogawa 2013). Eloborating these systems is beyond the scope of this chapter. However, these interactions lead to the lipidation of ATG8 (LC3B in mammals), a typical marker of autophagy, with phosphatidylethanolamine (PE) that insets ATG8 into autophagosome membranes. Once the autophagosomes are formed, they fuse with lysosomes to form autolysosomes in which the autophagic substrates are degraded and recycled (Fig. 1.1).

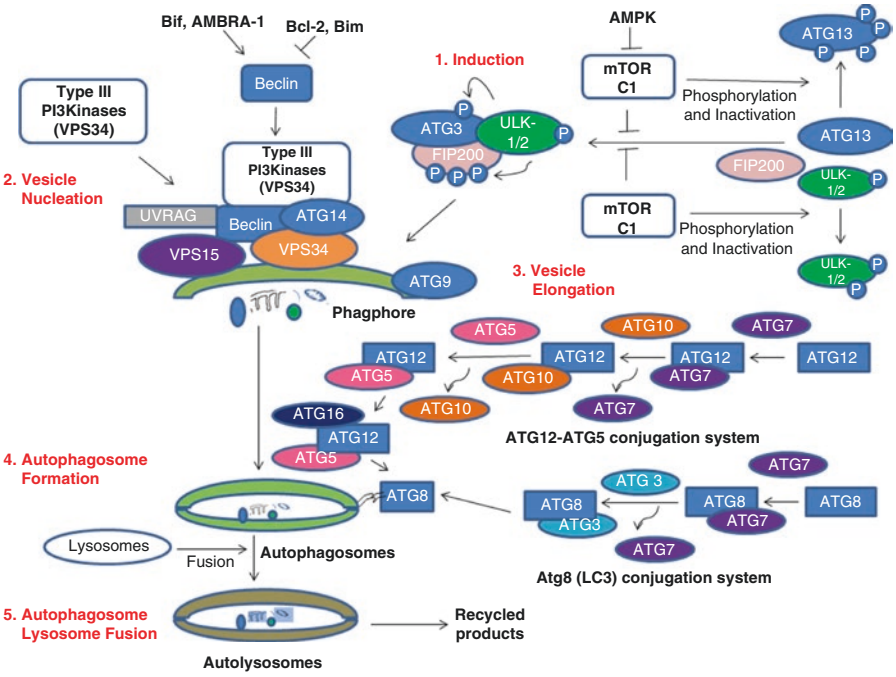

Fig. 1.1 The autophagic process. Formation of autophagosomes involves the various steps. The process begins with the formation of double membrane structures (phagophores) around the cargo molecules (damaged organelles and macromolecules, misfolded proteins, pathogens, etc.) in the cytoplasm. The process is induced by formation of ATG13-ULK-1/2-FIP200 complex at the autophagosome assembly site in mammalian cells. mTOR promotes the phosphorylation and inactivation of ATG13 and ULK-1/2 under nutrient-rich conditions. Inhibition of mTOR triggers the formation of this complex and induces autophagy in cells. Next step is vesicle nucleation which is performed by type III PI3 kinases along with Beclin and VPS15. Type III PI3 kinase subunit, VPS34, interacts with autophagic proteins Beclin and VPS15 and forms Beclin-VPS34-VPS15 complex which is recruited to the phagophore by ATG14. Beclin-VPS34-VPS15 complex acts as a nucleation site for most of ATGs involved in autophagosome formation. Vesicle elongation involves two ubiquitin-like conjugation systems, ATG5-ATG12 conjugation systems and ATG8 (LC3B in mammals) conjugation systems, which promote the lipidation of LC3B with phosphatidyl ethanolamine and complete the formation of autophagosomes. Once autophagosomes are formed, it fused with lysosomes to form autolysosomes which degrade its inner constituents via lysosomal hydrolases

1.3 Autophagy in Cancer

Autophagy dysregulation is a common phenomenon in almost all cancers, and modulating this process is an area of great interest in cancer drug discovery. The expression of many autophagy-associated genes is altered in cancer cells that leads to tumor progression. The first direct link between autophagy and cancer was established in early 1999 by Aita et al., demonstrating that the mono-allelic deletions of beclin (an ortholog of yeast ATG6 present on chromosome 17q21) are found in human breast and ovarian carcinoma cell lines (Aita et al. 1999). The same year Liang and co-workers reported that beclin promotes autophagy in autophagy-defective yeast having targeted the disruption of *agp6/vps30* and in breast cancer cell line MCF-7 (Liang et al. 1999). Beclin-induced autophagy inhibits cellular proliferation, clonogenic survival, and tumorigenesis in mice (Liang et al. 1999). Further work by Qu et al. (2003) demonstrated that heterozygous disruption of beclin increases cellular proliferation, spontaneous tumorigenesis, and development of HBV-induced premalignant lesions in mouse tumor models. Southern blots and mutational analysis of genomic DNA of beclin (+/−) mice did not reveal any deletions or rearrangements in the remaining normal beclin allele, and hence inactivation of only one allele is sufficient to promote tumorigenesis. These results show that Beclin is a haploinsufficient tumor suppressor gene and it does not follow Knudson two-hit hypothesis where mutations in both alleles are required for tumor suppressor gene to lose its function (Qu et al. 2003). These findings also revealed a new role of autophagy in preventing dysregulated growth of tumor cells besides maintaining homeostasis. Beclin is a Bcl-2 homology (BH)-3 domain-only protein localized throughout cytoplasm including mitochondria, ER, and nucleus (Kang et al. 2011). Beclin gene maps to a region of 150 kb centromeric to BRCA1 gene present on chromosome 17q21 and encodes a 2098-bp transcript, with a 120-bp 5′ UTR, 1353-bp coding region, and 625-bp 3′ UTR (Aita et al. 1999). Beclin contains three domains: N-terminal short BH3 domain (105–125 residues), a central coiled-coil segment (114–269), and C-terminal evolutionary conserved domain (244–337) (Sinha and Levine 2008; Huang et al. 2012). The Bcl-2 member proteins Bcl-2 and Bcl-xL regulate autophagy via binding to BH3 domain of Beclin and inhibit its association with class III PI3 kinases, an important autophagy regulator in cells (Pattingre et al. 2005; Ku et al. 2008). Furthermore, phosphorylation of BH3 domain at threonine 308 residue by proapoptotic kinase Mst1 stabilizes Beclin-Bcl-2 interactions and inhibits the formation of ATG14L-Beclin1-Vps34 autophagic complex (Maejima et al. 2013). Another proapoptotic BH3 only protein BCL2L11 (also known as BIM) inhibits autophagy in cells by interacting with Beclin and facilitates its binding with dynein protein DYNLL1. Starvation induces BIM phosphorylation through MAPK8/JNK pathway and abolishes BIM-DYNLL1 interactions, allowing dissociation of Beclin from BIM, and induces autophagy (Luo and Rubinsztein 2013). The central coiled-coil domain of Beclin is required for its interactions with autophagy proteins ATG14 and ultraviolet radiation resistance-associated gene (UVRAG) which forms Beclin-ATG14 or Beclin-UVRAG heterodimers during autophagy (Li et al. 2012). The third evolutionary conserved domain of Beclin is required for its binding with Vps34 and lipid membranes (Furuya et al. 2005; Huang et al. 2012). Mutations in this region hinder Beclin binding with membranes and compromise autophagy;

however, it has no effects on Beclin interactions with other autophagy mediators like UVRAG and ATG14 (Huang et al. 2012; Fu et al. 2013). Mutations in Beclin gene are found in many cancer types and have been associated with the poor prognosis. In prostate carcinoma tumor, low oxygen and androgen deprivation trigger AMPK activation which induces autophagy via beclin activation. The induced autophagy is protective in nature and its pharmacological or genetic inhibition induces apoptosis in these tumor cells (Chhipa et al. 2011). Furthermore, Beclin and its counterpart LC3B are involved in the pathogenesis of benign prostatic hyperplasia (BPH) cells and promote androgen independence in prostate cancer cells (Liu et al. 2013a).

1.4 Autophagy-Associated Genes: Mutations and Role in Cancer

There are 32 autophagy-related genes or ATG genes till discovered, out of which 18 ATGs are directly involved in autophagosome formation upon starvation (Mizushima et al. 2011). Most of ATG genes are evolutionary conserved between yeast and mammals. Somatic mutations in ATG genes are frequently observed in different cancers. Frameshift mutations in ATG2B, ATG5, ATG9B, and ATG12 are found in gastric and colorectal carcinomas (Kang et al. 2009). The frequency of such mutations is immense in carcinomas with high microsatellite instability as compared to those with low microsatellite instability. DNA sequence analysis of gastric and colorectal carcinoma patients found single-base deletion mutations in exon 20 of ATG2B and in exons 8 and 10 of ATG5, identical deletion mutations in exon 1-1, and three identical deletion mutations in exon 1-2 of *ATG9B* (Kang et al. 2009). Another autophagy gene, ATG12, is commonly mutated in breast cancer cells targeted against HER2-based therapies. ATG12 is upregulated in trastuzumab-resistant HER2-positive breast cancer cell lines as compared to trastuzumab-sensitive cell lines (Cufi et al. 2012). Quantitative real-time PCR-based arrays of 84 autophagy genes in trastuzumab-responsive SKBR3 and trastuzumab refractory JIMT1 breast cancer cell lines reveal the overexpression of ATG12 in JIMT1 cells as compared to SKBR3 (Cufi et al. 2012). Genetic knockdown of ATG12 by small hairpin RNA sensitizes breast JIMT1 cells to trastuzumab and HER1/HER2 tyrosine kinase inhibitors. Trastuzumab treatment showed strong tumor growth inhibitory effect in ATG12-shRNA/JIMT1 xenograft animal models as compared to wild-type ATG12 expressing JIMT xenografts (Cufi et al. 2012). Additionally, autophagy gene UVRAG is found to be mutated in many cancers (Liang and Jung 2010; Ionov et al. 2004). Frameshift mutations in UVRAG are present in colorectal and gastric carcinomas with high microsatellite instability (Kim et al. 2008). UVRAG is a Beclin-interacting protein that associates with Beclin-VPS34-VPS15 complex and promotes vesicular trafficking and autophagosome formation. UVRAG suppresses the proliferation and tumorigenicity in human colon cancer cells (Liang et al. 2006). Genetic silencing of UVRAG or Beclin by specific siRNA increases radiation-induced DNA double-stranded breaks and apoptotic cell death in 5-fluorouracil (5-FU)-treated irradiated colorectal cancer cells. UVRAG and Beclin interact with each other during DNA repair, and UVRAG mutants that are unable to bind to Beclin show the greater extent of DNA damage during irradiation as compared to normal UVRAG expressed cells. Furthermore, Beclin and UVRAG suppression increases centrosome number in cells that leads to spindle malformations and chromosome segregation errors (Park et al. 2014). Abnormal UVRAG expression along

with BRCA1, BECN1, CCND1, and PTEN genes has been associated with human breast carcinogenesis. The mRNA levels of these genes are downregulated in breast cancer cells as compared to normal breast tissues and linked with the pathogenesis of the disease (Wu et al. 2012). In non-medullary thyroid carcinoma (NMTC), genetic variations in autophagy genes increase susceptibility for cancer progression and outcome. NMTC patients show the statistically significant relation between ATG5 genetic variants and susceptibility for NMTC. G allele of the ATG5 rs2245214 SNP is mutated in NMTC patients and shows association with prognosis of the disease (Plantinga et al. 2014). Another important link between autophagy and thyroid carcinogenesis is single nucleotide polymorphism known as Thr300Ala polymorphism (threonine at position 300 is replaced by alanine, rs2241880) in ATG16L gene, which increases susceptibility for thyroid cancer. One possible mechanism for such association between ATG16L and thyroid cancer is the modulation of pro-inflammatory cytokine IL-1β by ATG16L which hinders its antiproliferative effect in thyroid cancer cells (Huijbers et al. 2012). Furthermore, Thr300Ala polymorphism in ATG16L is also associated with the increased risk of developing colorectal carcinoma, and patients carrying the less common GG genotype are at higher risk than those carrying more conmen AA genotype (Nicoli et al. 2014). Additionally, colorectal cancer patients show enhanced expression of ATG10, which is associated with lymphovascular invasion and lymph node metastasis. ATG10 is highly upregulated in patients with sporadic colorectal cancer and is involved in metastasis and tumor invasion (Jo et al. 2012). Patients who did not express ATG10 have significantly higher disease-free survival and overall survival rate than those bearing ATG10-expressing tumors. Colorectal carcinoma cell lines AMC5, LoVo, SW480, SW48, HCT15, DLD1, RKO, and CaCo2 show higher ATG10 expression as compared to normal colorectal cancer cell line CCD841. Silencing of ATG10 by using siRNA approach suppresses cell proliferation in HCT116 cells (Jo et al. 2012). Furthermore, mutations in 5q14 regions of ATG10 are found in ovarian (Ramus et al. 2003), gastric (Oga et al. 2001), and pancreatic cancer (Shiraishi et al. 2001). ATG10 linked two SNPs, rs1864182 and rs10514231, which are associated with risk factors in developing breast cancer. These genetic variants of ATG10 increase the susceptibility of breast cancer in Chinese population (Qin et al. 2013). Furthermore, ATG10 expression is upregulated in mesenchymal stem cells along with ATG12 and LC3B in serum-starved breast cancer cells. The increased expression of these proteins supports cell survival and growth by providing energy and secreting anti-apoptotic proteins (Sanchez et al. 2011). Serum starvation decreases the proliferation in breast cancer cell line MCF-7; however, its co-culture with normal or serum-starved mesenchymal stem cells increases their survival rate and proliferation. Inhibition of autophagy with autophagy inhibitors chloroquinone (CQ) and bafilomycin or by Beclin silencing decreases cell survival to a great extent signifying the protective role of autophagy during stress (Sanchez et al. 2011).

1.5 Genetic Regulation of Autophagy in Tumor Suppression and Promotion

After the discovery of Beclin having both tumor-suppressive and autophagy-inducing functions, many autophagy genes have been discovered, and mutations in these genes have been linked to tumor progression and growth. Mutations and deletions of three critical autophagic genes, ATG2A (1%), ATG7 (2%), and ATG13 (5%), are

found in nasopharyngeal carcinoma (NPC) patients (Lin et al. 2014). Although these changes are not significant, this is the first report about such genetic lesions in the ATG genes of cancer cells. Another important autophagy protein that has tumor-suppressive functions is ULK-1, a member of ATG13L-ULK-1/2-FIP200 complex. The mRNA and protein levels of ULK-1 are lower in breast cancer tissues as compared to matched normal tissues. Immunohistochemical staining of ULK-1 in 298 nonmetastatic invasive breast cancer tissues revealed lesser expression of ULK-1 in 70% of cases, whereas the adjacent noncancerous tissues have moderate to strong expression. The diminished ULK-1 expression is associated with reduced autophagic capacity and the progression of the disease (Tang et al. 2012). These findings also suggest the use of ULK-1 as a novel prognostic biomarker for breast cancer patients. ULK-1 and its counterpart ULK-2 are the transcriptional targets of tumor suppressor protein p53 in cells. Their transcription is upregulated by p53 during DNA damage which induces autophagy in cells. Similarly, DNA-damaging agents etoposide and camptothecin induce autophagy through this mechanism triggering cell death (Gao et al. 2011). Furthermore, treatment of human colon cancer cell lines with different p53 status including HCT116 (p53 wild-type cells), HCT116/p53KO (p53 knockout cells), RKO and RKO-E6 (p53-blunted), and human bone osteosarcoma epithelial cells (wild type or knock out p53) with DNA-damaging agent camptothecin shows reduced expression of ULK-1 in p53 null cells as compared to p53-positive cells. Additionally, ectopic expression of ULK-1 enhanced autophagy in U2OS cells and shows additive effect with rapamycin on autophagic cell death. Additionally, ULK1 knockdown attenuates ectopically expressed p53-mediated autophagy and cytotoxicity in these cells (Gao et al. 2011). However, the role of ULK-1 in cancer is controversial, and the linearity of its expression with disease prognosis varies in different tumor types. In some tumors, the high expression of ULK-1 is associated with the severity of the disease and overall survival time in patients (Jiang et al. 2011). ULK-1 protein levels are upregulated in esophageal squamous cell carcinoma (ESCC) cell lines and tumor samples as compared to normal esophageal cells and tissues. ESCC cell lines EC109, KYSE140, KYSE510, and KYSE520 have high ULK-1 expression as compared to normal esophageal cell line NE1 (Jiang et al. 2011). Tumor to normal ratio of ULK-1 mRNA isolated from ECACC patients and normal persons is approximately 0.68–1.44-fold high signifying the correlation of ULK-1 with cancer progression. Additionally, the upregulated expression of ULK-1 is inversely correlated with the overall less survival time in patients. Silencing of ULK-1 by gene-specific mRNA induces cytotoxicity and triggers apoptosis in ESCC cell lines (Jiang et al. 2011). These results further point out the protective role of autophagy in cancer cells and the response of tumors against cancer therapy. The upregulation of ULK-1 expression is also found in hepatocellular carcinoma (HCC) patients, and it is significantly associated with tumor size and progression. Patients with low ULK-1 expression have longer survival time than those with high ULK-1 expression (Xu et al. 2013). One of the reasons behind increase in transcription of ULK-1 in solid tumors is the creation of hypoxia. Exposure of cells to hypoxic conditions induces unfolded protein response (UFR) and HIF-1 activation that triggers ULK-1 mRNA transcription (Schaaf et al. 2013). UFR or ER stress activates ATF4 or activating transcription factor 4 which directly binds to the promoter region of ULK-1 DNA and increases its

transcription (Pike et al. 2013). Upregulated ULK-1 induces autophagy which promotes growth and survival of tumor cells during hypoxia. Ablation of ULK-1 or ATF4 in epidermoid carcinoma cell line A431 and breast cancer cell line MCF-7 suppresses autophagy and reduces clonogenic survival of cells, decreases cellular ATP levels, increases cell apoptosis, and reduces spheroid growth (Pike et al. 2013). Loss of function of autophagy protein, FIP200, has been found in many cancers. FIP200 is involved in various cellular process including cell survival, cell growth, cell proliferation, embryonic development, metastasis, and differentiation (Gan and Guan 2008). FIP200 is located on 8q11 chromosome started from 53,535,016 bp from pter to 53,658,403 bp from pter and comprises of 123,388 bases. This region contains several loci of presumptive tumor suppressor genes, and heterozygosity of this region has been linked with various tumor types. Loss of function of this region is present in prostate cancer (Perinchery et al. 1999), breast cancer (Dahiya et al. 1998), colorectal cancer (Staub et al. 2006), hepatocellular carcinoma (Katoh et al. 2005), and ovarian cancer (Dimova et al. 2009). FIP200 deletion in mammary epithelial cells suppresses breast cancer initiation, progression, and metastasis (Wei et al. 2011). Conditional knockout of FIP200 gene *in* mouse model of breast cancer decreases tumor burden and increases overall survival time as compared to control mice containing functional FIP200 gene (Wei et al. 2011). Furthermore, these mice show fewer metastatic nodules as compared to control mice. Deletion of FIP200 causes autophagy defects in MMTV-PyMT transgenic mice (conditional knockout mice or CKO) like accumulation of large ubiquitin-positive or p62-positive aggregates, deformed mitochondria, and deficient LC3 accumulation. These mice show reduced cell proliferation, cell cycle arrest, decreased anchorage-independent growth in soft agar, and glycolysis as compared to control mice. Additionally, FIP200 deletion in Ras-transformed primary mouse embryonic fibroblasts inhibits their proliferation, cell cycle progression, glucose uptake, lactate formation, and anchorage-independent growth in soft agar (Wei et al. 2011). These tumors also show defective autophagy and increased expression of several chemokines including CXCL9 and CXCL10 that initiates increased immune surveillance (Wei et al. 2011). Downregulation of FIP200 by using small interfering RNA triggers apoptotic induction in human glioblastoma cells, immortalized human astrocytes, and primary human brain MvEC. FIP200 directly interacts and inhibits proline-rich tyrosine kinase 2 (Pyk2) expression in these cells and abrogates Pyk2-mediated regulation of calcium ion channels and activation of MEK-ERK signaling (Wang et al. 2011). Another important tumor promoter autophagy gene is lysosomal-associated membrane protein 1 or LAMP-1. It is located on the surface of lysosomes and endosomes and assists lysosomal and autophagosome fusion (Eskelinen 2006). About one third of ovarian serous adenocarcinoma tumors show LAMP-1 over expression in their cytoplasm. Expression analysis of LAMP-1 protein in normal ovarian tissue and ovarian adenocarcinomas of stages IIb, III, and IV reveals high expression of LAMP-1 as compared to normal ovarian tissue (Marzinke et al. 2013). Immunohistochemistry of these tumor lysates shows the presence of LAMP-1 in epithelial cell cytoplasm and few on the surface of plasma membrane. Furthermore, about 73 percent of LAMP-positive adenocarcinomas are positively stained for epidermal growth factor receptor (EFGR) and exhibit moderate to strong EGFR

signaling. Co-incubation of normal OV90 epithelial ovarian cancer cells with ovarian cancer ascites increases LAMP-1 expression and promotes cancer cell migration and proliferation suggesting the role of LAMP-1 in tumor progression (Marzinke et al. 2013). LAMP-1 and LAMP-2 overexpression and its correlation with tumor prognosis are also found in pancreatic carcinoma patients (Kunzli et al. 2002). LAMP-1 and LAMP-2 mRNA is moderately present in normal islet cells and weekly in most acinar cells, while it is highly upregulated in pancreatic ductal carcinoma cells. The upregulated expression of LAMP-1 is associated with tumor survival, and pancreatic carcinoma patients exhibiting high levels of LAMP-1 have significantly higher postoperative tumor survival period than patients whose tumors have low to moderate LAMP-1 mRNA levels (Kunzli et al. 2002). The role of other autophagy genes in tumor suppression and promotion is summarized in Table 1.1.

Table 1.1 Role of autophagy-associated genes in tumor suppression and promotion

Gene name	Functions	Role in cancer
ATG3	Catalyzes the conjugation of LC3B and phosphatidylethanolamine (PE)	Overexpression of ATG3 inhibits cell growth and promotes apoptosis in leukemic SKM-1 cells (Wang et al. 2014a)
		Inhibition of starvation-induced autophagy by ATG3 silencing suppresses epithelial-mesenchymal transition (EMT) and abrogates invasiveness in HCC cell lines HepG2 and BEL7402 (Li et al. 2013)
ATG14	Compete with UVRAG for Beclin binding and ATG14-Beclin complex is required for their localization to autophagosomes	Silencing of ATG14 sensitizes osteosarcoma cells to cisplatin-induced apoptosis (Zhao et al. 2014)
ATG16	ATG16 form complex with ATG5-ATG12 conjugate and form ATG16-ATG5-ATG12 complex which is required for LC3 lipidation and autophagosome formation	Hypermethylation of Atg16L is associated with poor prognosis in CLL against imatinib treatment (Dunwell et al. 2010)
Bif-1	Form complex with Beclin and enhances Beclin-VPS34 activity during autophagy	Bif-1$^{-/-}$ mice are more prone to tumorigenesis as compared to wild type (Takahashi et al. 2007)
		Reduced expression of Bif-1 is associated with short survival period in CRC stage I and II patients (Ko et al. 2013)
		Bif haploinsufficiency suppresses mitophagy, induces chromosomal damage, and inhibits apoptosis (Takahashi et al. 2013)
AMBRA-1 (activating molecule in BECN1-regulated autophagy protein-1)	AMBRA-1 binds to Beclin and potentiates the lipid kinase activity of Beclin/VS34 complex during autophagy	Starvation-induced autophagy is associated with AMBRA-1 upregulation in colorectal cancer cells, and its genetic silencing enhances apoptotic effects of etoposide and staurosporine colorectal cancer cell lines (Gu et al. 2014)

Table 1.1 (continued)

Gene name	Functions	Role in cancer
		AMBRA-1 is overexpressed in cholangiocarcinoma patients and associated with poor prognosis. AMBRA-1 increases Snail expression in cholangiocarcinoma cell lines and promotes EMT (Nitta et al. 2014)
P62	P62 binds with ubiquitinated protein aggregates and promote their degradation by autophagy pathway via interacting with LC3B	P62 accumulation is correlated with lymph node metastasis in NSCLC adenocarcinoma patients (Inoue et al. 2012)
		P62 is overexpressed in cisplatin-resistant SKOV3/DDP (ovarian cancer) cells and knockdown of p62 sensitizes cells to cisplatin treatment. P62 activates Keap1-Nrf2-ARE pathway and induces antioxidant gene expression in SKOV3/DDP cells (Xia et al. 2014)
		P62 is overexpressed in breast tumors (Thompson et al. 2003)
Etoposide-induced protein 2.4 homologue (EI24)	Autophagosome formation and clearance of protein aggregates	Mutated in aggressive breast cancers and function as tumor suppressor (Zhao et al. 2005)
		Inhibition of EI24/PIG8 in fibroblasts and breast cancer cells abrogates the apoptotic effect of etoposide (Mork et al. 2007)
Immunity-related GTPase family M (IRGM)	IRGM involves in autophagy-mediated immunity against pathogens. It interacts with ATG5, ATG10, SH3GLB1, and LC3 and promotes phagosome maturation (Petkova et al. 2012; Singh et al. 2006)	IRGM rs4958847 polymorphism is associated with susceptibility to gastric cancer. Carriers of the rs4958847 A allele is protected against gastric cancer development (Burada et al. 2012). IRGM genetic polymorphism, rs13361189TC and polymorphic rs13361189CC genotype, increases in glioma patients as compared to healthy individuals and is associated with increased expression of IFN-γ and IL-4 which play an important role in glioma development (Ge et al. 2014)

1.6 MicroRNAs and Autophagy: Role in Cancer

MicroRNAs (miRNAs) are small noncoding RNA molecules of about 20–25 nucleotides in length distributed throughout the animal and plant kingdom. miRNAs regulate gene expression at posttranscriptional and translational levels, and dysregulation in this process is linked with various diseases including cancer. The first direct link between miRNA and cancer was observed in chronic lymphocytic leukemia

where a deletion in a small region of 13q14 chromosome leads to the downregulation of two miRNAs, miR15 and miR16. This deletion is present in 65% of CLL and correlated with the pathogenesis of the disease (Calin et al. 2002). Aberrant changes in miRNA expression have been detected in many types of tumors, and its role in disease prognosis and pathogenesis is continuously explored (Price and Chen 2014). The abnormally expressed miRNAs target the transcripts of various tumor suppressor or promoter genes in cancer. Somatic mutations associated with miRNA processing have been found in many cancers. MiRNA profile of tumor cells is different from normal cells, and they can be used in cancer diagnostics (Paranjape et al. 2009). Many autophagy genes are regulated by miRNA, and any dysregulation in this process may lead to defective autophagy (Frankel and Lund 2012). Since autophagy defects are common in cancer, its regulation by miRNAs is an important area of interest that may provide a potential therapeutic advantage in cancer treatment. The first report about miRNA regulation of autophagy was published by Zhu et al. in 2009, where they demonstrated that miR-30a directly binds to 3′-UTR of Beclin and negatively regulates its expression. Inhibition of autophagy by overexpression of miR-30a sensitizes tumor cells to chemotherapy-induced apoptosis. Chemotherapeutic drug taxol induces protective autophagy by decreasing miR-30a levels in cancer cells. Overexpression of miR-30a represses beclin-dependent autophagy and enhances apoptosis in cells (Zou et al. 2012). MiR-30d regulates many autophagy genes including BECN1, BNIP3L, ATG12, ATG5, and ATG2 in cells and found frequently mutated in human epithelial cancers. MiR-30d binds to 3′-UTR sequences of these genes and inhibits their expression at transcriptional and translational levels (Zhang et al. 2014). Inhibition of beclin by miR-30d suppresses autophagy and sensitizes human anaplastic thyroid carcinoma (ATC) cells to cisplatin treatment both *in vitro* and *in vivo* (Zhang et al. 2014). Furthermore, miRNA-induced autophagy promotes radioresistance in cancer cells. Overexpression of miR-23b in pancreatic tumors undergoing radiation therapy decreases ATG12 levels and autophagy which sensitizes these cells to treatment (Wang et al. 2013). Additionally, dysregulation of miR-130a increases protective autophagic flux in cancer cells (Kovaleva et al. 2012). Ectopic expression of miR-130a inhibits cell proliferation and induces apoptosis in chronic lymphocytic leukemia (CLL) cell lines (Kovaleva et al. 2012). MiR-130a has also been shown to inhibit autophagy in CLL cells by downregulating the transcription of autophagy genes DICER1 and ATG2B. It has been observed that CLL patients nonresponsive to chemo treatment show reduced miR-130a expression and enhanced autophagy (Kovaleva et al. 2012). Another miRNA, miR-101, inhibits autophagy and sensitizes breast cancer cells to tamoxifen treatment (Frankel et al. 2011). MiR-101 downregulates the activation of STMN1 (Stathmin/Oncoprotein 18), RAB5A, and ATG4D (Frankel et al. 2011) genes in breast cancer cells. They are important regulators of autophagy and their expression is highly upregulated in many tumors (Marklund et al. 1996; Rana et al. 2008). All three genes have miR-101-binding motifs in their 3′UTRs, and point mutations in this region significantly extirpate their miR-101-dependent downregulation. Furthermore, cotreatment of miR-101 with chemotherapeutic drug 4 hydroxytamoxifen significantly reduces autophagy and cell survival in breast cancer

cell lines (Frankel et al. 2011). Additionally, miR-22 suppresses proliferation in osteosarcoma cells by targeting high-mobility group box 1 (HMGB1) protein-mediated autophagy. Osteosarcoma cells show high HMGB1 expression after anti-cancer therapy and induce protective autophagy which contributed to chemoresistance (Li et al. 2014). Since the role of autophagy in cancer cells is controversial, its modulation by micro-RNA also has different impacts in cancer cells. There are many other examples where inhibition of autophagy by miRNA prevents autophagic cell death in cancer cells and hence promotes their survival. Inhibition of autophagy by miR-25 prevents autophagic cell death in breast cancer cells. MiR-25 directly binds to the 3′UTR region of autophagy genes ATG14 and ULK-1 and inhibits their transcription (Wang et al. 2014b). Furthermore, miR290–295 cluster prevents melanoma cancer cells from starvation-induced autophagic cell death. miR290–295 inhibits autophagy induction by downregulating several essential autophagy genes involved in the formation of class III PI3 kinase complex, ATG12 and ATG8 conjugation systems, and ULK1/ATG1 complex. Overexpression of miR290–295 in B16F1 melanoma cells undergoing glucose deprivation prevents persistent autophagy and hence promotes survival (Chen et al. 2012). Additionally, miR-17 prevents autophagy induction in glioma cells treated with temozolomide and promotes their survival. MiR-17 directly binds to 3′UTR region of ATG7 and prevents ATG7-mediated autophagy induction. Inhibition of miR-17 by using anti-miR-17 administration in glioma cells enhanced temozolomide-induced cytotoxicity and radiosensitivity (Comincini et al. 2013) as described in Fig. 1.2.

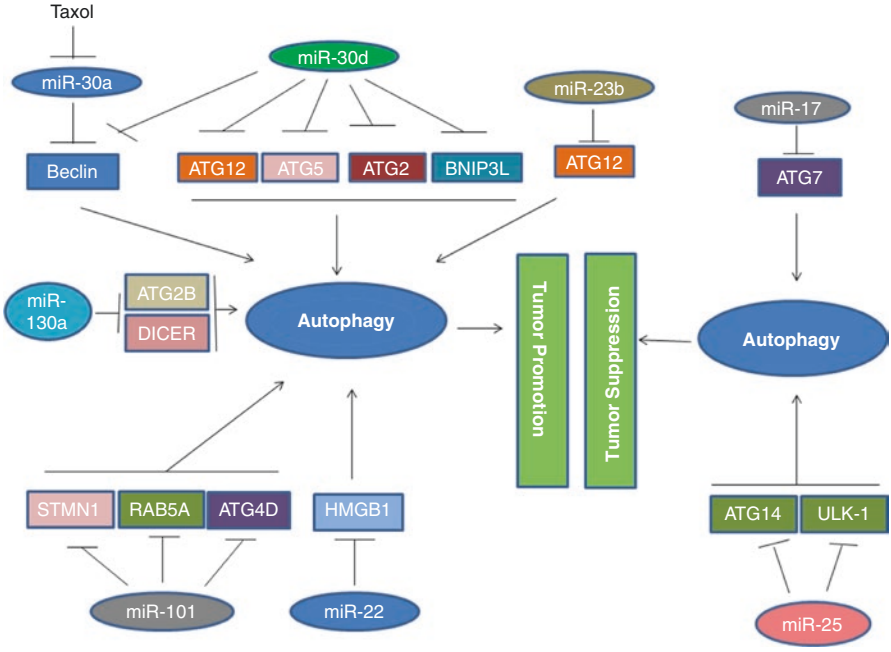

Fig. 1.2 Regulation of autophagy by miRNAs and its role in cancer

1.7 Epigenetic Regulation of Autophagy in Cancer

Epigenetic regulation of DNA is an important process in cells and plays a critical role in replication, transcription, repair, and development. Similar to genetic changes, epigenetic defects play an important role in the genesis of cancer. Almost all cancers show epigenetic defects, and along with genetic alterations, they are fully involved in initiation and progression of the disease. There are two primary epigenetic mechanisms that occur in the cells. One is DNA methylation and another is the covalent modifications of the histones. These modifications are performed by chromatin-modifying enzymes in a highly regulated manner. DNA methylation process involves the methylation of CpG islands that are mostly found in or near the promoter region of mammalian genes or in the region of repetitive DNA sequences in chromosome-like centromere, retrotransposons, etc. CpG islands are short DNA stretches of about 1000 bp that are rich in CpG dinucleotides. In normal cells, they usually demethylate in the regions near promoter sequences and heavily methylated in other parts of the genome in order to prevent chromosome instability. DNA methylation is catalyzed by a family of enzymes called DNA methyltransferases that transfer methyl group from donor S-adenosylmethionine (SAM) to carbon-5 of the cytosine residues (5mC) in CpG dinucleotides. Hypermethylation of CpG dinucleotides present within the promoter regions of genes leads to their silencing by preventing the binding of transcription factors to the DNA. Silencing of genes involved in tumor suppression by DNA hypermethylation promotes uncontrolled cell division, leading to tumorigenesis (Baylin 2005). Furthermore, the hypomethylation of CpG dinucleotides present in other parts of the genome led to the reactivation of suppressed elements that cause genomic instability. The second mechanism of epigenetic regulation involves histone modifications like acetylation, methylation, ADP ribosylation, SUMOylation, phosphorylation, ubiquitination, O-GlcNAcylation, etc. of histone residues. These modifications regulate chromatin condensation and DNA accessibility in cells. However, unlike DNA methylation, different types of histone modifications have different impacts on the transcription of histone DNA. For example, acetylation of ε-amino group of lysine side chains of histones catalyzed by histone acetyltransferases (HAT) neutralizes the lysine-positive charge and weakens the interactions between histones and DNA, thus making DNA accessible for transcription, whereas methylation of histones can lead to transcriptional repression or activation of DNA depending upon the mode of methylation. Dysregulation in acetylation and deacetylation process are most common posttranslational histone modifications found in human cancers. Unlike genetic changes, epigenetic abnormalities are reversible in nature, and the agents that restore this balance are of great interest in cancer therapy. The most validated targets in cancer epigenetics are histone deacetylases (HDACs) and DNA methyltransferases. Although structural mutations in HDACs are rare in cancer, their elevated expression has been associated with many tumor types. HDAC1 activity is found upregulated in various cancers like, gastric, stomach, esophagus, colon, prostate, breast, ovary, lung, pancreatic, and thyroid cancer, therefore, associated with the poor prognosis of the disease (Nakagawa et al. 2007; Choi et al. 2001). Similarly, aberrant expression of other HDACs is also found

in many tumor types. HDAC2 is upregulated in cervical (Huang et al. 2005), gastric (Song et al. 2005), colorectal (Zhu et al. 2004), and prostate cancer (Yin and Fu 2013), HDAC3 in colorectal (Spurling et al. 2008) and prostate cancer (Weichert et al. 2008), HDAC4 in breast cancer (Ozdag et al. 2006), HDAC5 and HDAC7 in colorectal cancer (Ozdag et al. 2006; Zhu et al. 2011), and HDAC6 in breast and oral squamous cell carcinoma (Zhang et al. 2004; Sakuma et al. 2006). The overexpression of these HDACs directly contributes to tumorigenicity and is associated with the poor prognosis of the disease. Most of HDACs are involved in the regulation of autophagy in cells. HDAC1 and HDAC2 are involved in autophagosome formation and regulate autophagy in skeletal muscles (Moresi et al. 2012). Deletion of HDAC1 and HDAC2 blocks autophagic flux and causes progressive myopathy in mice (Moresi et al. 2012). HDAC1 and HDAC2 double-knockout mice showed impaired autophagosome formation, mitochondrial abnormalities, and accumulation of p62. Conversely, ectopic expression of both alleles restores these functions in mice and promotes their survival (Moresi et al. 2012). HDAC6 is required for fusion of autophagosomes to lysosomes (Lee et al. 2010). Double-knockout HDAC6 mice show enhanced accumulation of autophagosomes and defective autophagy (Lee et al. 2010). HDAC6-mediated autophagy involves F-actin cytoskeleton structures that form specialized F-actin network over vesicles and lysosomes that undergo fusion (Gao et al. 2007; Jahraus et al. 2001). F-actin polymers are also present on protein aggregates that undergo autophagy-dependent degradation (Lee et al. 2010). The assembly of F-actin polymers over protein aggregates is facilitated by HDAC6 substrate cortactin. Knockdown of cortactin by siRNA prevents the formation of these polymers over protein aggregates and inhibits their degradation. However, cortactin silencing has little impact on the distribution of F-actin polymers on lysosomes (Lee et al. 2010). Furthermore, HDAC6 deacetylates LC3B and increases its transcription during serum starvation in cervical cancer cell line HeLa (Liu et al. 2013b). HDAC6 expression is lower in HCC patients and is associated with the poor prognosis of the disease (Jung et al. 2012). The recurrence-free survival rate of HCC patients having low HDAC6 expression is lower than those with high HDAC6 activity. Ectopic expression of HDAC6 induces autophagic cell death in HCC cell lines. The mechanism of autophagy induction by HDAC6 in cells involves the activation of Beclin via c-JunNH2-terminal kinase pathway. Inhibition of autophagy with autophagy inhibitor 3-MA or with SP600125, a JNK-specific inhibitor, effectively blocks HDAC6-induced cell death (Jung et al. 2012). Another member of histone deacetylases, SIRT-1 (Sirtuin family of proteins, an NAD-dependent deacetylases), induces autophagy in normal prostate cells and is required for prostate gland development and maintenance (Powell et al. 2011). SIRT-1 regulates the expression of late autophagy proteins ATG4, ATG7, and ATG8 by promoting their deacetylation. Deletion of SIRT-1 causes deficient autophagy and abnormal prostate development that leads to prostate intraepithelial neoplasia (Powell et al. 2011). Furthermore, $SIRT^{-/-}$ embryonic stem cells (ESCs) of human and mouse origin show lower expression of Beclin and LC3B and enhanced phosphorylation of mTOR substrates P70/85-S6 kinase and ribosomal-S6. Treatment of these cells with H_2O_2 shows enhanced mitochondrial membrane potential loss and apoptosis as compared to

wild-type ESCs reflecting the protective role of autophagy in ESCs during oxidative stress (Ou et al. 2014). Contrary to SIRT-1, SIRT-2 inhibits the basal levels of autophagy and abrogates the chemoresistance developed against microtubule inhibitors due to prolonged mitotic arrest (Inoue et al. 2014). Knockdown of SIRT-2 increases the basal levels of autophagy and triggers mitotic arrest for longer periods, which confer resistance to microtubule inhibitors. Such type of resistance is also seen in rapamycin-treated or mild-starved cells in the presence of microtubule inhibitors, which delays post-slippage death. Silencing of autophagy genes ATG5-ATG7 or Fox01 (forkhead O family protein) or HDAC6 abolishes such resistance and triggers cell death. HDAC6 forms complex with cytoplasmic Fox01 and inhibits its activation. Disruption of HDAC6-Fox01 complex by stimuli-like stress releases Fox01 and promotes its acetylation. Acetylated Fox01 binds to ATG7 and induces autophagy in cells (Inoue et al. 2014). Furthermore, HDAC counterparts, histone acetyltransferases (HATs), regulate the expression of many proteins via autophagy, and dysregulation in this process promotes tumor. One example of such regulation is the acetylation of glycolytic enzyme pyruvate kinase M2 (PKM2) on lysine 305 residue which decreases its activity and promotes degradation via chaperone-mediated autophagy (Lv et al. 2011). PKM2 is highly expressed in cancer cells and is involved in metabolic reprogramming which switch cells from oxidative phosphorylation to aerobic glycolysis and carcinogenesis (Christofk et al. 2008). PKM2 acetylation is performed by p300 (E1A-binding protein, 300 kDa) acetyltransferases, and it facilitates its binding with HSC70 (heat shock protein), a chaperone which recruits target proteins to lysosomes for chaperone-mediated autophagic degradation (Lv et al. 2011). PKM2 lysine 305 mutants accumulate glycolytic intermediates and show enhanced cell proliferation and tumor growth as compared to their wild-type counterparts. Hence, autophagic degradation of PKM2 through acetylation regulates cell cycle control and prevents tumorigenicity in cells (Lv et al. 2011). Additionally, the turnover of another acetyltransferase, hMOF, is controlled by autophagy in many cancer cells. hMOF along with other acetyltransferases KAT8 and MYST1 catalyzed the acetylation of lysine 16 residue of histone H4 (H4K16ac) which influences chromatin structure and transcription (Fullgrabe et al. 2013). Autophagy inhibits the acetylation of H4 in various cancer cell lines U1810, HeLa, and U2OS. Additionally, starvation or rapamycin treatment inhibits the acetylation of H4K16 in MEF cells. Cells deficient in ATG1, ATG5, or ATG7 show less inhibition on H4K16ac histone modifications after treatment with autophagy inducing stimuli-like starvation or treatment with rapamycin or torin (mTOR inhibitor) as compared to wild types. Treatment with autophagy inhibitors CQ or 3-MA abrogates autophagy-mediated inhibition of hMOF and induces H4K16 acetylation. Additionally, deacetylation of H4K16ac promotes cell death in cancer cell lines HeLa or U1810, which is prevented by autophagy inhibitors (Fullgrabe et al. 2013). Furthermore, autophagy is involved in the degradation of cytoplasmic chromatin fragments during cellular senescence. Senescent cells have chromatin fragments associated with γ-H2AX and H3K27me3 histone in their cytoplasm. Autophagic process degrades these fragments and contributes to the stability of senescence which may play a role in tumor suppression (Ivanov et al. 2013). During serum starvation, glycogen synthase kinase 3 (GSK3)

phosphorylates and activates acetyltransferase KAT5/TIP60 in cells, which in turn acetylates and activates ULK-1 at lysine 162 and lysine 606 residues and induces protective autophagy (Lin et al. 2012). GSK3 inhibition abrogates KAT5/TIP60 activation in serum-deprived cells. Additionally, cells with mutant KAT5/TIP60 that cannot be phosphorylated by GSK3 are resistant to serum starvation induced autophagy. Cells that have acetylation-defective ULK-1 fail to rescue autophagy in ULK-1-silenced MEF cells indicating that acetylation is a necessary event in ULK-1 activation (Lin Sy 2012). The loss of TIP60 leads to an accumulation of double-stranded breaks and induces genomic instability in cells (Chailleux et al. 2010; Murr et al. 2006). TIP60 mutations are found in many cancers including head and neck squamous cell carcinoma, breast cancer, colorectal cancer, and lymphoma (Sakuraba et al. 2009; Gorrini et al. 2007).

The other major epigenetic alteration found in tumor cells is DNA methylation (Kulis and Esteller 2010). DNA hypermethylation of certain tumor suppressor genes silenced their activation, whereas hypomethylation on the heterochromatin regions of DNA induces genomic instability and promotes tumorigenesis (Kulis and Esteller 2010). DNA methylation controls the expression of many autophagy genes in various tumor types. DNA methyltransferase G9a controls the transcription of key autophagy genes involved in autophagosome formation and modulates autophagy under normal growth conditions. Methyltransferase G9a is ubiquitously expressed in somatic cells and mainly localized in the euchromatin region of DNA. It is expressed in a variety of tumors including leukemia (Lehnertz et al. 2014), prostate cancer (Kondo et al. 2008), lung cancer (Chen et al. 2010), and neuroblastoma (Ke et al. 2014). G9a expression is higher in stage 4 neuroblastoma tumors as compared to stage 3 or 4S and is associated with tumor-related deaths (Ke et al. 2014). DNA methyltransferase G9a forms complex with another methyl transferase G9a-like protein (GLP) and promotes the methylation of CpG islands in the promoter region of autophagy genes which repress their activation (Artal-Martinez de Narvajas et al. 2013). Under normal conditions, G9a associates with the promoters of LC3B, WIPIs (WD repeat domain phosphoinositide-interacting propeller proteins), and DOR (LC3B-interacting protein diabetes and obesity regulated) and epigenetically represses them. Silencing of G9a by siRNA in cervical cancer cell line HELA induces LC3B accumulation which promotes autophagy (Artal-Martinez de Narvajas et al. 2013). Pharmacological or genetic inhibition of G9a inhibits cell proliferation, decreases tumorigenicity, and induces autophagy in neuroblastoma cells (Ke et al. 2014). Furthermore, hypermethylation of autophagy gene ATG16L is found in 69% of CP-CML (chronic phase chronic myeloid leukemia) patients. Patients having methylated ATG16L show significantly decreased major molecular response rate at 12 and 18 months of imatinib treatment in comparison with patients with unmethylated ATG16L gene (Dunwell et al. 2010). Additionally, hypermethylation and inactivation of the promoter region of tumor suppressor gene ARH1 (Aplasia Ras homologue member I, also known as DIRAS3) are found in ovarian cancer patients (Feng et al. 2008). ARH1 inhibits cell proliferation and motility by targeting PI3 kinase/Akt pathway and induces autophagy in cells via ATG4 upregulation. Serum starvation or mTOR treatment inhibits cell

proliferation in ovarian cancer cells via ARH1 activation which also induces protective autophagy in cells. Inhibition of autophagy with CQ enhanced cytotoxicity and reduces regrowth of xenografted tumors (Lu et al. 2014).

Another important tumor suppressor and autophagy-regulating protein, PCDH1, is frequently methylated and silenced in almost all gastric and colorectal tumor cell lines as well as in 95% of primary tumors but not in normal gastric or colorectal tissues. PCDH17 deletion is found in only 18% of gastric and 12% of colorectal cancer tissues signifying the importance of its epigenetic silencing in cancer (Hu et al. 2013). Furthermore, ectopic expression of PCDH17 inhibits tumor growth in these cancer cell lines *in vitro* and *in vivo* and promotes apoptosis and autophagic cell death (Hu et al. 2013).

1.8 Epigenetic Modifiers in Cancer and Role of Autophagy

Several drugs targeting epigenetic regulators are already approved by FDA for cancer therapy, and many are under preclinical investigation. Almost all anticancer agents that target such epigenetic defects induce either apoptotic or autophagic cell death in cancer cells. However, contrary to apoptosis, autophagy induction by these inhibitors has dual functions, and thus modulating this process is of great interest in cancer therapy. The most common HDAC inhibitor, suberoylanilide hydroxamic acid (SAHA), which is used in the treatment of cutaneous T cell lymphoma, induces robust autophagy in many cancer cells. SAHA induces caspase-independent autophagic cell death in endometrial stromal cells, chondrosarcoma cells, and tamoxifen-resistant breast cancer cells (Banreti et al. 2013; Lee et al. 2012). Another HDAC inhibitor, valproic acid, induces autophagic cell death and not apoptosis in glioma cells (Dong et al. 2013). The mechanism of autophagy induction involves the generation of oxidative stress after valproic acid treatment which further activates ERK pathway. The blockage of ERK signaling inhibits autophagy and induces apoptosis in cells (Fu et al. 2010). Additionally, the combination of valproic acid with temsirolimus inhibits tumor cell growth and triggers autophagic cell death in Burkitt leukemia/lymphoma cell lines (Dong et al. 2013). On the other hand, autophagy induced by SAHA in acute myeloid leukemia (AML) cell lines promotes chemoresistance, and inhibition of this process reduces the cell viability and colony-forming ability in AML cells (Torgersen et al. 2013). Additionally, genetic or pharmacological inhibition of autophagy-enhanced SAHA induced apoptosis in glioblastoma cells (Chiao et al. 2013). Co-treatment of HDAC inhibitor vorinostat with another anticancer drug sorafenib induces protective autophagy in hepatocellular carcinoma cells, and its inhibition by using 3-MA or beclin silencing enhances their synergistic effects (Yuan et al. 2014). Combination of pan-HDAC inhibitor, panobinostat, with CQ enhances its antitumor effects against human estrogen/progesterone receptor and HER2 (triple)-negative breast cancer (TNBC) cells. CQ co-treatment with panobinostat results in reduced tumor burden and higher survival rate in MD-AMB-231 breast cancer xenografts (Rao et al. 2012). Some of the examples of the autophagy induction during epigenetic targeting in cancer are summarized in Table 1.2.

Table 1.2 Epigenetic targets and the role of autophagy in cancer therapy

Anticancer agents	Specificity	Role of autophagy in chemo treatment
Trichostatin	HDAC1 and HDAC2	Autophagy induces cytoprotection in neuroblastoma cell lines against trichostatin treatment (Francisco et al. 2012)
		Inhibition of autophagy enhances trichostatin-induced apoptosis and radiosensitivity in colon cancer cells (He et al. 2014)
Romidepsin	HDAC1 and HDAC2	Inhibition of autophagy with CQ enhances romidepsin-induced cell death in malignant rhabdoid tumors (Watanabe et al. 2009)
Apicidin	HDAC2 and HDAC3	Apicidin induces apoptosis and autophagy in human oral squamous carcinoma cells, and inhibition of autophagy enhances apicidin-mediated apoptosis (Ahn et al. 2011)
Abexinostat (PCI-24781)	HDAC1 and HDAC2	Inhibition of HIF-1α attenuates abexinostat-induced autophagy in B-cell lymphoma cells and decreases cell survival (Bhalla et al. 2013)
		Autophagy blockage sensitizes resistant malignant peripheral nerve sheath tumors to abexinostat-induced apoptosis (Lopez et al. 2011)
Sirtinol	HDAC3	Sirtinol induces autophagic cell death in MCF-7 cells by downregulating Sirt1/2 (Wang et al. 2012a)
Resveratrol	HDAC3	Induces autophagic cell death in lung cancer cell line A549 which is rescued by genetic or pharmacological inhibition of autophagy (Zhang et al. 2013)
		Induces autophagic cell death in chronic myelogenous leukemia cells by activating JNK pathway (Puissant et al. 2010)
MGCD0103 (Mocetinostat)	HDAC1, HDAC2, and HDAC11	Induction of apoptosis and inhibition of autophagy mediate the therapeutic effect of MGCD0103 in B-cell chronic lymphocytic leukemia (El-Khoury et al. 2014)
GSK343	Histone methyltransferase EZH2	Induces autophagic cell death in MDA-MB-231, HepG2, and A549 cells (Liu et al. 2014)
BIX-01294 (BIX)	Euchromatic histone lysine N-methyltransferase 2 (EHMT2)	Induces autophagic cell death in estrogen receptor (ESR)-negative SKBr3, ESR-positive MCF-7, and HCT116 colon cancer cells (Kim et al. 2013)

1.9 Posttranslational Modifications in Autophagy Genes: Impact in Cancer

Posttranslational modifications are important part of gene regulation and have been associated with early and late stages of cancer. The important posttranslational modifications in cells are phosphorylation, acetylation, methylation, ubiquitination, farnesylation, glycosylation, and sialylation. The expression of many autophagy genes is controlled by these posttranslational modifications, and their dysregulation promotes tumorigenesis. The main posttranslational modifications linked to autophagy proteins are phosphorylation, ubiquitination, and SUMOylation (McEwan and Dikic 2011). Autophagy is regulated by a wide variety of kinases involved in PI3K-Akt-mTOR pathways (Jaber and Zong 2013; Wang et al. 2012b), MAPK-ERK pathway (Wang et al. 2009), Wnt signaling (Petherick et al. 2013), JAK-STATs (Jonchere et al. 2013), FAK (focal adhesion kinases) pathways (Tuloup-Minguez et al. 2011), and TGF-β (Kiyono et al. 2009). Phosphorylation of Beclin at multiple tyrosine residues by the epidermal growth factor receptor (EFGR) tyrosine kinase enhances its binding with negative regulators of autophagy such as Bcl2 and RUBICON and prevents the formation of Beclin type III PI3 kinase complex (Wei et al. 2013). The inhibition of EGFR by tyrosine kinase inhibitors induces autophagy in non-small cell lung cancer (NSCLC) cell lines which induces cytotoxicity. The overexpression of EFGR enhances Beclin inactivation and promotes tumor growth, chemoresistance, and metastasis in NSCLC (Wei et al. 2013). Phosphorylation and inactivation of ULK-1 and ATG13 by mTOR during nutrient-rich conditions inhibit autophagy in cells (Jung et al. 2009). The second common posttranslational modification, ubiquitination, is involved in the selective targeting of proteins to autophagosomes which promotes their degradation upon autophagosome lysosome fusion. Ubiquitin-binding proteins are selected by p62 which directly binds to LC3B via its LC3-interacting binding motif, bringing these proteins to the autophagosomes. Another mechanism of autophagic degradation of ubiquitinated proteins involves HDAC6, which directly interacts with polyubiquitinated proteins and facilitates their interaction with MTOC (microtubule-organizing center) and forms aggresomes that are cleared by autophagy (Shaid et al. 2013). Ubiquitin-dependent autophagic degradation is an essential process in cancer cells to removes large amount of misfolded proteins. Inhibiting cancer cell growth by using proteasomal inhibitors that disrupt this mechanism is an important area of research in cancer therapy. The third important posttranslational modification in autophagy proteins is SUMOylation. SUMOylation is a proteasomal degradation process which involves the attachment of a small ubiquitin-like modifier to target proteins similar to ubiquitination. SUMOylation is involved in various cellular processes such as protein stability, nuclear transport, cell cycle progression, and transcriptional regulation (Geiss-Friedlander and Melchior 2007). SUMOylation of type III PI3 kinase, VPS34, increases the activity of Beclin-VPS34 complex and enhances autophagy during stress (Yang et al. 2013). Autophagy-inducing stress (starvation or treatment with HDAC inhibitor panobinostat) triggers the acetylation of heat shock protein 70, which binds and recruits

SUMO E3 ligase KAP1 to Beclin-VPS34 complex and promotes Vps34 SUMOylation at lysine 840 residue. Vps34 SUMOylation increases its lipid kinase activity and further promotes its association with Beclin in cells (Yang et al. 2013). Moreover, the knockdown of hsp70 abolished Vps34-beclin interactions and induces of autophagy in cells.

1.10 Perspective

Autophagy is a tightly regulated catabolic process which is involved in variety of physiological process including removal of damaged or abnormal cellular constituents, recycling of biomolecules for stress adaptation, host defense mechanism, cell death, and embryonic development. Autophagy process comprises wide protein network and interlinked cellular pathways. Defective autophagy is found in much human pathology including cancer, neurodegenerative diseases, and cardiovascular, metabolic, and infectious diseases. Defects in autophagy-linked genes are commonly found in cancer that leads to tumor progression and therapeutic resistance. Autophagy-associated genes act as both tumor suppressors and promoters, and their aberrant expression is associated with the dysregulation of cellular homeostasis and initiation of tumor growth. In addition to genetic and epigenetic defects, the role of microRNAs associated with autophagy is vital in cancer development. However, the mechanisms associated with such epigenetic alterations are largely unknown though our understanding in this area is continuously growing. Based on different genetic and epigenetic backgrounds, the influence of autophagy on the fate of cancer cells can vary between different cancers or among the cells of the same cancer. Keeping in view of the paradoxical role of autophagy in cancer development and progression, careful designing of the autophagy inhibitors or promoters can increase the therapeutic efficacy of anticancer agents. Thus, an increase understanding with respect to the functional aspects of autophagy is important in various cancers in order to exploit it for therapeutic advantage.

Competing Interests No competing interests to be disclosed.

References

Ahn MY, Ahn SG, Yoon JH. Apicidin, a histone deaceylase inhibitor, induces both apoptosis and autophagy in human oral squamous carcinoma cells. Oral Oncol. 2011;47(11):1032–8. doi:10.1016/j.oraloncology.2011.07.027.

Aita VM, et al. Cloning and genomic organization of beclin 1, a candidate tumor suppressor gene on chromosome 17q21. Genomics. 1999;59(1):59–65. doi:10.1006/geno.1999.5851.

Alexander DD, Cushing CA. Red meat and colorectal cancer: a critical summary of prospective epidemiologic studies. Obes Rev. 2011;12(5):e472–93. doi:10.1111/j.1467-789X.2010.00785.x.

Artal-Martinez de Narvajas A, et al. Epigenetic regulation of autophagy by the methyltransferase G9a. Mol Cell Biol. 2013;33(20):3983–93. doi:10.1128/mcb.00813-13.

Banreti A, Sass M, Graba Y. The emerging role of acetylation in the regulation of autophagy. Autophagy. 2013;9(6):819–29. doi:10.4161/auto.23908.

Baylin SB. DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol 2005, 2 Suppl 1:S4–11. doi:10.1038/ncponc0354.

Bhalla S, Evens AM, Prachand S, Schumacker PT, Gordon LI. Paradoxical regulation of hypoxia inducible factor-1alpha (HIF-1alpha) by histone deacetylase inhibitor in diffuse large B-cell lymphoma. PLoS One. 2013;8(11):e81333. doi:10.1371/journal.pone.0081333.

Burada F, Plantinga TS, Ioana M, Rosentul D, Angelescu C, Joosten LA, Netea MG, Saftoiu A. IRGM gene polymorphisms and risk of gastric cancer. J Dig Dis. 2012;13(7):360–5. doi:10.1111/j.1751-2980.2012.00602.x.

Calin GA, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524–9. doi:10.1073/pnas.242606799.

Chailleux C, et al. Physical interaction between the histone acetyl transferase Tip60 and the DNA double-strand breaks sensor MRN complex. Biochem J. 2010;426(3):365–71. doi:10.1042/bj20091329.

Chen MW, et al. H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. Cancer Res. 2010;70(20):7830–40. doi:10.1158/0008-5472.can-10-0833.

Chen Y, Liersch R, Detmar M. The miR-290-295 cluster suppresses autophagic cell death of melanoma cells. Sci Rep. 2012;2:808. doi:10.1038/srep00808.

Chhipa RR, Wu Y, Ip C. AMPK-mediated autophagy is a survival mechanism in androgen-dependent prostate cancer cells subjected to androgen deprivation and hypoxia. Cell Signal. 2011;23(9):1466–72. doi:10.1016/j.cellsig.2011.04.008.

Chiao MT, Cheng WY, Yang YC, Shen CC, Ko JL. Suberoylanilide hydroxamic acid (SAHA) causes tumor growth slowdown and triggers autophagy in glioblastoma stem cells. Autophagy. 2013;9(10):1509–26. doi:10.4161/auto.25664.

Choi JH, Kwon HJ, Yoon BI, Kim JH, Han SU, Joo HJ, Kim DY. Expression profile of histone deacetylase 1 in gastric cancer tissues. Jpn J Cancer Res. 2001;92(12):1300–4.

Choi KS. Autophagy and cancer. Exp Mol Med. 2012;44(2):109–20. doi:10.3858/emm.2012.44.2.033.

Christofk HR, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature. 2008;452(7184):230–3. doi:10.1038/nature06734.

Comincini S, et al. microRNA-17 regulates the expression of ATG7 and modulates the autophagy process, improving the sensitivity to temozolomide and low-dose ionizing radiation treatments in human glioblastoma cells. Cancer Biol Ther. 2013;14(7):574–86. doi:10.4161/cbt.24597.

Cufi S, et al. Autophagy-related gene 12 (ATG12) is a novel determinant of primary resistance to HER2-targeted therapies: utility of transcriptome analysis of the autophagy interactome to guide breast cancer treatment. Oncotarget. 2012;3(12):1600–14.

Dahiya R, Perinchery G, Deng G, Lee C. Multiple sites of loss of heterozygosity on chromosome 8 in human breast cancer has differential correlation with clinical parameters. Int J Oncol. 1998;12(4):811–6.

Deretic V. Autophagy in infection. Curr Opin Cell Biol. 2010;22(2):252–62. doi:10.1016/j.ceb.2009.12.009.

Dimova I, et al. Genomic markers for ovarian cancer at chromosomes 1, 8 and 17 revealed by array CGH analysis. Tumori. 2009;95(3):357–66.

Dong LH, et al. Histone deacetylase inhibitor potentiated the ability of MTOR inhibitor to induce autophagic cell death in Burkitt leukemia/lymphoma. J Hematol Oncol. 2013;6:53. doi:10.1186/1756-8722-6-53.

Du J, Teng RJ, Guan T, Eis A, Kaul S, Konduri GG, Shi Y. Role of autophagy in angiogenesis in aortic endothelial cells. Am J Phys Cell Physiol. 2012;302(2):C383–91. doi:10.1152/ajpcell.00164.2011.

Dunwell T, et al. A genome-wide screen identifies frequently methylated genes in haematological and epithelial cancers. Mol Cancer. 2010;9:44. doi:10.1186/1476-4598-9-44.

El-Khoury V, et al. Disruption of autophagy by the histone deacetylase inhibitor MGCD0103 and its therapeutic implication in B-cell chronic lymphocytic leukemia. Leukemia. 2014;28(8):1636–46. doi:10.1038/leu.2014.19.

Eskelinen EL. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. Mol Asp Med. 2006;27(5–6):495–502. doi:10.1016/j.mam.2006.08.005.

Feng W, et al. Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation. Cancer. 2008;112(7):1489–502. doi:10.1002/cncr.23323.

Francisco R, Perez-Perarnau A, Cortes C, Gil J, Tauler A, Ambrosio S. Histone deacetylase inhibition induces apoptosis and autophagy in human neuroblastoma cells. Cancer Lett. 2012;318(1):42–52. doi:10.1016/j.canlet.2011.11.036.

Frankel LB, Lund AH. MicroRNA regulation of autophagy. Carcinogenesis. 2012;33(11):2018–25. doi:10.1093/carcin/bgs266.

Frankel LB, et al. MicroRNA-101 is a potent inhibitor of autophagy. EMBO J. 2011;30(22):4628–41. doi:10.1038/emboj.2011.331.

Fu J, Shao CJ, Chen FR, Ng HK, Chen ZP. Autophagy induced by valproic acid is associated with oxidative stress in glioma cell lines. Neuro-Oncology. 2010;12(4):328–40. doi:10.1093/neuonc/nop005.

Fu LL, Cheng Y, Liu B. Beclin-1: autophagic regulator and therapeutic target in cancer. Int J Biochem Cell Biol. 2013;45(5):921–4. doi:10.1016/j.biocel.2013.02.007.

Fullgrabe J, et al. The histone H4 lysine 16 acetyltransferase hMOF regulates the outcome of autophagy. Nature. 2013;500(7463):468–71. doi:10.1038/nature12313.

Furuya N, Yu J, Byfield M, Pattingre S, Levine B. The evolutionarily conserved domain of Beclin 1 is required for Vps34 binding, autophagy and tumor suppressor function. Autophagy. 2005;1(1):46–52.

Gan B, Guan JL. FIP200, a key signaling node to coordinately regulate various cellular processes. Cell Signal. 2008;20(5):787–94. doi:10.1016/j.cellsig.2007.10.021.

Gao W, Shen Z, Shang L, Wang X. Upregulation of human autophagy-initiation kinase ULK1 by tumor suppressor p53 contributes to DNA-damage-induced cell death. Cell Death Differ. 2011;18(10):1598–607. doi:10.1038/cdd.2011.33.

Gao YS, Hubbert CC, Lu J, Lee YS, Lee JY, Yao TP. Histone deacetylase 6 regulates growth factor-induced actin remodeling and endocytosis. Mol Cell Biol. 2007;27(24):8637–47. doi:10.1128/mcb.00393-07.

Ge J, Li L, Jin Q, Liu YC, Zhao L, Song HH. Functional IRGM polymorphism is associated with language impairment in glioma and upregulates cytokine expressions. Tumour Biol. 2014;35(8):8343–8. doi:10.1007/s13277-014-2091-x.

Geiss-Friedlander R, Melchior F. Concepts in SUMOylation: a decade on. Nat Rev Mol Cell Biol. 2007;8(12):947–56. doi:10.1038/nrm2293.

Goldenberg D, Lee J, Koch WM, Kim MM, Trink B, Sidransky D, Moon CS. Habitual risk factors for head and neck cancer. Otolaryngol Head Neck Surg. 2004;131(6):986–93. doi:10.1016/j.otohns.2004.02.035.

Gorrini C, et al. Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. Nature. 2007;448(7157):1063–7. doi:10.1038/nature06055.

Gu W, Wan D, Qian Q, Yi B, He Z, Gu Y, Wang L, He S. Ambra1 is an essential regulator of autophagy and apoptosis in SW620 cells: pro-survival role of Ambra1. PLoS One. 2014;9(2):e90151. doi:10.1371/journal.pone.0090151.

He G, Wang Y, Pang X, Zhang B. Inhibition of autophagy induced by TSA sensitizes colon cancer cell to radiation. Tumour Biol. 2014;35(2):1003–11. doi:10.1007/s13277-013-1134-z.

Hu X, et al. Protocadherin 17 acts as a tumour suppressor inducing tumour cell apoptosis and autophagy, and is frequently methylated in gastric and colorectal cancers. J Pathol. 2013;229(1):62–73. doi:10.1002/path.4093.

Huang BH, et al. Inhibition of histone deacetylase 2 increases apoptosis and p21Cip1/WAF1 expression, independent of histone deacetylase 1. Cell Death Differ. 2005;12(4):395–404. doi:10.1038/sj.cdd.4401567.

Huang W, et al. Crystal structure and biochemical analyses reveal Beclin 1 as a novel membrane binding protein. Cell Res. 2012;22(3):473–89. doi:10.1038/cr.2012.24.

Huijbers A, et al. The effect of the ATG16L1 Thr300Ala polymorphism on susceptibility and outcome of patients with epithelial cell-derived thyroid carcinoma. Endocr Relat Cancer. 2012;19(3):L15–8. doi:10.1530/erc-11-0302.

Inoue D, et al. Accumulation of p62/SQSTM1 is associated with poor prognosis in patients with lung adenocarcinoma. Cancer Sci. 2012;103(4):760–6. doi:10.1111/j.1349-7006.2012.02216.x. Epub 2012 Feb 21.

Inoue T, et al. SIRT2 knockdown increases basal autophagy and prevents postslippage death by abnormally prolonging the mitotic arrest that is induced by microtubule inhibitors. FEBS J. 2014;281(11):2623–37. doi:10.1111/febs.12810. Epub 2014 Apr 30.

Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. Manipulation of nonsense mediated decay identifies gene mutations in colon cancer cells with microsatellite instability. Oncogene. 2004;23(3):639–45. doi:10.1038/sj.onc.1207178.

Ivanov A, et al. Lysosome-mediated processing of chromatin in senescence. J Cell Biol. 2013;202(1):129–43. doi:10.1083/jcb.201212110.

Jaber N, Zong WX. Class III PI3K Vps34: essential roles in autophagy, endocytosis, and heart and liver function. Ann N Y Acad Sci. 2013;1280:48–51. doi:10.1111/nyas.12026.

Jahraus A, et al. ATP-dependent membrane assembly of F-actin facilitates membrane fusion. Mol Biol Cell. 2001;12(1):155–70.

Jiang S, et al. Intensive expression of UNC-51-like kinase 1 is a novel biomarker of poor prognosis in patients with esophageal squamous cell carcinoma. Cancer Sci. 2011;102(8):1568–75. doi:10.1111/j.1349-7006.2011.01964.x.

Jo YK, et al. Increased expression of ATG10 in colorectal cancer is associated with lymphovascular invasion and lymph node metastasis. PLoS One. 2012;7(12):e52705. doi:10.1371/journal.pone.0052705.

Jonchere B, Belanger A, Guette C, Barre B, Coqueret O. STAT3 as a new autophagy regulator. JAKSTAT. 2013;2(3):e24353. doi:10.4161/jkst.24353.

Jung CH, et al. ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery. Mol Biol Cell. 2009;20(7):1992–2003. doi:10.1091/mbc.E08-12-1249.

Jung KH, et al. Histone deacetylase 6 functions as a tumor suppressor by activating c-Jun NH2-terminal kinase-mediated beclin 1-dependent autophagic cell death in liver cancer. Hepatology. 2012;56(2):644–57. doi:10.1002/hep.25699.

Kang MR, et al. Frameshift mutations of autophagy-related genes ATG2B, ATG5, ATG9B and ATG12 in gastric and colorectal cancers with microsatellite instability. J Pathol. 2009;217(5):702–6. doi:10.1002/path.2509.

Kang R, Zeh HJ, Lotze MT, Tang D. The Beclin 1 network regulates autophagy and apoptosis. Cell Death Differ. 2011;18(4):571–80. doi:10.1038/cdd.2010.191.

Katoh H, et al. Genetic profile of hepatocellular carcinoma revealed by array-based comparative genomic hybridization: identification of genetic indicators to predict patient outcome. J Hepatol. 2005;43(5):863–74. doi:10.1016/j.jhep.2005.05.033.

Ke XX, Zhang D, Zhu S, Xia Q, Xiang Z, Cui H. Inhibition of H3K9 methyltransferase G9a repressed cell proliferation and induced autophagy in neuroblastoma cells. PLoS One. 2014;9(9):e106962. doi:10.1371/journal.pone.0106962.

Kim MS, Jeong EG, Ahn CH, Kim SS, Lee SH, Yoo NJ. Frameshift mutation of UVRAG, an autophagy-related gene, in gastric carcinomas with microsatellite instability. Hum Pathol. 2008;39(7):1059–63. doi:10.1016/j.humpath.2007.11.013.

Kim Y, et al. BIX-01294 induces autophagy-associated cell death via EHMT2/G9a dysfunction and intracellular reactive oxygen species production. Autophagy. 2013;9(12):2126–39. doi:10.4161/auto.26308.

Kiyono K, et al. Autophagy is activated by TGF-beta and potentiates TGF-beta-mediated growth inhibition in human hepatocellular carcinoma cells. Cancer Res. 2009;69(23):8844–52. doi:10.1158/0008-5472.can-08-4401.

Klionsky DJ. Autophagy: from phenomenology to molecular understanding in less than a decade. Nat Rev Mol Cell Biol. 2007;8(11):931–7. doi:10.1038/nrm2245.

Ko YH, et al. Stage-stratified analysis of prognostic significance of Bax-interacting factor-1 expression in resected colorectal cancer. Biomed Res Int. 2013;2013:329839. doi:10.1155/2013/329839.

Kondo Y, Shen L, Ahmed S, Boumber Y, Sekido Y, Haddad BR, Issa JP. Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. PLoS One. 2008;3(4):e2037. doi:10.1371/journal.pone.0002037.

Kovaleva V, et al. miRNA-130a targets ATG2B and DICER1 to inhibit autophagy and trigger killing of chronic lymphocytic leukemia cells. Cancer Res. 2012;72(7):1763–72. doi:10.1158/0008-5472.can-11-3671.

Ku B, Woo JS, Liang C, Lee KH, Jung JU, Oh BH. An insight into the mechanistic role of Beclin 1 and its inhibition by prosurvival Bcl-2 family proteins. Autophagy. 2008;4(4):519–20.

Kulis M, Esteller M. DNA methylation and cancer. Adv Genet. 2010;70:27–56. doi:10.1016/b978-0-12-380866-0.60002-2.

Kunzli BM, et al. Influences of the lysosomal associated membrane proteins (Lamp-1, Lamp-2) and Mac-2 binding protein (Mac-2-BP) on the prognosis of pancreatic carcinoma. Cancer. 2002;94(1):228–39.

Lee EY, Muller WJ. Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol. 2010;2(10):a003236. doi:10.1101/cshperspect.a003236.

Lee JY, et al. HDAC6 controls autophagosome maturation essential for ubiquitin-selective quality-control autophagy. EMBO J. 2010;29(5):969–80. doi:10.1038/emboj.2009.405.

Lee YJ, et al. Molecular mechanism of SAHA on regulation of autophagic cell death in tamoxifen-resistant MCF-7 breast cancer cells. Int J Med Sci. 2012;9(10):881–93. doi:10.7150/ijms.5011.

Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. 2011;11(1):9–22. doi:10.1038/nrc2982.

Lehnertz B, et al. The methyltransferase G9a regulates HoxA9-dependent transcription in AML. Genes Dev. 2014;28(4):317–27. doi:10.1101/gad.236794.113.

Levine B, Mizushima N, Virgin HW. Autophagy in immunity and inflammation. Nature. 2011;469(7330):323–35. doi:10.1038/nature09782.

Li X, et al. Imperfect interface of Beclin1 coiled-coil domain regulates homodimer and heterodimer formation with Atg14L and UVRAG. Nat Commun. 2012;3:662. doi:10.1038/ncomms1648.

Li J, et al. Autophagy promotes hepatocellular carcinoma cell invasion through activation of epithelial-mesenchymal transition. Carcinogenesis. 2013;34(6):1343–51. doi:10.1093/carcin/bgt063.

Li X, Wang S, Chen Y, Liu G, Yang X. miR-22 targets the 3′ UTR of HMGB1 and inhibits the HMGB1-associated autophagy in osteosarcoma cells during chemotherapy. Tumour Biol. 2014;35(6):6021–8. doi:10.1007/s13277-014-1797-0.

Liang C, Feng P, Ku B, Dotan I, Canaani D, Oh BH, Jung JU. Autophagic and tumour suppressor activity of a novel Beclin1-binding protein UVRAG. Nat Cell Biol. 2006;8(7):688–99. doi:10.1038/ncb1426.

Liang C, Jung JU. Autophagy genes as tumor suppressors. Curr Opin Cell Biol. 2010;22(2):226–33. doi:10.1016/j.ceb.2009.11.003.

Liang XH, et al. Induction of autophagy and inhibition of tumorigenesis by beclin 1. Nature. 1999;402(6762):672–6. doi:10.1038/45257.

Lin SY, et al. Protein phosphorylation-acetylation cascade connects growth factor deprivation to autophagy. Autophagy. 2012;8(9):1385–6. doi:10.4161/auto.20959. Epub 2012 Jun 21.

Lin DC, et al. The genomic landscape of nasopharyngeal carcinoma. Nat Genet. 2014;46(8):866–71. doi:10.1038/ng.3006.

Liu C, Xu P, Chen D, Fan X, Xu Y, Li M, Yang X, Wang C. Roles of autophagy-related genes Beclin-1 and LC3 in the development and progression of prostate cancer and benign prostatic hyperplasia. Biomed Rep. 2013a;1(6):855–60. doi:10.3892/br.2013.171.

Liu KP, et al. LC3B-II deacetylation by histone deacetylase 6 is involved in serum-starvation-induced autophagic degradation. Biochem Biophys Res Commun. 2013b;441(4):970–5. doi:10.1016/j.bbrc.2013.11.007.

Liu TP, Lo HL, Wei LS, Hao-Yun Hsiao H, Yang PM. S-Adenosyl-L-methionine-competitive inhibitors of the histone methyltransferase EZH2 induce autophagy and enhance drug sensitivity in cancer cells. Anti-Cancer Drugs. 2014; doi:10.1097/cad.0000000000000166.

Lopez G, et al. Autophagic survival in resistance to histone deacetylase inhibitors: novel strategies to treat malignant peripheral nerve sheath tumors. Cancer Res. 2011;71(1):185–96. doi:10.1158/0008-5472.can-10-2799.

Lu Z, et al. ARHI (DIRAS3) induces autophagy in ovarian cancer cells by downregulating the epidermal growth factor receptor, inhibiting PI3K and Ras/MAP signaling and activating the FOXo3a-mediated induction of Rab7. Cell Death Differ. 2014;21(8):1275–89. doi:10.1038/cdd.2014.48.

Luo S, Rubinsztein DC. BCL2L11/BIM: a novel molecular link between autophagy and apoptosis. Autophagy. 2013;9(1):104–5. doi:10.4161/auto.22399.

Lv L, et al. Acetylation targets the M2 isoform of pyruvate kinase for degradation through chaperone-mediated autophagy and promotes tumor growth. Mol Cell. 2011;42(6):719–30. doi:10.1016/j.molcel.2011.04.025.

Maejima Y, et al. Mst1 inhibits autophagy by promoting the interaction between Beclin1 and Bcl-2. Nat Med. 2013;19(11):1478–88. doi:10.1038/nm.3322.

Marklund U, Larsson N, Gradin HM, Brattsand G, Gullberg M. Oncoprotein 18 is a phosphorylation-responsive regulator of microtubule dynamics. EMBO J. 1996;15(19):5290–8.

Marzinke MA, Choi CH, Chen L, Shih Ie M, Chan DW, Zhang H. Proteomic analysis of temporally stimulated ovarian cancer cells for biomarker discovery. Mol Cell Proteomics. 2013;12(2):356–68. doi:10.1074/mcp.M112.019521.

McEwan DG, Dikic I. The Three Musketeers of Autophagy: phosphorylation, ubiquitylation and acetylation. Trends Cell Biol. 2011;21(4):195–201. doi:10.1016/j.tcb.2010.12.006.

Mizushima N, Yoshimori T, Ohsumi Y. The role of Atg proteins in autophagosome formation. Annu Rev Cell Dev Biol. 2011;27:107–32. doi:10.1146/annurev-cellbio-092910-154005.

Moresi V, et al. Histone deacetylases 1 and 2 regulate autophagy flux and skeletal muscle homeostasis in mice. Proc Natl Acad Sci U S A. 2012;109(5):1649–54. doi:10.1073/pnas.1121159109.

Mork CN, Faller DV, Spanjaard RA. Loss of putative tumor suppressor EI24/PIG8 confers resistance to etoposide. FEBS Lett. 2007;581(28):5440–4. doi:10.1016/j.febslet.2007.10.046.

Murr R, et al. Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. Nat Cell Biol. 2006;8(1):91–9. doi:10.1038/ncb1343.

Nakagawa M, et al. Expression profile of class I histone deacetylases in human cancer tissues. Oncol Rep. 2007;18(4):769–74.

Nakatogawa H. Two ubiquitin-like conjugation systems that mediate membrane formation during autophagy. Essays Biochem. 2013;55:39–50. doi:10.1042/bse0550039.

Nicoli ER, Dumitrescu T, Uscatu CD, Popescu FD, Streata I, Serban Sosoi S, et al. Determination of autophagy gene ATG16L1 polymorphism in human colorectal cancer. Romanian J Morphol Embryol. 2014;55(1):57–62.

Nitta T, et al. Autophagy may promote carcinoma cell invasion and correlate with poor prognosis in cholangiocarcinoma. Int J Clin Exp Pathol. 2014;7(8):4913–21.

Nixon RA, Yang DS. Autophagy and neuronal cell death in neurological disorders. Cold Spring Harb Perspect Biol. 2012;4(10). doi:10.1101/cshperspect.a008839.

Oga A, Kong G, Ishii Y, Izumi H, Park CY, Sasaki K. Preferential loss of 5q14-21 in intestinal-type gastric cancer with DNA aneuploidy. Cytometry. 2001;46(1):57–62.

Ou X, Lee MR, Huang X, Messina-Graham S, Broxmeyer HE. SIRT1 positively regulates autophagy and mitochondria function in embryonic stem cells under oxidative stress. Stem Cells. 2014;32(5):1183–94. doi:10.1002/stem.1641.

Ozdag H, et al. Differential expression of selected histone modifier genes in human solid cancers. BMC Genomics. 2006;7:90. doi:10.1186/1471-2164-7-90.

Paranjape T, Slack FJ, Weidhaas JB. MicroRNAs: tools for cancer diagnostics. Gut. 2009;58(11):1546–54. doi:10.1136/gut.2009.179531.

Park JM, Tougeron D, Huang S, Okamoto K, Sinicrope FA. Beclin 1 and UVRAG confer protection from radiation-induced DNA damage and maintain centrosome stability in colorectal cancer cells. PLoS One. 2014;9(6):e100819. doi:10.1371/journal.pone.0100819.

Pattingre S, et al. Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. Cell. 2005;122(6):927–39. doi:10.1016/j.cell.2005.07.002.

Peng YF, et al. Promoting colonization in metastatic HCC cells by modulation of autophagy. PLoS One. 2013;8(9):e74407. doi:10.1371/journal.pone.0074407.

Perinchery G, et al. Loss of two new loci on chromosome 8 (8p23 and 8q12-13) in human prostate cancer. Int J Oncol. 1999;14(3):495–500.

Petherick KJ, et al. Autolysosomal beta-catenin degradation regulates Wnt-autophagy-p62 crosstalk. EMBO J. 2013;32(13):1903–16. doi:10.1038/emboj.2013.123.

Petkova DS, Viret C, Faure M. IRGM in autophagy and viral infections. Front Immunol. 2012;3:426. doi:10.3389/fimmu.2012.00426.
Pike LR, et al. Transcriptional up-regulation of ULK1 by ATF4 contributes to cancer cell survival. Biochem J. 2013;449(2):389–400. doi:10.1042/bj20120972.
Plantinga TS, et al. Role of genetic variants of autophagy genes in susceptibility for non-medullary thyroid cancer and patients outcome. PLoS One. 2014;9(4):e94086. doi:10.1371/journal.pone.0094086.
Powell MJ, et al. Disruption of a Sirt1-dependent autophagy checkpoint in the prostate results in prostatic intraepithelial neoplasia lesion formation. Cancer Res. 2011;71(3):964–75. doi:10.1158/0008-5472.can-10-3172.
Price C, Chen J. MicroRNAs in cancer biology and therapy: current status and perspectives. Genes Dis. 2014;1(1):53–63. doi:10.1016/j.gendis.2014.06.004.
Puissant A, et al. Resveratrol promotes autophagic cell death in chronic myelogenous leukemia cells via JNK-mediated p62/SQSTM1 expression and AMPK activation. Cancer Res. 2010;70(3):1042–52. doi:10.1158/0008-5472.can-09-3537.
Qin Z, et al. Potentially functional polymorphisms in ATG10 are associated with risk of breast cancer in a Chinese population. Gene. 2013;527(2):491–5. doi:10.1016/j.gene.2013.06.067.
Qu X, et al. Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. J Clin Invest. 2003;112(12):1809–20. doi:10.1172/jci20039.
Ramus SJ, et al. BRCA1/2 mutation status influences somatic genetic progression in inherited and sporadic epithelial ovarian cancer cases. Cancer Res. 2003;63(2):417–23.
Rana S, Maples PB, Senzer N, Nemunaitis J. Stathmin 1: a novel therapeutic target for anticancer activity. Expert Rev Anticancer Ther. 2008;8(9):1461–70. doi:10.1586/14737140.8.9.1461.
Rao R, et al. Combination of pan-histone deacetylase inhibitor and autophagy inhibitor exerts superior efficacy against triple-negative human breast cancer cells. Mol Cancer Ther. 2012;11(4):973–83. doi:10.1158/1535-7163.mct-11-0979.
Sakuma T, et al. Aberrant expression of histone deacetylase 6 in oral squamous cell carcinoma. Int J Oncol. 2006;29(1):117–24.
Sakuraba K, et al. Down-regulation of Tip60 gene as a potential marker for the malignancy of colorectal cancer. Anticancer Res. 2009;29(10):3953–5.
Sanchez CG, et al. Activation of autophagy in mesenchymal stem cells provides tumor stromal support. Carcinogenesis. 2011;32(7):964–72. doi:10.1093/carcin/bgr029.
Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. Lung Cancer 2004, 45 Suppl 2:S3–S9. doi:10.1016/j.lungcan.2004.07.998.
Schaaf MB, et al. The autophagy associated gene, ULK1, promotes tolerance to chronic and acute hypoxia. Radiother Oncol. 2013;108(3):529–34. doi:10.1016/j.radonc.2013.06.015.
Shaid S, Brandts CH, Serve H, Dikic I. Ubiquitination and selective autophagy. Cell Death Differ. 2013;20(1):21–30. doi:10.1038/cdd.2012.72.
Shiraishi K, et al. A comparison of DNA copy number changes detected by comparative genomic hybridization in malignancies of the liver, biliary tract and pancreas. Oncology. 2001;60(2):151–61.
Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science. 2006;313(5792):1438–41. doi:10.1126/science.1129577.
Sinha S, Levine B. The autophagy effector Beclin 1: a novel BH3-only protein. Oncogene 2008, 27 Suppl 1:S137–S148. doi:10.1038/onc.2009.51.
Song J, et al. Increased expression of histone deacetylase 2 is found in human gastric cancer. APMIS. 2005;113(4):264–8. doi:10.1111/j.1600-0463.2005.apm_04.x.
Spurling CC, et al. HDAC3 overexpression and colon cancer cell proliferation and differentiation. Mol Carcinog. 2008;47(2):137–47. doi:10.1002/mc.20373.
Staub E, et al. A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer. Mol Cancer. 2006;5:37. doi:10.1186/1476-4598-5-37.
Takahashi Y, et al. Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. Nat Cell Biol. 2007;9(10):1142–51. doi:10.1038/ncb1634.

Takahashi Y, et al. Bif-1 haploinsufficiency promotes chromosomal instability and accelerates Myc-driven lymphomagenesis via suppression of mitophagy. Blood. 2013;121(9):1622–32. doi:10.1182/blood-2012-10-459826.

Tang J, et al. Low expression of ULK1 is associated with operable breast cancer progression and is an adverse prognostic marker of survival for patients. Breast Cancer Res Treat. 2012;134(2):549–60. doi:10.1007/s10549-012-2080-y.

Thompson HG, Harris JW, Wold BJ, Lin F, Brody JP. p62 overexpression in breast tumors and regulation by prostate-derived Ets factor in breast cancer cells. Oncogene. 2003;22(15):2322–33. doi:10.1038/sj.onc.1206325.

Torgersen ML, Engedal N, Boe SO, Hokland P, Simonsen A. Targeting autophagy potentiates the apoptotic effect of histone deacetylase inhibitors in t(8;21) AML cells. Blood. 2013;122(14):2467–76. doi:10.1182/blood-2013-05-500629.

Tuloup-Minguez V, Greffard A, Codogno P, Botti J. Regulation of autophagy by extracellular matrix glycoproteins in HeLa cells. Autophagy. 2011;7(1):27–39.

Wang J, et al. A non-canonical MEK/ERK signaling pathway regulates autophagy via regulating Beclin 1. J Biol Chem. 2009;284(32):21412–24. doi:10.1074/jbc.M109.026013.

Wang D, et al. Downregulation of FIP200 induces apoptosis of glioblastoma cells and microvascular endothelial cells by enhancing Pyk2 activity. PLoS One. 2011;6(5):e19629. doi:10.1371/journal.pone.0019629.

Wang J, et al. Sirtinol, a class III HDAC inhibitor, induces apoptotic and autophagic cell death in MCF-7 human breast cancer cells. Int J Oncol. 2012a;41(3):1101–9. doi:10.3892/ijo.2012.1534.

Wang RC, et al. Akt-mediated regulation of autophagy and tumorigenesis through Beclin 1 phosphorylation. Science. 2012b;338(6109):956–9. doi:10.1126/science.1225967.

Wang P, et al. MicroRNA 23b regulates autophagy associated with radioresistance of pancreatic cancer cells. Gastroenterology. 2013;145(5):1133–43. doi:10.1053/j.gastro.2013.07.048.

Wang L, Song J, Zhang J, Zhu C, Ma Y, Xu X. Lentiviral vector-mediate ATG3 overexpression inhibits growth and promotes apoptosis of human SKM-1 cells. Mol Biol Rep. 2014a;41(4):2093–9. doi:10.1007/s11033-014-3058-0.

Wang Z, et al. MicroRNA-25 regulates chemoresistance-associated autophagy in breast cancer cells, a process modulated by the natural autophagy inducer isoliquiritigenin. Oncotarget. 2014b;5(16):7013–26.

Watanabe M, et al. Induction of autophagy in malignant rhabdoid tumor cells by the histone deacetylase inhibitor FK228 through AIF translocation. Int J Cancer. 2009;124(1):55–67. doi:10.1002/ijc.23897.

Wei H, Wei S, Gan B, Peng X, Zou W, Guan JL. Suppression of autophagy by FIP200 deletion inhibits mammary tumorigenesis. Genes Dev. 2011;25(14):1510–27. doi:10.1101/gad.2051011.

Wei Y, et al. EGFR-mediated Beclin 1 phosphorylation in autophagy suppression, tumor progression, and tumor chemoresistance. Cell. 2013;154(6):1269–84. doi:10.1016/j.cell.2013.08.015.

Weichert W, et al. Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. Br J Cancer. 2008;98(3):604–10. doi:10.1038/sj.bjc.6604199.

Wu T, et al. Multi-step process of human breast carcinogenesis: a role for BRCA1, BECN1, CCND1, PTEN and UVRAG. Mol Med Rep. 2012;5(2):305–12. doi:10.3892/mmr.2011.634.

Xia M, et al. p62/SQSTM1 is involved in cisplatin resistance in human ovarian cancer cells via the Keap1-Nrf2-ARE system. Int J Oncol. 2014;45(6):2341–8. doi:10.3892/ijo.2014.2669.

Xu H, et al. UNC51-like kinase 1 as a potential prognostic biomarker for hepatocellular carcinoma. Int J Clin Exp Pathol. 2013;6(4):711–7.

Yang Y, et al. Acetylated hsp70 and KAP1-mediated Vps34 SUMOylation is required for autophagosome creation in autophagy. Proc Natl Acad Sci U S A. 2013;110(17):6841–6. doi:10.1073/pnas.1217692110.

Yin CK, Fu Q. Expressions of HDAC1 and HDAC2 in prostate cancer and their clinical implications. Zhonghua Nan Ke Xue. 2013;19(8):699–703.

Yuan H, et al. Inhibition of autophagy signi fi cantly enhances combination therapy with sorafenib and HDAC inhibitors for human hepatoma cells. World J Gastroenterol. 2014;20(17):4953–62. doi:10.3748/wjg.v20.i17.4953.

Zhang Z, et al. HDAC6 expression is correlated with better survival in breast cancer. Clin Cancer Res. 2004;10(20):6962–8. doi:10.1158/1078-0432.ccr-04-0455.

Zhang J, et al. Autophagic cell death induced by resveratrol depends on the Ca(2+)/AMPK/mTOR pathway in A549 cells. Biochem Pharmacol. 2013;86(2):317–28. doi:10.1016/j.bcp.2013.05.003.

Zhang Y, et al. Regulation of autophagy by miR-30d impacts sensitivity of anaplastic thyroid carcinoma to cisplatin. Biochem Pharmacol. 2014;87(4):562–70. doi:10.1016/j.bcp.2013.12.004.

Zhao X, et al. Apoptosis factor EI24/PIG8 is a novel endoplasmic reticulum-localized Bcl-2-binding protein which is associated with suppression of breast cancer invasiveness. Cancer Res. 2005;65(6):2125–9. doi:10.1158/0008-5472.can-04-3377.

Zhao Z, Tao L, Shen C, Liu B, Yang Z, Tao H. Silencing of Barkor/ATG14 sensitizes osteosarcoma cells to cisplatininduced apoptosis. Int J Mol Med. 2014;33(2):271–6. doi:10.3892/ijmm.2013.1578.

Zhu P, et al. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. Cancer Cell. 2004;5(5):455–63.

Zhu H, Wu H, Liu X, Li B, Chen Y, Ren X, Liu CG, Yang JM. Regulation of autophagy by a beclin 1-targeted microRNA, miR-30a, in cancer cells. Autophagy. 2009;5(6):816–23. Epub 2009 Aug 20.

Zhu C, Chen Q, Xie Z, Ai J, Tong L, Ding J, Geng M. The role of histone deacetylase 7 (HDAC7) in cancer cell proliferation: regulation on c-Myc. J Mol Med (Berl). 2011;89(3):279–89. doi:10.1007/s00109-010-0701-7.

Zou Z, et al. MicroRNA-30a sensitizes tumor cells to cis-platinum via suppressing beclin 1-mediated autophagy. J Biol Chem. 2012;287(6):4148–56. doi:10.1074/jbc.M111.307405.

Dr. Fayaz Malik Dr. Fayaz Malik is an early career investigator involved in cancer research and discovery for the past several years. His major research area is to explore the role of autophagy and associated signaling pathways driving cancer stem cell plasticity during tumor invasion and metastasis. He is also interested in the discovery of new therapeutic agents targeting cancer stem cell-driven pathways to address drug resistance and tumor metastasis. Dr. Malik has authored more than 40 publications, several patents, and book chapters. He has also been conferred with several prestigious national and international awards and fellowships.

Cancer Genomics and Precision Medicine: A Way Toward Early Diagnosis and Effective Cancer Treatment

2

Ritika Tiwari and Bushra Ateeq

2.1 Introduction

Since the first draft of the Human Genome Project was completed in April 2003, biomedical researchers have been mining and extrapolating genomic data toward the goal of improving human health and realizing medical benefits. The promise of "personalized oncomedicine," the matching of therapeutics to appropriate molecular targets in individual cancer patients, lies in the convergence of cancer researchers, computational biologist, and clinicians to identify the driving mutations involved in tumor progression and metastasis and pursue appropriate therapies. The virtual concept of "cancer genome" in the development of uncontrolled cell growth was conceived as early as late nineteenth and early twentieth century by Theodor Boveri (Boveri 2008). Boveri hypothesized that malignant tumors could be the result of a certain abnormal condition of the chromosomes arising from multipolar mitosis. Several decades later the discovery of the Philadelphia chromosome as the genetic driver of chronic myeloid leukemia (CML) provided the experimental evidence for Boveri's hypothesis (Nowell and Hungerford 1961).

The first description of the translocation between chromosomes 9 and 22 in the Philadelphia chromosome was reported by Janet D. Rowley in 1980 (Rowley 1980); however, it was another 10 years before the genes involved in the rearrangement were identified as breakpoint cluster region (BCR; chromosome 22) and v-abl Abelson murine leukemia viral oncogene homolog (ABL; chromosome 9) (Groffen et al. 1984). BCR-ABL fusion protein was demonstrated to function as a constitutively activated tyrosine kinase that stimulated proliferation of myeloid cells, leading to the development of CML (Lugo et al. 1990). Subsequently, a new therapeutic agent, imatinib mesylate (Gleevec), was developed that targets the kinase domain of

R. Tiwari • B. Ateeq (✉)
Department of Biological Sciences and Bioengineering,
Indian Institute of Technology, Kanpur 208016, UP, India
e-mail: bushra@iitk.ac.in

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_2

the fusion protein which essentially reversed the high mortality rate of CML patients (Druker et al. 1996).

The discovery of the Philadelphia chromosome, the identification of BCR-ABL fusion protein, and the development of Gleevec targeting the oncoprotein are a classic early example of personalized medicine. While the identification of BCR-ABL fusion to the development of Gleevec in CML represents a linear path from basic molecular discovery to medical success, most cancers are far more complex. Unlike CML (and most hematological cancers) where there is only a single causative genetic lesion, most solid tumors are highly heterogeneous and many harbor private mutation(s). However, over the last decade or so the field of medical oncology has experienced several remarkable breakthroughs. The amplification of the *HER2/neu* gene that was identified in ~20% of the breast cancer patients (Schechter et al. 1984; Slamon et al. 1987) led to the development of a monoclonal antibody, trastuzumab (Herceptin; Genentech), to treat HER2-positive breast cancer women (Robertson 1998); and lung cancer patients that harbor specific *EGFR* mutations were found to respond to gefitinib (Iressa) and erlotinib (Tarceva) that target these mutations (Lynch et al. 2004; Pao et al. 2004). *PARP1* inhibitor, olaparib, demonstrated antitumor activity in cancers associated with specific *BRCA1* or *BRCA2* mutations that impair the DNA repair pathway (ClinicalTrials.gov number, NCT00516373) (Fong et al. 2009).

The recent advances in high-throughput sequencing technologies that now make clinical sequencing economically feasible, combined with advanced computational approaches and higher-resolution analyses, have allowed us to obtain an unprecedented view of the genomic landscape of mutations/aberrations in individual cancer patients. Integrative sequencing strategies including whole-genome sequencing, targeted whole-exome sequencing, transcriptome sequencing (RNA-Seq), and shallow (5X–15X) paired-end whole-genome sequencing can be applied to uncover clinically significant genetic alterations in tumor specimens of patients and identify targets for existing therapies or direct patients to appropriate clinical trials.

2.2 Integrative Sequencing Strategy

Cancer arises from various genetic/molecular alterations including nucleic acid substitutions, gene fusions and rearrangements, amplifications and deletions, and a host of other aberrations that perturb gene expression (Stratton et al. 2009). Although the cancer genome can harbor multiple mutations, only few are "drivers" that confer clonal growth advantage, are positively selected, and are causally implicated in cancer development in a background of "passenger" mutations. Tumor specimens are often admixtures with varying fractions of normal tissue, or they may contain tumor subclones; therefore high sequencing depth is required for the detection of variants (Fig. 2.1). While whole-genome sequencing could be employed to identify copy number alterations (CNAs) and structural rearrangements at relatively shallow depth (Stephens et al. 2009), accurate identification of

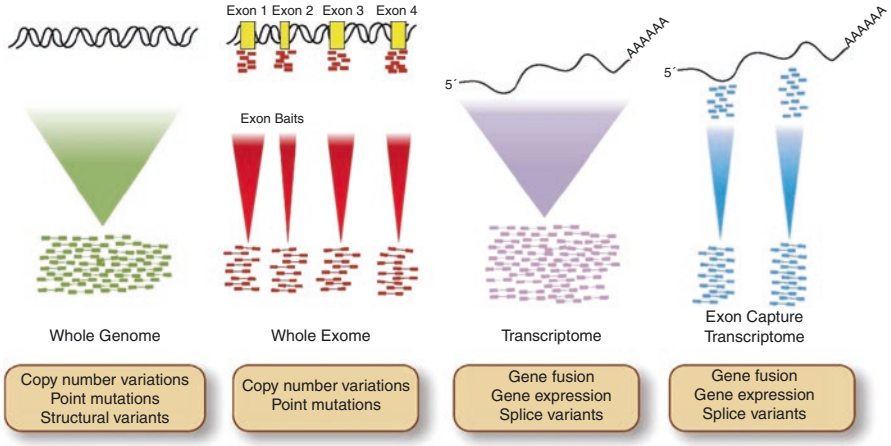

Fig. 2.1 Integrative sequencing of tumors for personalized oncology. Schema representing integration of whole-genome sequencing (*green*), whole-exome sequencing (*red*) for 1–2% of the genome, transcriptome or mRNA sequencing (*purple*), and targeted exome-capture sequencing (*blue*). Each sequencing strategy can be integrated for detecting genetic aberrations in cancer tissues including structural variants, CNVs, splice variants, point mutations, and gene expression

point mutations requires greater coverage and depth (Meyerson et al. 2010). Somatic mutations distinct from inherited DNA variants are identified by filtering out commonly inherited variants in human populations (>5% allele frequency) that have been registered in databases (Stratton et al. 2009); however, some rare inherited single nucleotide polymorphisms (SNPs) and structural variants may not be registered. Somatic mutations that are highly represented among cancer genes include protein kinases families in various signaling pathways; MAPK/ERK pathway is an example where upstream mutations are found in cell membrane-bound receptor tyrosine kinases such as *EGFR*, *ERBB2*, *FGFR1*, *FGFR2*, *FGFR3*, *PDGFRA*, and *PDGFRB* as well as in the downstream cytoplasmic components *NF1*, *PTPN11*, *HRAS*, *KRAS*, *NRAS*, and *BRAF* (Johnson and Lapadat 2002).

The International Cancer Genome Consortium (ICGC: http://www.icgc.org/home) undertook the task of comprehensively characterizing somatically acquired genetic alterations in at least 50 different classes of cancer, including those with the highest global incidence and mortality (Stratton et al. 2009). Currently, the ICGC has received commitments from various funding organizations in Asia, Australia, Europe, North America, and South America for 74 project teams in 17 jurisdictions to study over 25,000 tumor genomes that include cancers of the biliary tract, bladder, blood, bone, brain, breast, cervix, colon, eye, head and neck, kidney, liver, lung, nasopharynx, oral cavity, ovary, pancreas, prostate, rectum, skin, soft tissues, stomach, thyroid, and uterus. The genomic data generated by the participating ICGC members listed in Table 2.1 are made available by the Data Coordination Center through the ICGC website (www.icgc.org).

Table 2.1 List of the participating International Cancer Genome Consortium (ICGC) members for generating genomic data from multiple cancer types

ICGC members	Analyzed tumors
Australia	Ovarian and pancreatic cancer
Canada	Pancreatic, pediatric brain, and prostate cancer
China	Bladder, esophageal, gastric, and renal cancer
European Union/ France	Renal cancer
France	Liver cancer
Germany	Blood, brain, and prostate cancer
India	Oral cancer
Japan	Liver cancer
Saudi Arabia	Thyroid cancer
South Korea	Blood and lung cancer
Spain	Blood cancer
UK	Blood, bone, breast, esophageal, lung, prostate, and skin cancer
USA	Bladder, blood, brain, breast, cervical, colon, gastric, head and neck, liver, lung, ovarian, pancreatic, prostate, rectal, renal, skin, thyroid, and uterine cancer

2.3 Whole-Genome Sequencing

Whole-genome sequencing provides comprehensive characterization of somatic and germline mutations in a specimen. The most commonly used methods to make single nucleotide variant calls are (1) comparison with other sequenced genomes via Single Nucleotide Polymorphism database (dbSNP) and other resources for variant discovery such as the 1000 Genomes Project (www.1000genomes.org) and (2) critical assessment of remaining variant sites by comparison of tumor and matched normal genome. This approach also takes into consideration two primary measures to distinguish high- from low-quality variants (Mardis and Wilson 2009): first, a cumulative base-calling quality value that is summed from the individual quality values of each base identifying the putative variant (assigned by the Illumina's analysis pipeline known as Consensus Assessment of Sequence and Variation—CASAVA software) and second, a mapping quality value assigned by MAQ (Mapping and Assembly with Quality) assessing the genome-wide uniqueness of each aligned read (Li et al. 2008). CASAVA enables genomic builds, SNP calls, insertions/deletions (indels) detection, and count reads from the data generated from one or more runs across a broad range of sequencing applications (Table 2.2). Additionally, MAQ enables both read mapping and genotype calling from simulated and real data by utilizing mate-pair information and estimates the error probability of each read alignment. Recently, whole-genome sequencing of bladder cancers at a median depth of ~80X revealed recurrent protein-inactivating mutations in *CDKN1A* and *FAT1*. Moreover, the Stampy and Platypus programs have been used for mapping and aligning and somatic base substitution/single

Table 2.2 Key terminologies commonly used in next-generation sequencing

Key terminologies	Definition
Number of reads	Total amount of sequence data output by the instrument
Coverage	The average number of reads that align to or "cover" known reference bases
Sequencing depth	The total number of bases sequenced and aligned at a given reference base position
Read length	Number of base pairs of a given read
Error rate	Overall error rates are calculated by dividing the total number of errors by the number of known bases in the reference genome
Paired-end reads	A technology that obtains sequence reads from both ends of a DNA fragment template to generate high-quality, alignable sequence data. Paired-end sequencing facilitates detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts
Mate-pair reads	Mate-pair sequencing is similar to paired-end sequencing; however, the size of the DNA fragments used as sequencing templates is much longer (1000–10,000 bp). Mate-pair methods are particularly valuable for joining contigs in de novo sequencing and for detecting translocations and large deletions (structural variants)
Multiplexing	Processing of a large number of samples on a high-throughput instrument. Sample multiplexing is a useful technique when targeting specific genomic regions or working with smaller genomes

nucleotide variant (SNV) calls with high confidence ranging from 27,490 to 121,016 (Cazier et al. 2014). A recent study that employed exome sequencing of 50 lethal, heavily treated metastatic castration-resistant prostate cancers (CRPCs) demonstrated recurrent (8.6%) mutations in multiple chromatin- and histone-modifying genes such as *MLL2* and the AR collaborating factor *FOXA1* (3.4%) (Grasso et al. 2012). Nevertheless, evaluating somatic mutations in cancer specimens is often challenging for samples with very low tumor content (percentage of tumor cells in a given specimen) and can limit analysis as the differential variants between tumor and normal sample that could be detected would be very low.

2.4 Transcriptome Sequencing

Transcriptome sequencing (RNA-Seq) provides a comprehensive landscape of the expressed genome that includes all unique RNA transcripts from both coding and noncoding regions. In addition to the sequence and identity of RNA species in a sample, RNA-Seq can identify genomic rearrangements, copy number variations (CNVs), focused indels, and single nucleotide mutations. Importantly, transcriptome sequencing can also provide gene expression data with more sensitivity than microarray experiments (Marioni et al. 2008; Sultan et al. 2008). Utilizing RNA-Seq, Maher et al. discovered novel fusion transcripts by employing single-end reads of various lengths. This approach nominated multiple candidates or chimeras such as *SLC45A3-ELK4* that were independently confirmed as a common "read-through" transcript identified in prostate cancer (Maher et al. 2009). A combination of two

next-generation sequencing platforms was utilized, and the data were integrated to identify fusion transcripts from cancer cell lines; first long reads from RNA-Seq data that partially aligned to the reference genome (Roche 454) were identified as putative fusion transcripts, and then short reads that spanned fusion junctions obtained from a second RNA-Seq dataset (Illumina Genome Analyzer) were integrated with the first dataset to nominate candidate gene fusions. Using this approach, Maher et al. successfully "rediscovered" previously known and novel fusion transcripts in the prostate cancer cell lines LNCaP and VCaP and various prostate tumor samples (Maher et al. 2009). Recently, Kalyana-Sundaram et al. developed a bioinformatics pipeline for explicitly detecting pseudogene transcripts from RNA-Seq data and demonstrated genome-wide expression of pseudogenes, which are ubiquitously expressed in a lineage and/or cancer-specific manner (Kalyana-Sundaram et al. 2012). Briefly, they discovered breast-specific unprocessed pseudogene ATP8A2Ψ, which possibly arouses from the duplication of wild-type ATP8A2, therefore likely harbors similar promoter elements. Similarly, a prostate cancer-specific pseudogene, CXADR-Ψ, was also revealed using the same bioinformatics framework. CXADR-Ψ is a processed pseudogene located on chromosome 15, and parental CXADR protein demonstrates putative tumor suppressor functions, and its loss has been implicated in α-catenin downregulation (Pong et al. 2003).

Chromosomal rearrangements leading to generation of gene fusions represent one of the common mechanisms for the expression of oncogenes in epithelial cancers (Chinnaiyan and Palanisamy 2010). Oncogenic genetic rearrangements were initially thought to be confined to hematological cancers (Mitelman 2000; Mitelman et al. 2007). In 2005, Chinnaiyan and colleagues reported recurrent gene fusions between the transmembrane protease serine 2 (*TMPRSS2*) gene and members of the ETS family of transcription factors, predominantly *ERG* (v-ets erythroblastosis virus E26 oncogene homolog (avian)), in prostate cancer, representing the first discovery of a gene fusion in a solid tumor (Tomlins et al. 2005). Subsequently, various gene fusions in a variety of cancers including breast and lung were discovered (Stephens et al. 2009; Martelli et al. 2009; Natrajan et al. 2014). Gene rearrangements of *SLC45A3-BRAF* (solute carrier family 45, member 3–v-raf murine sarcoma viral oncogene homolog B1) and *ESRP1-RAF1* (epithelial splicing regulatory protein-1–v-raf-1 murine leukemia viral oncogene homolog-1) were discovered in prostate cancer by employing paired-end massively parallel transcriptome (Palanisamy et al. 2010). Importantly, these fusions are potentially "druggable." Furthermore, identification of these RAF pathway gene rearrangements in a variety of cancer types—prostate and gastric cancers and melanoma—supports the notion that cancers should be stratified by the driving molecular alterations/genetic events rather than by organ site.

Using whole-exome and transcriptome sequencing, a genetic rearrangement between transcriptional repressor *NAB2* and the transcriptional activator *STAT6* was detected in all solitary fibrous tumors (SFT)/hemangiopericytoma cases tested, establishing *NAB2-STAT6* as the causative mutation of SFT (Robinson et al. 2013a). In addition to driving mutations, clinical sequencing can also uncover mechanisms of treatment resistance and disease progression. Recently, ER-positive, metastatic

breast cancer patients underwent sequence analysis that revealed mutations in the ligand-binding domain (LBD) of the estrogen receptor (ESR1) that resulted in constitutive activity and continued responsiveness to anti-estrogen therapies in vitro. These results suggest that activating mutations in ESR1 are a key mechanism in acquired endocrine resistance in breast cancer therapy (Robinson et al. 2013b). Moreover, recently identified novel variants of *CDK4*, *LARP1*, *ADD3*, and *PHLPP2* in breast cancer are adding to the repertoire of the cancer transcriptome as well as uncovering novel therapeutic targets (Eswaran et al. 2013). Transcriptome analysis of various cancers also identified recurrent novel fusion involving kinase receptors that can potentially serve as promising drug targets (Stransky et al. 2014).

A distinct advantage of unbiased transcriptome sequencing is the ability to study noncoding RNA species whose role in cellular processes and disease state is becoming increasingly appreciated. Long noncoding RNAs (lncRNAs) play a role in normal cellular processes and are also implicated in cancer progression and metastasis (Crea et al. 2014). Noncoding RNAs (ncRNAs) can be categorized into small (under 200 nucleotides) and large ncRNAs. The small ncRNAs include small nucleolar RNAs (snoRNAs), PIWI-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and microRNAs (miRNAs) (Amaral et al. 2008). Earlier, the lncRNA known as HOTAIR was found to be aberrantly over-expressed in advanced breast and colorectal cancer, and repressing HOTAIR expression in cancer cells attenuated the invasive potential of the cancer cells (Kogo et al. 2011; Wang and Chang 2011). Thus, RNA-Seq analysis is a powerful tool for understanding the transcriptome landscape of cancers and to molecularly stratify subsets of cancer by mutational classes of genetic aberrations. Further, the resulting datasets from various RNA-Seq methodologies can provide a wide range of information such as gene expression, methylation status, histone modifications, and genomic occupancy of transcription factors and other regulatory protein-binding positions.

2.5 Methylated DNA Immunoprecipitation (MeDIP) Sequencing

The term “epigenetics” was originally coined by Conrad Waddington to describe heritable changes in a cellular phenotype that were independent of alterations in the DNA sequence. All epigenetic changes such as chromatin remodeling, histone modifications, and DNA methylation are highly regulated by a group of chromatin-modifying enzymes. There are at least four known DNA modifications (Baylin and Jones 2011; Wu et al. 2011) and 16 classes of histone modifications (Kouzarides 2007; Tan et al. 2011). Initially, MethylC-seq, a bisulfite conversion method, was developed to analyze the methylome at single-base resolution (Cokus et al. 2008). In this approach, sodium bisulphite converts unmethylated cytosines to uracils leaving 5-methylated cytosines unchanged, and upon amplification by polymerase chain reaction (PCR), unmethylated cytosines appear as thymines and methylated cytosines appear as cytosines (Frommer et al. 1992). A combination of next-generation sequencing (NGS) platforms with established techniques such as chromatin

immunoprecipitation (ChIP-Seq) has yielded an unparalleled view of the epigenome. Importantly, NGS introduced a novel approach to assess genome-wide epigenetic changes in an unbiased manner without the limitations of probe-based microarray platforms. One of the most prevalent epigenetic alterations in cancer is the methylation changes that occur within CpG islands that are present in 70% of all mammalian promoters. CpG island methylation plays a critical role in transcriptional regulation, and it is commonly altered during malignant transformation (Baylin and Jones 2011). Genome-wide mapping of CpG methylation using NGS platforms has confirmed that ~5–10% of normally unmethylated CpG promoter islands become abnormally methylated in various cancer genomes (Ateeq et al. 2008; Szyf 2005). Moreover, CpG hypermethylation of promoters not only affects the expression of protein-coding genes but also the expression of various noncoding RNAs, some of which have a role in oncogenesis (Baylin and Jones 2011).

Kim et al. employed a novel deep-sequencing technique named MethylPlex to enrich for methylated regions of the genome to characterize the DNA methylome map of prostate cancer cells and tissues requiring minimal DNA input. Massively parallel sequencing of the enriched products identified differentially methylated regions (DMRs) and revealed novel insights regarding the genomic loci and functional consequences of DNA methylation in cancer (Kim et al. 2011). This study uncovered 6691 methylated promoters in prostate tissues, 2481 cancer-specific DMRs including several novel DMRs such as WFDC2 promoter that displayed increased levels of methylation in cancer tissues compared to benign tissues and normal prostate epithelial cells.

Whole-genome sequencing in a variety of cancers has identified recurrent somatic mutations in numerous epigenetic regulators as well (Forbes et al. 2011; Stratton et al. 2009). Targeted NGS resequencing of cancer genomes found mutations within EZH2 (enhancer of zeste 2 polycomb repressive complex 2), the catalytic subunit of polycomb repressive complex 2 that is over-expressed in multiple cancers (Cao et al. 2008; Li et al. 2009; Varambally et al. 2008), including lymphoid and myeloid cancers (Khan et al. 2013; Yoshida et al. 2013). Moreover, heterozygous missense mutations resulting in the substitution of tyrosine 641 (Y641) within the SET domain of EZH2 were observed in 22% of patients with diffuse large B-cell lymphoma (Morin et al. 2010). Recurrent mutations in the histone methyltransferase, MLL2, have been discovered in ~90% of follicular lymphoma patients (Morin et al. 2011). Similarly, mutations in UTX, a histone demethylase, were observed in up to 12 histologically distinct cancers (van Haaften et al. 2009). Given these findings, several new drugs against these epigenetic targets are in development (Arrowsmith et al. 2012).

Conclusions

The current genomics era holds tremendous promise for "personalized oncomedicine." The recent rapid advances in high-throughput sequencing technologies and the field of cancer genomics is catalyzing the discovery of novel "druggable" molecular targets that include oncogenes, protein pathways involved in signaling cascade, and networks shown to be involved in the pathogenesis of cancer as well as the development of therapies against these targets. In the not too distant future,

clinical sequencing of patient tumor specimens to inform therapeutic intervention is likely to be adopted as standard clinical practice.

Acknowledgments We thank Jyoti Athanikar for critically reading the manuscript. B.A. is an Intermediate Fellow of the Wellcome Trust-DBT India Alliance [IA/I(S)/12/2/500635 to BA] and a Young Investigator of the DST-FAST Track scheme.

Competing Interests No competing interests to be disclosed.

References

Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. Science. 2008;319:1787–9.

Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012;11:384–400.

Ateeq B, Unterberger A, Szyf M, Rabbani SA. Pharmacological inhibition of DNA methylation induces proinvasive and prometastatic genes in vitro and in vivo. Neoplasia. 2008;10:266–78.

Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. Nat Rev Cancer. 2011;11:726–34.

Boveri T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Harris H. J Cell Sci. 2008;121 Suppl 1:1–84.

Cao Q, et al. Repression of E-cadherin by the polycomb group protein EZH2 in cancer. Oncogene. 2008;27:7274–84.

Cazier JB, et al. Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. Nat Commun. 2014;5:3756.

Chinnaiyan AM, Palanisamy N. Chromosomal aberrations in solid tumors. Prog Mol Biol Transl Sci. 2010;95:55–94.

Cokus SJ, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452:215–9.

Crea F, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget. 2014;5:764–74.

Druker BJ, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med. 1996;2:561–6.

Eswaran J, et al. RNA sequencing of cancer reveals novel splicing alterations. Sci Rep. 2013;3:1689.

Fong PC, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361:123–34.

Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39:D945–50.

Frommer M, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA. 1992;89:1827–31.

Grasso CS, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012;487:239–43.

Groffen J, Stephenson JR, Heisterkamp N, de Klein A, Bartram CR, Grosveld G. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell. 1984;36:93–9.

Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. Science. 2002;298:1911–2.

Kalyana-Sundaram S, et al. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell. 2012;149:1622–34.

Khan SN, et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. Leukemia. 2013;27:1301–9.

Kim JH, et al. Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. Genome Res. 2011;21:1028–41.
Kogo R, 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.
Kouzarides T. Chromatin modifications and their function. Cell. 2007;128:693–705.
Li X, Gonzalez ME, Toy K, Filzen T, Merajver SD, Kleer CG. Targeted overexpression of EZH2 in the mammary gland disrupts ductal morphogenesis and causes epithelial hyperplasia. Am J Pathol. 2009;175:1246–54.
Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18:1851–8.
Lugo TG, Pendergast AM, Muller AJ, Witte ON. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science. 1990;247:1079–82.
Lynch TJ, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350:2129–39.
Maher CA, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;458:97–101.
Mardis ER, Wilson RK. Cancer genome sequencing: a review. Hum Mol Genet. 2009;18:R163–8.
Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18:1509–17.
Martelli MP, et al. EML4-ALK rearrangement in non-small cell lung cancer and non-tumor lung tissues. Am J Pathol. 2009;174:661–70.
Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010;11:685–96.
Mitelman F. Recurrent chromosome aberrations in cancer. Mutat Res. 2000;462:247–53.
Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer. 2007;7:233–45.
Morin RD, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010;42:181–5.
Morin RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476:298–303.
Natrajan R, et al. Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. J Pathol. 2014;232:553–65.
Nowell PC, Hungerford DA. Chromosome studies in human leukemia. II. Chronic granulocytic leukemia. J Natl Cancer Inst. 1961;27:1013–35.
Palanisamy N, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med. 2010;16:793–8.
Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci USA. 2004;101:13306–11.
Pong RC, et al. Epigenetic regulation of coxsackie and adenovirus receptor (CAR) gene promoter in urogenital cancer cells. Cancer Res. 2003;63:8680–6.
Robertson D. Genentech's anticancer Mab expected by November. Nat Biotechnol. 1998;16:615.
Robinson DR, et al. Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing. Nat Genet. 2013a;45:180–5.
Robinson DR, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013b;45:1446–51.
Rowley JD. Ph1-positive leukaemia, including chronic myelogenous leukaemia. Clin Haematol. 1980;9:55–86.
Schechter AL, et al. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. Nature. 1984;312:513–6.
Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987;235:177–82.

Stephens PJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009;462:1005–10.
Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun. 2014;5:4846.
Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458:719–24.
Sultan M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008;321:956–60.
Szyf M. Therapeutic implications of DNA methylation. Future Oncol. 2005;1:125–35.
Tan M, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell. 2011;146:1016–28.
Tomlins SA, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310:644–8.
van Haaften G, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009;41:521–3.
Varambally S, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science. 2008;322:1695–9.
Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43:904–14.
Wu H, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011;25:679–84.
Yoshida K, et al. The landscape of somatic mutations in down syndrome-related myeloid disorders. Nat Genet. 2013;45:1293–9.

Dr. Bushra Ateeq Dr. Bushra Ateeq is an Assistant Professor at Department of Biological Sciences and Bioengineering (BSBE), IIT, Kanpur, India. The primary research focus of Dr. Bushra's laboratory is to understand the complex molecular events involved in prostate and breast cancer, identify early diagnostic markers and valuable therapeutic targets. Her lab focuses on experimental evaluation of the functional relevance of the genetic rearrangements, copy number changes, and somatic alterations identified by genomic approaches, using molecular and cellular approaches.

Genetics of Liver Diseases

3

Nirupama Trehanpati, Dinesh Rawat, and Hamed Hemati

… Liver disease is best explored during the time when genes involved in adaptation to life and disease are given their first "test drive"—childhood…

Saul J Karpen, Hepatology 2008 Aug;48(2):353–4

Pathophysiology of liver disorders involves an interaction between environmental genetic and host factors. This includes childhood liver diseases like familial cholestatic syndromes as well as liver diseases manifesting in adulthood like alcoholic liver disease, non-alcoholic liver diseases, viral hepatitis, etc. Genetic predisposition can be due to single-gene mutations (as in alpha-1 antitrypsin deficiency), susceptible single-nucleotide polymorphisms in intrahepatic cholestasis of pregnancy, or modifier genes in drug-induced liver disease and histocompatibility leucocyte antigen (HLA) association in complex diseases as autoimmune hepatitis (AH). Understanding the genetic contribution to liver diseases is important for clinicians as the same disorder may have wide phenotypic variability while this is vital for research workers due to therapeutic implications.

N. Trehanpati (✉) • H. Hemati
Laboratory of Cellular Transplant Immunology, Department of Molecular and Cellular Medicine, Institute of Liver and Biliary Sciences (ILBS), New Delhi, India
e-mail: trehanpati@gmail.com

D. Rawat
Department of Pediatric Hepatology, Institute of Liver and Biliary Sciences, D-1, Vasant Kunj, New Delhi, India

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_3

Table 3.1 Liver disorders

Age at common liver diseases genetic contribution presentation			
Age	Disease		Genetic mutations
Neonate/infants	Famlial cholestatic syndrome	Alagille syndrome	JAG 1 (jagged 1) mutation notch 2 mutation
		PFIC1	ATP8B1 mutation
		PFIC2	ABCB11 mutation
		PFIC3	ABCB4 mutation
	Metabolic conditions	Alpha-1 antitrypsin deficiency	SERPINA 1 gene mutation
		Cystic fibrosis	CFTR gene mutation
	3D structural	Cystic liver disease/ductal plate malformation	Fibrocystin gene mutation
Childhood		Wilson's disease	ATP7B mutation
Adults		Hereditary	
		Haemochromatosis	
		Alcoholic liver disease	
		Viral hepatitis	
		NAFD	

3.1 Liver Diseases in Neonate/Infants

The liver disorders are often categorised by their age at presentation-neonatal period/infancy, childhood or adulthood (Table 3.1).

3.1.1 Familial Cholestatic Syndromes

Alagille syndrome (ALGS)—ALGS is autosomal dominant disorder with variable gene mutation expression affecting cellular differentiation as well as tissue development (Dhorne-Pollet et al. 1994). Predominantly (approx 95%) this disorder is caused due to mutations in JAG1 (jagged 1) gene (Li et al. 1997). JAG1 is a ligand for notch. JAG1 as a ligand binds to transmembrane notch receptors and initiates the notch signalling pathway responsible for determining cell fate and normal organogenesis (Artavanis-Tsakonas et al. 1999). Notch signalling has an important role in skeletal development, intrahepatic bile ducts as well as cardiac development and ductal plate remodelling. Mutations result in premature termination of proteins with curtailed and inactive proteins to decrease in the amount of normal protein. This phenomenon of haploinsufficiency is the postulated mechanism for causing AGLS. These mutations are distributed across the entire coding region of the *JAG1* gene with more than 430 mutations identified so far. Mutations from an affected parent are seen in 30–50% while many mutations (50–70%) are sporadic without obvious genotype-phenotype correlation (Crosnier et al. 2000). These commonly include protein-truncating frameshift and nonsense mutation (69%), splicing mutations (16%), missense mutation (11%) and deletions (4%) (Spinner et al. 2001) (Table 3.2).

Table 3.2 Alagille syndrome-related genes

Gene name	Chromosomal locus	Protein name[a]	No. of exons	Total mutations[b]	Gene mutations proportion in syndrome[c]
JAG1	20p12.1–p11.23	Protein jagged 1	26	474	94–96%
notch 2	1p13–p11	Neurogenic locus notch homolog protein 2 (notch 2)	34	44	1–2%

[a]Uniprot
[b]HGMD
[c]NCBI gene review

Table 3.3 PFIC-associated genes

PFIC type	Gene (protein)	GGT level	Pathophysiology	Clinical considerations	
Type 1	ATP8B1(FIC1)	Low/normal	Decreased bile acid transport	PFIC1 patients	Liver failure before adulthood, diarrhoea, hearing loss
				BRIC patients	Episodic pruritus and jaundice
Type 2	ABCB11 (BSEP)	Low/normal	Mutation in bile acid export pump (BSEP)	Risk of hepatobiliary malignancies	
Type 3	ABCB4(MDR3)	High	Destabilised micelles	Cholestasis	

The main clinical characteristics of Alagille syndrome include: (1) cholestasis, due to lack of intrahepatic bile ducts which presents in the neonatal period; (2) characteristic facies; (3) skeletal abnormalities, butterfly thoracic vertebrae; (4) eye abnormalities, posterior embryotoxon; and (5) cardiac abnormalities like peripheral pulmonic stenosis.

3.1.2 Progressive Familial Intrahepatic Cholestasis (PFIC)

Formerly known as Byler's disease, this is the prototype of bile canalicular transport defects. PFIC is characterised by low or normal level of gamma-glutamyl transpeptidase (GGT) in PFIC type 1 or 2 and high GGT in PFIC type 3. The main clinical manifestations are pruritus and cholestasis presenting in early infancy but may have variable phenotype with presentation across all age groups (Table 3.3).

3.1.2.1 PFIC Type 1 (FIC1 Disease)

This condition results from an abnormal regulation of bile acid homeostasis due to recessive mutations in ATP8B1 gene on chromosome 18q21. FIC1 disorder is characterised by persistent or repeated jaundice (BRIC, benign recurrent intrahepatic cholestasis) with low GGT disproportionate to the level of cholestasis.

In PFIC 1 disorder, generally liver disease progresses to liver failure before adulthood, while BRIC patients are on the mild end of the spectrum with recurrent episodic pruritus and jaundice but preserved liver function. Extrahepatic manifestations are characteristically seen in PFIC 1 which includes diarrhoea, poor growth, pancreatic insufficiency and hearing loss.

Genetic linkage analysis in patients with FIC1 and BRIC has helped in mapping of the responsible gene to chromosome 18q21 (Carlton et al. 1995) and showed that defect in P-type ATPase of FIC1 gene results in the disease (Fig. 3.1). ATP8B1 gene consists 28 exons in the span over 77 kb of genomic sequence and is expressed in many tissues, mostly in the pancreas, intestine and liver, thus explaining the extrahepatic manifestations (Bull et al. 1999). The exact pathophysiology has not been identified though it is known that FIC1 is an amino phospholipid transporter and helps to maintain the asymmetric distribution of amino phospholipids between the inner and outer leaflets of plasma membrane.

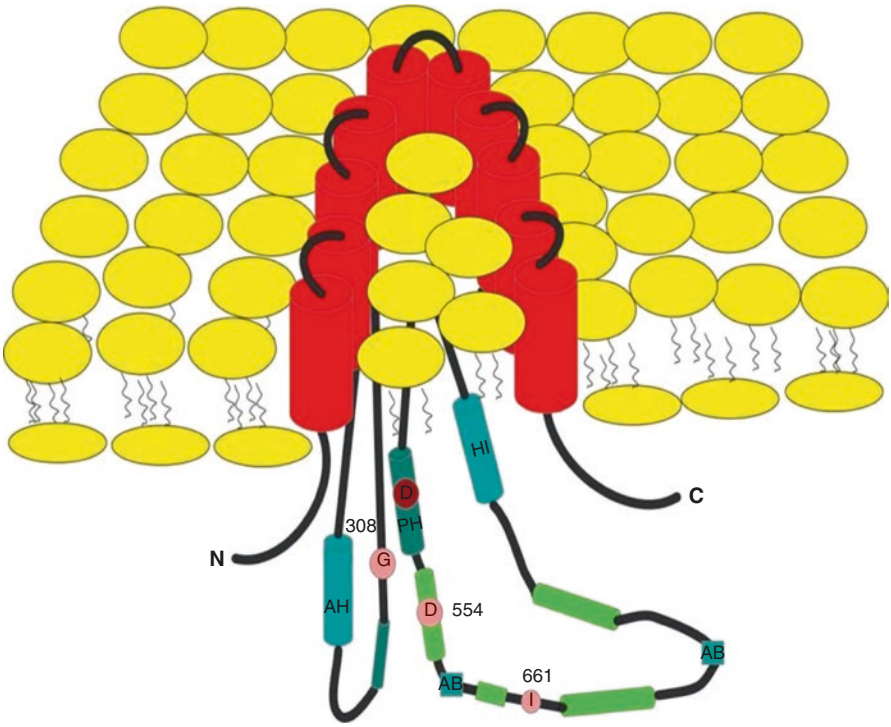

Fig. 3.1 Schematic representation of FIC1 protein topology. FIC1 is an integral component of the membrane with ten predicted transmembrane domains (*red*) and both the amino terminus (N) and the carboxy terminus (C) extending into the cytoplasm. The consensus domains for all classes of P-type ATPases are depicted in yellow. Consensus sequences for the phosphorylation site (PH), in which the aspartate residue is depicted in *blue* (D), the ATP-binding domains (AB) and the Hinge (HI) domain are located. In orange circles, three disease-associated missense mutation sites are depicted

The inheritance of FIC1 disease is compounded, with several mutations identified in the *ATP8B1* gene, which may partially explain the marked clinical differences in FIC1 disease (Klomp et al. 2000). Mutational analysis of families with FIC1 and BRIC have shown 54 disease specific mutations with predominant missense mutations in 24 while others included splice site, nonsense or frameshift mutations (Klomp et al. 2004). Specific mutations may be responsible for the disease in genetically isolated populations like the Amish or the Greenland specific populations where specific missense mutations (G308V and D554N, respectively; Fig. 3.1) were identified (Klomp et al. 2000). There is some genotype-phenotype correlation as missense mutations were seen more commonly in BRIC1 (58%) as compared to PFIC1 (38%) while large deletion, nonsense and frameshift mutations are more common with severe phenotype PFIC1 (Klomp et al. 2004). One missense mutation (I661T; Fig. 3.1) seen in BRIC patients of Western European origin has shown a dramatic variable phenotype. These patients have shown marked variability in age of onset as well as in the frequency and duration of the cholestatic episodes (Klomp et al. 2004).

3.1.2.2 PFIC Type 2 (BSEP, Bile Salt Export Pump Defect)

PFIC2 is characterised by defective bile salt excretion across hepatocyte canalicular membrane due to the defect in bile salt export pump. PFIC2 is caused due to mutations in ABCB11 gene, which encodes the ATP-dependent canalicular bile salt export pump. It is found that ATP-binding cassette (ABC) family of proteins is important for the normal enterohepatic circulation of bile salts and bile salt-dependent bile flow (Fig. 3.2). PFIC2 usually presents as cholestasis in early infancy with rapid progression to cirrhosis in childhood. Pruritus is the predominant clinical symptom disproportionate to the level of jaundice while extrahepatic manifestations are absent (in contrast to PFIC1). Occurrence of malignancy (hepatocellular carcinoma) has been noted as early as 10 months of age.

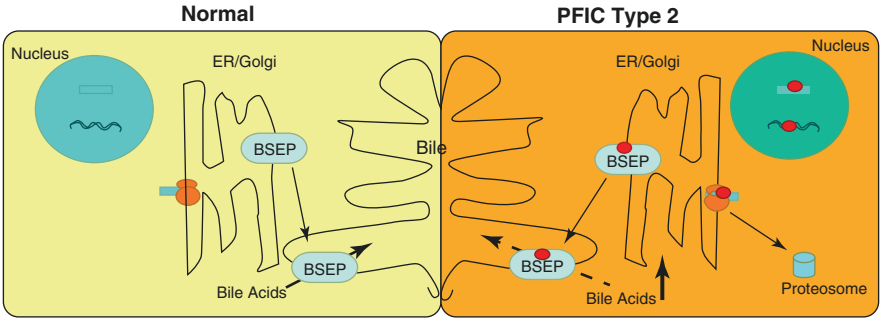

Fig. 3.2 Schematic representation of progressive familial intrahepatic cholestasis (PFIC) type 2. *Left panel*: under normal conditions bile salt export pump (BSEP) transports bile acids into bile. *Right panel*: mutation of the ABCB11/BSEP gene results in proteasomal degradation and/or expression of a protein with low/no function, leading to reduced transport of bile acids into bile and consequent accumulation of deleterious bile acids

Linkage analysis and homozygosity mapping in Middle Eastern families with low GGT cholestasis first identified this locus on chromosome 2q24 (Strautnieks et al. 1997). Several mutations have been identified in PFIC2, while some mutations have also been seen in BRIC, in drug-induced cholestasis and intrahepatic cholestasis during pregnancy. ABCB11 gene has 28 exons and, so far, about 100 mutations have been identified. Missense mutations are seen in majority, while others include splice site, deletion and insertion as well as nonsense mutations. Majority of children with BSEP mutations showed no canalicular BSEP protein expression (Jansen et al. 1999). Disease phenotype are often associated with mutations resulting in premature protein truncation or failure of protein production with little or no detectable canalicular BSEP expression and a higher risk of developing hepatocellular carcinoma. Missense mutations result in impaired BSEP protein processing and trafficking in the endoplasmic reticulum (i.e., p.E297G, p.D482G) or disrupt functional domains and protein structure. Moreover, mutations like p.N490D, p.G562D, p.R832C and p.A1110E showed detectable BSEP expression but with functional deficiency (Hayashi et al. 2005). However, in milder disease, such as BRIC2, missense mutations predominate over those leading to failure of protein production (van Mil et al. 2004).

3.1.2.3 PFIC Type 3 (PFIC3) or MDR3 Deficiency (Multidrug Resistance Protein 3 Deficiency)

PFIC3 disorder is an autosomal recessive, with similar features to PFIC1 and PFIC2 but with high GGT levels. PFIC3 is caused due to mutations in ABCB4 (previously called MDR3) located on chromosome 7q21. ABCB4 is a phospholipid translocator involved in biliary phospholipid (phosphatidylcholine) excretion and is expressed in the canalicular membrane of the hepatocyte (Jacquemin 2001). Cholestasis results from the toxicity of bile in which the bile salts are not solubilised due to absence of biliary phospholipids, leading to bile canaliculi and biliary epithelium injuries (Jacquemin et al. 2001).

The phenotypic continuum of PFIC3 spans from neonatal cholestasis to cirrhosis in young adults. In one large series of 31 patients, *ABCB4* sequence analysis revealed around 17 different mutations with 11 missense mutations and 6 mutations predicting a truncated protein (Jacquemin et al. 2001). Homozygous mutations in patients causing a truncated protein resulted in absence of canalicular MDR3 protein. MDR3 protein deficiency is seen when the truncated protein is broken down very rapidly after synthesis giving rise to extremely low steady levels of the protein. More likely, the premature stop codon may lead to instability and decay of the *ABCB4* mRNA. The absence of *ABCB4* mRNA is observed in several liver disease patients (de Vree et al. 1998). Missense mutations also lead to intracellular misprocessing of MDR3 and result in some residual MDR3 function with milder disease (Delaunay et al. 2009) (Fig. 3.3).

Therefore, it is now evidenced that in addition to PFIC3, an MDR3 defect can be involved in intrahepatic cholestasis of pregnancy (ICP3) (Dixon et al. 2000), cholesterol gallstone disease and drug-induced cholestasis (Rosmorduc et al. 2003; Lang et al. 2007).

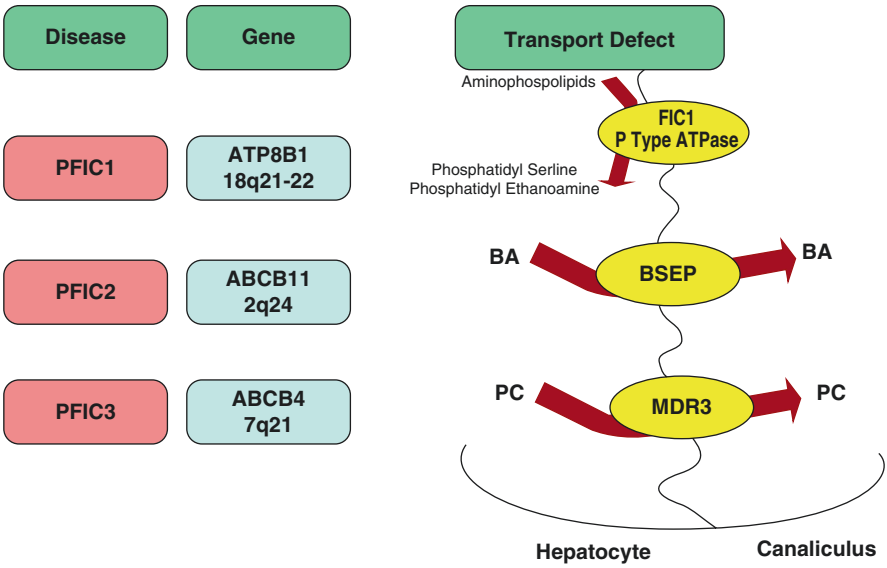

Fig. 3.3 Progressive familial intrahepatic cholestasis (PFIC): types, related genes and transport defects. *BA* bile acid, *PC* phosphatidylcholine

3.1.3 Cystic Fibrosis (CF)

Cystic fibrosis is also an autosomal recessive disease categorised as abnormal epithelial electrolyte transport with elevated sweat chloride concentrations, chronic lung disease and pancreatic dysfunctions. It is frequent potentially fatal genetic disorder in the Caucasians but is rarely seen in the Indian subcontinent. There are more than 2000 mutations identified in the cystic fibrosis transmembrane conductance regulator (CFTR) gene (CF Mutation Pubmed Database). CFTR causes neonatal cholestasis, steatosis, nodular or multilobular cirrhosis and biliary tract complications. CFTR gene encodes for a membrane-channel protein and is located on the long arm of chromosome 7. CFTR acts as a cAMP-dependent chloride channel in the apical membrane of most secretory epithelia, including biliary epithelial cells, or cholangiocytes. In cholangiocytes, CFTR has an important role in biliary secretion and bile flow.

The CFTR gene has 27 exons and 1480 amino acid membrane-bound glycoproteins, which is associated with ATP-binding cassette (ABC) superfamily proteins. CFTR protein contains two six-span membrane-bound regions each connected to a nuclear binding factor bound to ATP. A unique feature of CFTR is that in between these two units is an R-domain (Fig. 3.4).

The most common CFTR mutation is delta F508, (66%) widespread in all ethnicity worldwide. CFTR gene mutations are classified into five classes based on their effect on the CFTR protein. *Class I* mutations cause impairment of CFTR messenger RNA production (G542X); *Class II* (delta F508) mutations result in

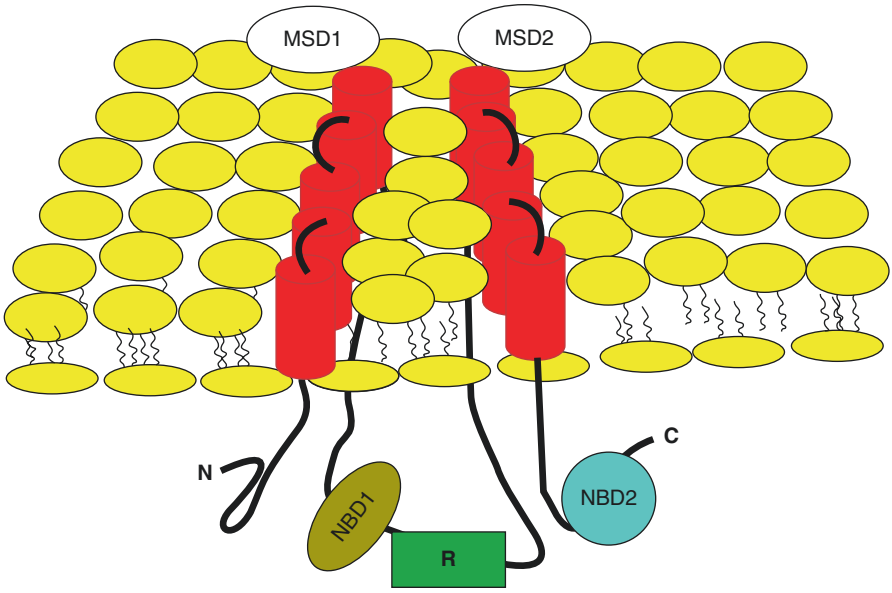

Fig. 3.4 CFTR domains

defective processing or trafficking of CFTR protein to the apical membrane; *Class III* mutation (G551D) is associated with defective regulation of CFTR, resulting in lack of response to cAMP agonists; *Class IV* mutations (R117H) demonstrate residual calcium channel conductance; and *Class V* mutations lead to abnormal splicing of CFTR with a reduction in the number and function of chloride channels (Fig. 3.5) (Koch et al. 2001). Class I, II and III mutations cause non-functioning of CFTR at plasma membrane and are considered as 'severe' mutation. Class IV and V mutations are considered as 'mild'. There is no correlation between the liver disease and genotype, which also indicate an important role of other genetic or environmental modifiers.

3.1.4 Alpha-1 Antitrypsin Deficiency

Alpha-1-antitrypsin (α-1AT) deficiency is an autosomal codominant disorder characterised by a mutant α-1 antitrypsin Z protein that fails to be secreted in body fluids due to abnormal folding. α-1AT is a 52 kDa glycoprotein formed in hepatocytes and secreted into the blood circulation, where it performs as main protease inhibitor specifically for destructing neutrophil protease, elastase, proteinase 3 and cathepsin G. The primary physiologic significance of the protein is in the lungs, where it protects the alveolar tissue from proteolytic damage by enzymes like neutrophil elastase and its deficiency results in emphysema in adults. Liver disease is due to the

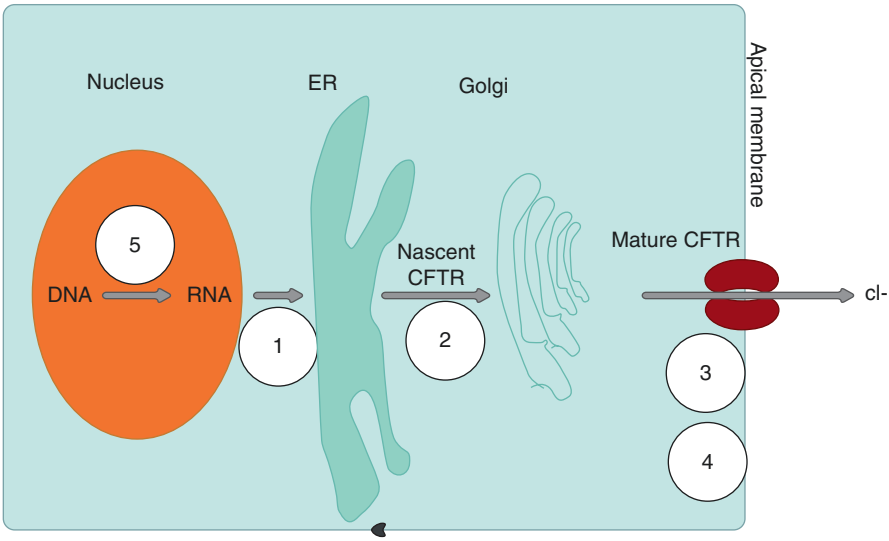

Fig. 3.5 Classes of CFTR mutations. (1) Early stop codons resulting in no CFTR protein. (2) Abnormal CFTR trafficking resulting in degradation in the endoplasmic reticulum. (3) Mature CFTR protein is refractory to normal activation. (4) CFTR activated normally but with reduction in single-channel conductance. (5) Splice site mutations resulting in decreased full length mRNAs and a decrease in functional CFTR at apical membrane. *ER* endoplasmic reticulum

hepatotoxic effect of the retained mutant protein within the hepatocytes (Fig. 3.6). Children usually present with neonatal cholestasis, chronic hepatitis with later development of cirrhosis, but pulmonary manifestations are not seen till adulthood. This is the most common genetic cause of liver disease in Caucasian children but not yet reported in Indian children.

The gene encoding α-1AT is present on chromosome 14q31–32.2 and is called the *SERPINA1* gene (Carrell and Lomas 2002). More than 100 allelic variants of the *SERPINA1* gene are known including normal variants (M1, M2, M3, etc.), null variants and deficiency variants. The α-1AT Z variant is due to single amino acid replacement at position 342 resulting in abnormal polymerisation of mutant protein within the endoplasmic reticulum (ER) and hepatocyte injury mediated via ER stress pathways. Hepatocyte nuclear factors (HNFs) especially HNF1α and HNF4 are important for expression of human SERPINA1 (Kalsheker et al. 2002), and humoral regulation is mediated by interleukin 6 (IL-6) and oncostatin M cytokines (Boutten et al. 1998).

3.1.5 Fibrocystic Hepatorenal Diseases

The congenital fibrocystic hepatorenal syndromes are autosomal recessive, monogenic disorders characterised by multiple defects in the liver and kidneys. Liver disease manifests as fibrosis often associated with cysts lined in biliary

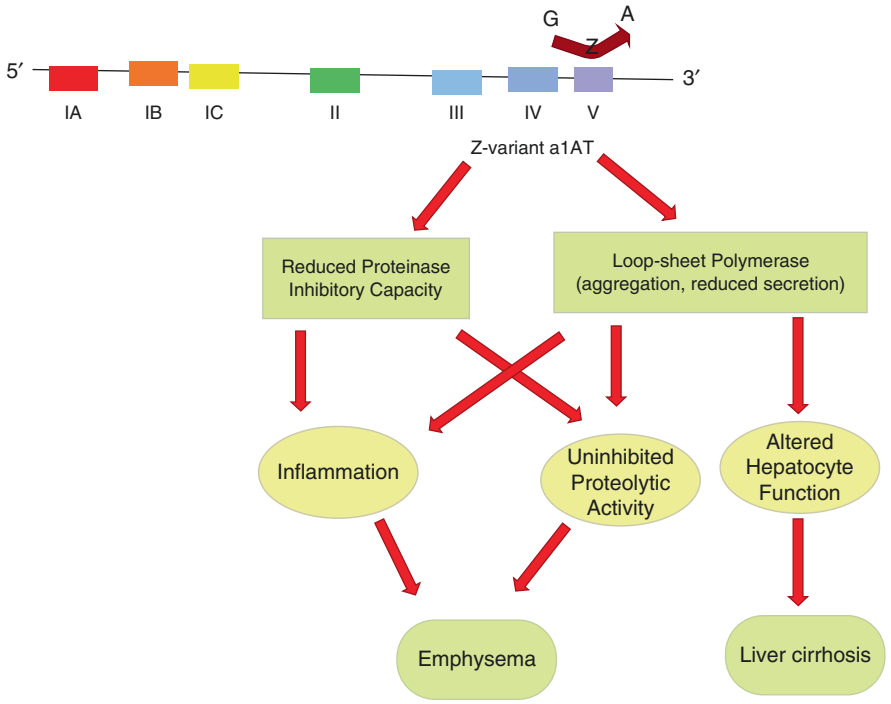

Fig. 3.6 The Z allele is the most important genetic defect in alpha-1 antitrypsin deficiency. It is a single mutation in exon 5 of the gene, leading to substitution of the amino acid glutamine (G) in position 342 in the protein for a lysine (A) amino acid. The Z allele results in hepatic polymerisation in both hepatocyte inclusions and decreased serum concentration. Therefore, strategies to augment the inherited deficiency as well as the development of small peptides that can selectively inhibit polymerisation of the Z allele of the AAT protein in the liver are central to therapeutic approach

epithelium leading to intrahepatic biliary tract dilatation. Cystic lesions also affect kidneys, and the severity determines the clinical presentation and long-term prognosis of these patients. Conditions like Joubert syndrome, Bardet-Biedl syndrome present during early infancy, while hereditary hepatic fibrosis and Caroli's disease are associated with autosomal recessive polycystic kidney disease (ARPKD), may manifest during childhood. Mutations in polycystic kidney and hepatic disease (*PKHD1*) gene located on chromosome 6p12.3-6p12.2 are responsible for ARPKD and Caroli's disease (Gunay-Aygun et al. 2013). *PKHD1* encodes for fibrocystin/polyductin (FPC), a type of membrane-associated receptor-like protein which is localised to primary cilia. FPC is predominantly expressed in the apical domain of renal tubule epithelial cells. This protein may play a vital role in renal collecting duct and biliary differentiation (Ward et al. 2002).

3.1.6 Wilson's Disease (WD)

Wilson's disease is also an autosomal recessive disorder and featured with toxic accumulation of copper in the liver, brain and other organs manifested into end-stage liver disease and acute liver failure (Tanzi et al. 1993). However, disease also has a wide spectrum of asymptomatic individuals. Symptoms are usually seen in the first decades of life, with the majority of cases in the ages of 5–35 years.

ATP7B gene present on chromosome 13 and containing 21 exons is responsible for WD (Koch et al. 2001). ATP7B is a copper-transporting P-type adenosine triphosphatase (ATPase) which is required for efficient excretion of the copper into bile. Hepatocytes release copper after integration into ceruloplasmin, a glycoprotein that carries six atoms of copper per molecule. In the absence of ATP7B function, copper is not incorporated efficiently resulting in accumulation of copper within the hepatocytes.

To date, over 500 mutations in ATP7B have been reported. The most common mutation is H1069Q, which has been seen in 35–45% European population (Thomas et al. 1995). Another mutation, R778L, is widely distributed among Asians (20%) (Nanji et al. 1997). In Indian patients, mostly C271X was found to be common with a frequency of 19% of the total mutations (Gupta et al. 2005), and the presence of R778L or H1069Q is completely absent. All 21 exons and promoter regions of ATP7B can be sequenced, by multiplex polymerase chain reaction (PCR), which has become standard practice in clinical molecular diagnosis (Fig. 3.7).

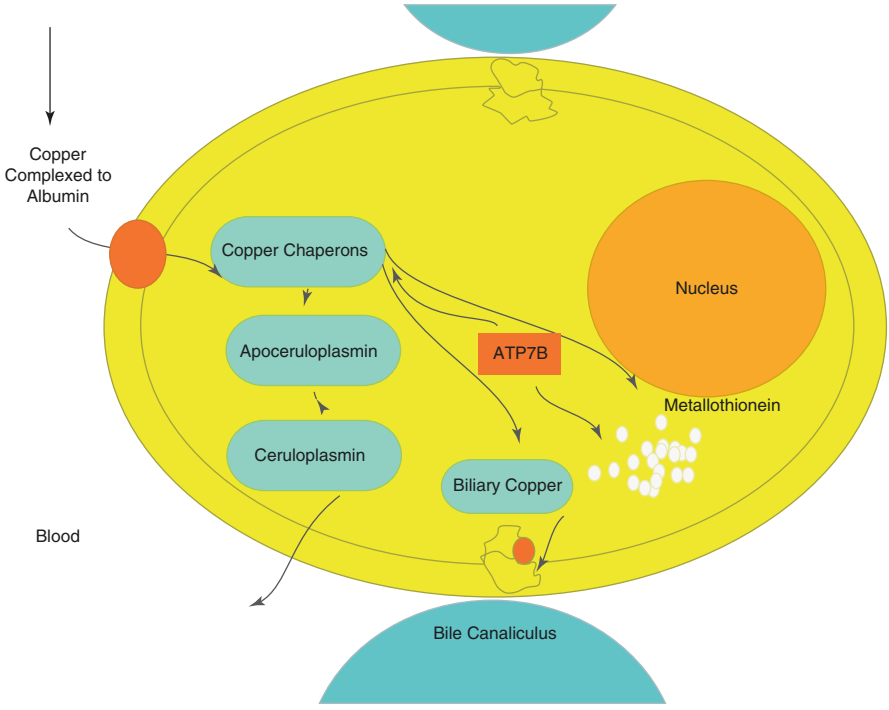

Fig. 3.7 Schematic representation of copper metabolism within the hepatocyte

3.1.7 Hereditary Haemochromatosis

Hereditary haemochromatosis (HH) is the most common autosomal recessive disorder mainly reported in Caucasians. HH is characterised by highly abnormal uptake of iron from the gastrointestinal tracts, and it manifests initially with symptoms of discomforts in joints, general fatigue, low libido and abdominal pain. If HH is left untreated, it leads to hypogonadism, cardiomyopathy liver fibrosis and end-stage liver diseases like cirrhosis and hepatocellular carcinoma. Therefore, early diagnosis of hereditary haemochromatosis and larger educational programs are required to increase people's awareness about hereditary haemochromatosis. Two major mutations of the human haemochromatosis gene, C282Y and His63Asp (H63D), cause iron overload due to less production of hepcidin. Hepcidin is regulator of iron homeostasis in hepatocytes, and insufficient hepcidin expression results to too much iron absorption and deposition in tissues, causing multiple organ damage and failure (Pelusi et al. 2014; Emanuele et al. 2014; Vujic 2014). However, despite increased store of iron in body, enterocyte continue strive to absorb dietary iron.

There is close association between HLA-A3 and HFE as both are on chromosome 6. There is also role of other mutations including TFR2 (transferrin receptor 2), SLC40A1 (encoding ferroportin), HAMP (encoding hepcidin) and HJV (encoding hemojuvelin), A736V variant of TMPRSS6 (regulating hepcidin) in hereditary haemochromatosis. However, genetic data exclude role of transferrin, transferrin receptor and ferritin since they are not expressed on chromosome 6 (Wang et al. 2013; Roetto et al. 2003; Papanikolaou et al. 2004; Panigrahi et al. 2006; Potekhina et al. 2005; Zamani et al. 2012; Del-Castillo-Rueda et al. 2012; Valenti et al. 2012a) (Table 3.4).

Table 3.4 Genes associated with iron accumulation

Gene	Loci	Mutations	Wild function
HFE	6p22.2	C282Y, recessive	Regulate interaction of transferrin receptor with transferrin
		H63D, recessive	
		Arg455Gln	
TFR2	7q22	Glu60Xaa,	Encoding transferrin receptor 2
		Met172Lys,	
		loss Ala-Val-Ala-Gln at 594-597	
SLC40A1	2q32	Asn144 His, dominant	Encoding ferroportin
HAMP	19q13.1	Frameshift mutation, G71D, recessive	Encoding hepcidin
HJV(HFE2)	1q21.1	L101P, G320 V, recessive	Encoding hemojuvelin
TMPRSS6	22q12.3	A736V	Encoding matriptase 2
PNPLA3	22q13.31	I148M	Encoding multifunctional enzyme having both triacylgycerol lipase and acylglycerol O-acyltransferase activity
TNF-alpha	6p21.3	−308G > A	Encoding multifunctional pro-inflammatory cytokine
BMP2	20p12	rs235756	

Other than idiopathic haemochromatosis in India and majority of Asians, frequencies of C282Y mutations are near to zero. Five percent of Russians showed homozygosity for C282Y with biochemical and clinical signs of HH (Roetto et al. 2003). Alcoholic liver disease patients also showed association with C282Y and H63D mutations, and mostly are homozygous for the H63D with increased total and low-density lipoprotein cholesterol (Raszeja-Wyszomirska et al. 2010).

Heterozygotic mutations of H63D and TFR2 (transferrin receptor 2) genes were found to be more common in Iranian patients; however, these mutations were not found to be significantly associated with severity of presentation in HH patients (Del-Castillo-Rueda et al. 2012).

However, in the rat model of hereditary haemochromatosis, sequencing of TFR2 revealed a novel Ala679Gly polymorphism which is highly conserved residue and, therefore, showed the involvement of TFR2 in haemochromatosis (Santos et al. 2011). In the same way, the importance of A736V and TMPRSS6 polymorphism which regulates hepcidin levels was associated with HH and hepatocellular carcinoma (Valenti et al. 2012a; Raszeja-Wyszomirska et al. 2010; Santos et al. 2011). Many other gene variants are also found to be associated with HH, as I148M protein variant of PNPLA3 and TNF-alpha −308G > A allelic variant was found to be associated with steatosis, fibrosis and cirrhosis in patients with C282Y+/+ hereditary haemochromatosis (Bartnikas et al. 2013; Valenti et al. 2012b; Krayenbuehl et al. 2006).

3.1.8 Alcoholic Liver Disease

Alcohol overconsumption for an extended period of time leads to alcoholic liver disease (ALD), which is the major cause of chronic liver disease in adults. The metabolism of alcohol is known to generate reactive intermediates which contribute to cell and tissue damage by altering various cell signalling pathways. In ALD, manifestation starts from simple steatosis, fibrosis to cirrhosis and hepatocellular carcinoma. Alcohol dependence (AD) is a heritable trait (>50%) and is strongly correlated with susceptibility to excessive alcohol consumption.

The metabolism of alcohol occurs mainly in the intestine and the liver, where alcohol is changed into acetaldehyde by cytochrome P450 2E1 (CYP2E1) and alcohol dehydrogenase (ADH). Acetaldehyde is a very toxic cancer-inducing substance. However, most of the acetaldehyde is converted into acetate which is less toxic.

Polymorphism in alcohol dehydrogenase (ADH) is very important, ADH2*2 (rs1229984), ADH3*2 (rs698) and aldehyde dehydrogenase (ALDH), which converts acetaldehyde to acetate. In human genome, out of 19 ALDH genes, ALDH2 which is mainly present in the mitochondrial matrix has an important role in alcohol metabolism. Candidate gene and linkage studies revealed that the alcohol dehydrogenase 1B (ADH1B) and aldehyde dehydrogenase 2 (ALDH2) genes mediate the risk for alcoholism in Asians. In many ethnic populations, increase in ADH1B*2 and ALDH2*1/*1 alleles showed genetic susceptibility for cirrhosis (Li et al. 2012). Other important mutations in Asia including India and other continents are

CYP2E1*1D, CYP2E1*5 (rs3813867 and rs2031920), TNF-α (rs1800629), TNF-α (rs361525), IL-1β (rs3087258), CD-14 (rs2569190), IL-10 (rs1800872) and PNPLA3 (rs738409), glutathione S-transferase P1 (GSTP1-Val allele) in alcoholics (Yokoyama et al. 2013; Roy et al. 2012; Dutta 2013; Bansal et al. 2013; Wan et al. 1998; Liu et al. 2012; Zeng et al. 2013; Agrawal and Bierut 2012; Nischalke et al. 2013; Giby and Ajith 2014; Adams et al. 2005). CYP2E1 is a major determinant of alcohol-induced toxicity in the liver, intestine, brain and other peripheral tissues where it is expressed (Liu et al. 2012). Meta-analysis suggested that CYP2E1 polymorphism is associated with alcohol-induced steatosis, hepatitis and fibrosis (Wan et al. 1998). Alcoholic cirrhotics also showed statistically significant increase in PNPLA3 allele (Yokoyama et al. 2013) (Table 3.5).

In addition, neurotransmitter aminobutyric acid receptor gene appears to have a role in the development of alcohol dependence (Dutta 2013). In fact, many intrahepatic chemokines and their receptors are upregulated in alcohol-induced liver

Table 3.5 Candidates genes associated with ALD

Function	Gene[a]
Alcohol metabolism	ADH1B
	ADH1C
	CYP2E1
	CYP1A1
	ALDH2
Oxidative stress	GSTM1
	GSTP1
	GSTT1
	MnSOD
	NAT
	HFE
Immune reactions	TNF-α
	IL-10
	IL-1R
	IL-1B
Fibrosis-associated factors	TGFβ1
	MMP3
Modulation of Steatosis	PPARg
	MTP
	ApoE
Other	NFkB1
	DRD2
	SLC6A4
	GSTP1
	PNPLA3
	GABRA2

[a]Some of these genes have poor association with the ALD

fibrosis. Polymorphism of CXC-chemokine ligand 1 (CXCL1), inflammatory cytokine gene (rs4074) is also considered as an independent factor for cirrhosis and HCC in ALD patients, even in the absence of liver damage (Zeng et al. 2013; Agrawal and Bierut 2012; Nischalke et al. 2013).

3.1.9 Non-alcoholic Fatty Liver Disease (NAFLD)

NAFLD is affecting approximately 20–40% of the population in western countries and becoming increasingly frequent in Asian subcontinent especially in India. NAFLD has wide spectrum, from steatosis, non-alcoholic steatohepatitis (NASH) and advanced fibrosis to end-stage liver disease cirrhosis and hepatocellular carcinoma (HCC) (Giby and Ajith 2014; Adams et al. 2005; Angulo 2002; Clark et al. 2002; Chitturi et al. 2004).

It is a metabolic syndrome, its pathophysiology centrally associated with insulin resistance (Day 2006; Guerrero et al. 2009). It is evident that majority of patients with NAFLD develop hepatic steatosis (Wilfred de Alwis and Day 2007), and only a small group develops the more advanced forms of non-alcoholic steatohepatitis (NASH).

However, other than obesity and insulin resistance, many factors are responsible to develop NAFLD or its advanced form, which still remain unclear. Environmental and genetic factors also play a part in developing the disease.

Therefore, determining the genetic factors which predispose an individual for developing NAFLD might help in taking preventive strategies in those at higher risk (Browning et al. 2004b). These genetic variants associated with disease risk should lead to the development of non-invasive biomarkers and identification of novel treatment targets.

Regardless of, equal susceptibility rate of obesity in African–Americans and European Americans of Hispanic origin has lower incidence of both steatosis and cryptogenic cirrhosis (Guerrero et al. 2009). However, there is also increasing data suggesting that US Hispanics are also reported to be more susceptible to NAFLD than those of European descent (Browning et al. 2004a; Williams et al. 2011).

Non-alcoholic steatohepatitis (NASH) is the progressive form of NAFLD which often leads from steatosis to steatohepatitis and cirrhosis and may progress to HCC. In steatosis, liver cell large vacuoles accumulate triglyceride fat via the process of lipogenesis. Therefore, mediators released from adipose tissue such as adipokines play integral role in NAFLD. Adipokines regulate homeostasis in maintenance of energy through lipogenesis, lipolysis and fatty acid oxidation. Fatty acid oxidation in the liver is activated via peroxisome proliferator-activated receptor (PPAR)-α by leptin and adiponectin (Williams et al. 2011; Namikawa et al. 2004; Bernard et al. 2000).

It has been shown that patients with NAFLD show mutations in microsomal triglyceride transfer protein (MTP). MTP is critical for the production of very-low-density lipoprotein (VLDL) in the liver as well as in the intestine. G/T SNP at

position −493 in the promoter region of MTP has been linked with low expression of MTP levels resulting in failure to secrete triacylglycerol (TG) in the liver. NAFLD patients with G/G genotype have increased chance of steatosis and NASH compared to T/T genotype (Namikawa et al. 2004; Bernard et al. 2000; Oliveira et al. 2010).

Partial loss of function in the *phosphatidylethanolamine N-methyltransferase* (*PEMT*) gene was also reported to be involved in the production of phosphatidylcholine needed for VLDL synthesis. In two different studies, Japanese and American patients with biopsy-proven NASH showed increased frequency of V175M allele of PEMT gene than the controls (Dong et al. 2007; Song et al. 2005). These patients had the lower BMI, indicating that they were more genetically predisposed to develop lean NASH (Song et al. 2005). SNPs in other genes which regulate intrahepatic free fatty acid (FFA) and TG synthesis, storage and export are also attractive candidates for NAFLD. In addition, pregnane X receptor (PXR) gene, a well-defined transcription factor, also has role in lipid homeostasis and hepatic detoxification mechanisms (Zhang et al. 2008; Zhou et al. 2006). SNPs in PXR gene were significantly associated with NAFLD and considered as predictor of disease severity (Sookoian et al. 2010). SNP in the promoter region of binding site of transcription factor HNF-4 affects the expression of PXR, *CYP3A4* and *ABCB1* genes. Therefore, there is a high possibility of its effect on lipid homeostasis which needs to be investigated.

Apolipoprotein (APO) genes, a constituent of VLDL, are also appealing candidates for their roles in NAFLD vulnerability and development; however, data is very limited. In Indian population, APOC3 is associated with triglyceridemia and considered as one of the strongest factors for NAFLD (Salamone et al. 2010). Two variant alleles in the promoter region of APOC3 are in linkage disequilibrium and associated with NAFLD. However, contradictory to this, in Caucasians, there was no association between APOC3 variants and hepatic TG content or insulin resistance (Kozlitina et al. 2011) (Table 3.6).

Table 3.6 Candidate genes associated with the NAFLD

Gene	SNP
APOC3	T455C, C482T
MTP	498G/T
PEMT	V175M
PNPLA3	rs738408, rs738409
FDFT1	rs2645424
COL13A1	rs1227756
EFCAB4B	rs887304
PZP	rs6487679
NCAN	rs2228603
PPP1R3B	rs4240624
GCKR	rs780094
LYPLAL1	rs12137855

3.2 Viral Hepatitis

Acute or chronic viral hepatitis due to hepatotropic viruses like hepatitis A, B, C, D and E is the most common cause of liver disease worldwide. In addition to hepatic viruses, other viruses like herpes simplex, cytomegalovirus, Epstein–Barr virus or yellow fever also cause liver inflammation (hepatitis).

HAV spreads by the faecal-oral route and is often related with ingestion of contaminated food or water. It is responsible for an acute form of hepatitis and does not cause chronicity. The patient's immune system develops antibodies against HAV which builds up lifelong immunity.

Hepatitis B virus belongs to Hepadnaviridae virus family and causes both acute and chronic hepatitis. More than 350 million people are chronic carrier of hepatitis B infection. Ninety to 95% individuals clear the virus by potent immunity; however, 5–10% acute infection turns into chronic hepatitis, and adults are unable to clear the virus. HBV infection is transmitted through blood transfusion, tattoos, unsafe sex or HBV-infected mother to child through transplacental crossing. But, in about half of cases, the source of infection cannot be determined.

Chronic hepatitis B patients develop specific antibodies against hepatitis B, but the titers of antibodies are not enough to clear the infection. Therefore, continued replication of virus and small amount of antibodies is the main cause of the immune complex disease in these patients.

Clinical outcome in HBV infection is mainly decided by viral, host immunological and genetic factors. HBV infection influences many cellular processes through genetic instability such as:

- Virus binding, entry, fusion, with cell membrane
- Modulation of host immune response
- Cause of pathological alterations in the liver
- Development of liver cirrhosis and HCC
- Mother-to-infant vertical transmission resistance to antiviral therapies

Many studies have reported association of HBV infection with HLA. HLA class II alleles such as DRB1*1302 or HLA-DR13 or DQA1*0501-DQB1*0301-DQB1*1102 are associated with acute and/or chronic HBV infection. Several pro-inflammatory (Th1) cytokines like IL-2, IFN-γ and TNF-α have a major role in boosting host immunity leading to viral clearance. On the contrary, IL-10 cytokine (Th2) serves as a potent inhibitor of Th1 effector cells in HBV infection.

Polymorphism in IL-16 gene was found to be associated with HBV-related HCC. TG and GG genotypes of rs11556218 T/G SNP in Il-16 were significantly associated with increased risk of HBV-related HCC compared to TT genotype (OR = 1.96 and OR = 3.33). Previous data revealed that subjects with the G allele appeared to have lower susceptibility to chronic hepatitis B infection than those with the T allele (OR = 0.46). Under the dominant model, genotype TG + GG seemed to have lesser association to chronic hepatitis B (OR = 0.44) (Li et al. 2011).

Another SNP, rs4072111C/T, with TT genotype prevalence was found to be associated with risk of HBV-related HCC compared to CC genotype (OR = 6.67).

Programmed death (PD1) receptor and ligand are important in viral clearance as well as treatment outcomes. Recently, PD1 antagonist is being considered as main therapy along with antivirals and interferons. It is observed that interaction between PD1 and T-cell immunoglobulin (Tim-3) is important in immune dysfunctions in chronic HBV infection. It was observed that PD1 and TIM3 polymorphisms differentially and interactively predispose cirrhosis and HCC in HBV-infected patients (Li et al. 2013) (Table 3.7).

In the large meta-analysis, it is observed that polymorphisms in *PD1* and *TIM3 genes at position* +8669 and −1516 have profound effect on HBV chronicity.

Multivariate analysis showed that, in addition to *PD1* +8669 genotype AA and *TIM3*-1516 genotypes GT + TT, gender, age, ALT, albumin and HBV DNA were associated with HBV cirrhosis compared to patients without cirrhosis. The combined presence of *PD1* +8669 AA/*TIM3*-1516 GT or TT was higher in cirrhosis and HCC pooled patients than in patients without cirrhosis (OR, 2.326; $p = 0.020$) and HCC.

Other important polymorphisms in the coding region of *MBP* gene is reported to be involved in chronic HBV infection. However, since genetic interactions are complex, reports from different studies showed inconsistencies with respect to the effects of host genetic factors on HBV clearance or persistence.

Therefore, it is unlikely that a single allelic variant is responsible for HBV persistence or clearance. It may be that collective influence of several single-nucleotide polymorphisms (SNPs) or haplotype(s) underlies the synergistic protection against HBV (Wang 2003). Combined panel of genes in large cohort of HBV-infected patients may provide insight into the pathogenesis of HBV infection and a unique rationale for new methods of diagnosis and therapeutic strategies.

Table 3.7 Genes associated with viral hepatitis clearance and persistence

Viral hepatitis	HLA	Cytokine	IL	Other
B	DRB1*1302,DR2, DR6,DR13,DQA1*0501-DQB1*0301-DQB1*1102	TNF-alpha, IFN-gamma	IL-2 IL16 IL28B IL18	PD-1 Tim-3 SPP1 CCR5 GNLY MBL
C	DQB1*0301, DQB1*0201, DQB1*03, DQA1*03, DRB1*01, DRB1*0101,DRB1*0301, DRB1*04,DRB1*0401, DRB1*11, DRB1*12, DRB1*1101,DRB1*1104, DRB1*1302,DRB3*03, DRQ1*1101,DQA1*03, DQA1*0501,DQB1*0302, A*1101, B*57,Cw*0102, DR15	TNF-alpha, TGF-b1	IL28B	HFE TYK2 SVR

Conclusions

Present genetic data of the distributions and functions of the implicated allele polymorphisms mostly lie in small patient groups and in few ethnic groups. Future elaborate studies should be constituted internationally including multi-cohorts to clarify gene associations and identify other potential candidate genes in liver diseases.

Competing Interests There are no competing interests to be disclosed.

References

Adams LA, Lymp JF, St Sauver J, Sanderson SO, Lindor KD, Feldstein A, Angulo P. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. Gastroenterology. 2005;129(1):113–21.

Agrawal A, Bierut LJ. Identifying genetic variation for alcohol dependence. Alcohol Res. 2012;34(3):274–81.

Angulo P. Nonalcoholic fatty liver disease. N Engl J Med. 2002;346(16):1221–31. doi:10.1056/NEJMra011775.

Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science. 1999;284(5415):770–6.

Bansal S, Anandatheerthavarada HK, Prabu GK, Milne GL, Martin MV, Guengerich FP, Avadhani NG. Human cytochrome P450 2E1 mutations that alter mitochondrial targeting efficiency and susceptibility to ethanol-induced toxicity in cellular models. J Biol Chem. 2013;288(18):12627–44. doi:10.1074/jbc.M113.452367.

Bartnikas TB, Wildt SJ, Wineinger AE, Schmitz-Abe K, Markianos K, Cooper DM, Fleming MD. A novel rat model of hereditary hemochromatosis due to a mutation in transferrin receptor 2. Comp Med. 2013;63(2):143–55.

Bernard S, Touzet S, Personne I, Lapras V, Bondon PJ, Berthezene F, Moulin P. Association between microsomal triglyceride transfer protein gene polymorphism and the biological features of liver steatosis in patients with type II diabetes. Diabetologia. 2000;43(8):995–9. doi:10.1007/s001250051481.

Boutten A, Venembre P, Seta N, Hamelin J, Aubier M, Durand G, Dehoux MS. Oncostatin M is a potent stimulator of alpha1-antitrypsin secretion in lung epithelial cells: modulation by transforming growth factor-beta and interferon-gamma. Am J Respir Cell Mol Biol. 1998;18(4):511–20. doi:10.1165/ajrcmb.18.4.2772.

Browning JD, Kumar KS, Saboorian MH, Thiele DL. Ethnic differences in the prevalence of cryptogenic cirrhosis. Am J Gastroenterol. 2004a;99(2):292–8.

Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, Grundy SM, Hobbs HH. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. Hepatology. 2004b;40(6):1387–95. doi:10.1002/hep.20466.

Bull LN, Juijn JA, Liao M, van Eijk MJ, Sinke RJ, Stricker NL, DeYoung JA, Carlton VE, Baharloo S, Klomp LW, Abukawa D, Barton DE, Bass NM, Bourke B, Drumm B, Jankowska I, Lovisetto P, McQuaid S, Pawlowska J, Tazawa Y, Villa E, Tygstrup N, Berger R, Knisely AS, Freimer NB, et al. Fine-resolution mapping by haplotype evaluation: the examples of PFIC1 and BRIC. Hum Genet. 1999;104(3):241–8.

Carlton VE, Knisely AS, Freimer NB. Mapping of a locus for progressive familial intrahepatic cholestasis (Byler disease) to 18q21-q22, the benign recurrent intrahepatic cholestasis region. Hum Mol Genet. 1995;4(6):1049–53.

Carrell RW, Lomas DA. Alpha1-antitrypsin deficiency—a model for conformational diseases. N Engl J Med. 2002;346(1):45–53. doi:10.1056/NEJMra010772.

Chitturi S, Farrell GC, George J. Non-alcoholic steatohepatitis in the Asia-Pacific region: future shock? J Gastroenterol Hepatol. 2004;19(4):368–74.

Clark JM, Brancati FL, Diehl AM. Nonalcoholic fatty liver disease. Gastroenterology. 2002;122(6):1649–57.

Crosnier C, Attie-Bitach T, Encha-Razavi F, Audollent S, Soudy F, Hadchouel M, Meunier-Rotival M, Vekemans M. JAGGED1 gene expression during human embryogenesis elucidates the wide phenotypic spectrum of Alagille syndrome. Hepatology. 2000;32(3):574–81. doi:10.1053/jhep.2000.16600.

Day CP. From fat to inflammation. Gastroenterology. 2006;130(1):207–10. doi:10.1053/j.gastro.2005.11.017.

de Vree JM, Jacquemin E, Sturm E, Cresteil D, Bosma PJ, Aten J, Deleuze JF, Desrochers M, Burdelski M, Bernard O, Oude Elferink RP, Hadchouel M. Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. Proc Natl Acad Sci USA. 1998;95(1):282–7.

Delaunay JL, Durand-Schneider AM, Delautier D, Rada A, Gautherot J, Jacquemin E, Ait-Slimane T, Maurice M. A missense mutation in ABCB4 gene involved in progressive familial intrahepatic cholestasis type 3 leads to a folding defect that can be rescued by low temperature. Hepatology. 2009;49(4):1218–27. doi:10.1002/hep.22775.

Del-Castillo-Rueda A, Moreno-Carralero MI, Cuadrado-Grande N, Alvarez-Sala-Walther LA, Enriquez-de-Salamanca R, Mendez M, Moran-Jimenez MJ. Mutations in the HFE, TFR2, and SLC40A1 genes in patients with hemochromatosis. Gene. 2012;508(1):15–20. doi:10.1016/j.gene.2012.07.069.

Dhorne-Pollet S, Deleuze JF, Hadchouel M, Bonaiti-Pellie C. Segregation analysis of Alagille syndrome. J Med Genet. 1994;31(6):453–7.

Dixon PH, Weerasekera N, Linton KJ, Donaldson O, Chambers J, Egginton E, Weaver J, Nelson-Piercy C, de Swiet M, Warnes G, Elias E, Higgins CF, Johnston DG, McCarthy MI, Williamson C. Heterozygous MDR3 missense mutation associated with intrahepatic cholestasis of pregnancy: evidence for a defect in protein trafficking. Hum Mol Genet. 2000;9(8):1209–17.

Dong H, Wang J, Li C, Hirose A, Nozaki Y, Takahashi M, Ono M, Akisawa N, Iwasaki S, Saibara T, Onishi S. The phosphatidylethanolamine N-methyltransferase gene V175M single nucleotide polymorphism confers the susceptibility to NASH in Japanese population. J Hepatol. 2007;46(5):915–20. doi:10.1016/j.jhep.2006.12.012.

Dutta AK. Genetic factors affecting susceptibility to alcoholic liver disease in an Indian population. Ann Hepatol. 2013;12(6):901–7.

Emanuele D, Tuason I, Edwards QT. HFE-associated hereditary hemochromatosis: overview of genetics and clinical implications for nurse practitioners in primary care settings. JAm Assoc Nurse Pract. 2014;26(3):113–22. doi:10.1002/2327-6924.12106.

Giby VG, Ajith TA. Role of adipokines and peroxisome proliferator-activated receptors in nonalcoholic fatty liver disease. World J Hepatol. 2014;6(8):570–9. doi:10.4254/wjh.v6.i8.570.

Guerrero R, Vega GL, Grundy SM, Browning JD. Ethnic differences in hepatic steatosis: an insulin resistance paradox? Hepatology. 2009;49(3):791–801. doi:10.1002/hep.22726.

Gunay-Aygun M, Font-Montgomery E, Lukose L, Tuchman Gerstein M, Piwnica-Worms K, Choyke P, Daryanani KT, Turkbey B, Fischer R, Bernardini I, Sincan M, Zhao X, Sandler NG, Roque A, Douek DC, Graf J, Huizing M, Bryant JC, Mohan P, Gahl WA, Heller T. Characteristics of congenital hepatic fibrosis in a large cohort of patients with autosomal recessive polycystic kidney disease. Gastroenterology. 2013;144(1):112–21. e112 doi:10.1053/j.gastro.2012.09.056.

Gupta A, Aikath D, Neogi R, Datta S, Basu K, Maity B, Trivedi R, Ray J, Das SK, Gangopadhyay PK, Ray K. Molecular pathogenesis of Wilson disease: haplotype analysis, detection of prevalent mutations and genotype-phenotype correlation in Indian patients. Hum Genet. 2005;118(1):49–57. doi:10.1007/s00439-005-0007-y.

Hayashi H, Takada T, Suzuki H, Akita H, Sugiyama Y. Two common PFIC2 mutations are associated with the impaired membrane trafficking of BSEP/ABCB11. Hepatology. 2005;41(4):916–24. doi:10.1002/hep.20627.

Jacquemin E. Role of multidrug resistance 3 deficiency in pediatric and adult liver disease: one gene for three diseases. Semin Liver Dis. 2001;21(4):551–62. doi:10.1055/s-2001-19033.

Jacquemin E, De Vree JM, Cresteil D, Sokal EM, Sturm E, Dumont M, Scheffer GL, Paul M, Burdelski M, Bosma PJ, Bernard O, Hadchouel M, Elferink RP. The wide spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. Gastroenterology. 2001;120(6):1448–58.

Jansen PL, Strautnieks SS, Jacquemin E, Hadchouel M, Sokal EM, Hooiveld GJ, Koning JH, De Jager-Krikken A, Kuipers F, Stellaard F, Bijleveld CM, Gouw A, Van Goor H, Thompson RJ, Muller M. Hepatocanalicular bile salt export pump deficiency in patients with progressive familial intrahepatic cholestasis. Gastroenterology. 1999;117(6):1370–9.

Kalsheker N, Morley S, Morgan K. Gene regulation of the serine proteinase inhibitors alpha1-antitrypsin and alpha1-antichymotrypsin. Biochem Soc Trans. 2002;30(2):93–8.

Klomp LW, Bull LN, Knisely AS, van Der Doelen MA, Juijn JA, Berger R, Forget S, Nielsen IM, Eiberg H, Houwen RH. A missense mutation in FIC1 is associated with greenland familial cholestasis. Hepatology. 2000;32(6):1337–41. doi:10.1053/jhep.2000.20520.

Klomp LW, Vargas JC, van Mil SW, Pawlikowska L, Strautnieks SS, van Eijk MJ, Juijn JA, Pabon-Pena C, Smith LB, DeYoung JA, Byrne JA, Gombert J, van der Brugge G, Berger R, Jankowska I, Pawlowska J, Villa E, Knisely AS, Thompson RJ, Freimer NB, Houwen RH, Bull LN. Characterization of mutations in ATP8B1 associated with hereditary cholestasis. Hepatology. 2004;40(1):27–38. doi:10.1002/hep.20285.

Koch C, Cuppens H, Rainisio M, Madessani U, Harms H, Hodson M, Mastella G, Navarro J, Strandvik B, McKenzie S, Investigators of the E. European epidemiologic registry of cystic fibrosis (ERCF): comparison of major disease manifestations between patients with different classes of mutations. Pediatr Pulmonol. 2001;31(1):1–12.

Kozlitina J, Boerwinkle E, Cohen JC, Hobbs HH. Dissociation between APOC3 variants, hepatic triglyceride content and insulin resistance. Hepatology. 2011;53(2):467–74. doi:10.1002/hep.24072.

Krayenbuehl PA, Maly FE, Hersberger M, Wiesli P, Himmelmann A, Eid K, Greminger P, Vetter W, Schulthess G. Tumor necrosis factor-alpha −308G>A allelic variant modulates iron accumulation in patients with hereditary hemochromatosis. Clin Chem. 2006;52(8):1552–8. doi:10.1373/clinchem.2005.065417.

Lang C, Meier Y, Stieger B, Beuers U, Lang T, Kerb R, Kullak-Ublick GA, Meier PJ, Pauli-Magnus C. Mutations and polymorphisms in the bile salt export pump and the multidrug resistance protein 3 associated with drug-induced liver injury. Pharmacogenet Genomics. 2007;17(1):47–60. doi:10.1097/01.fpc.0000230418.28091.76.

Li S, Deng Y, Chen ZP, Huang S, Liao XC, Lin LW, Li H, Peng T, Qin X, Zhao JM. Genetic polymorphism of interleukin-16 influences susceptibility to HBV-related hepatocellular carcinoma in a Chinese population. Infect Genet Evol. 2011;11(8):2083–8. doi:10.1016/j.meegid.2011.09.025.

Li L, Krantz ID, Deng Y, Genin A, Banta AB, Collins CC, Qi M, Trask BJ, Kuo WL, Cochran J, Costa T, Pierpont ME, Rand EB, Piccoli DA, Hood L, Spinner NB. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat Genet. 1997;16(3):243–51. doi:10.1038/ng0797-243.

Li Z, Li N, Zhu Q, Zhang G, Han Q, Zhang P, Xun M, Wang Y, Zeng X, Yang C, Liu Z. Genetic variations of PD1 and TIM3 are differentially and interactively associated with the development of cirrhosis and HCC in patients with chronic HBV infection. Infect Genet Evol. 2013;14:240–6. doi:10.1016/j.meegid.2012.12.008.

Li D, Zhao H, Gelernter J. Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. Hum Genet. 2012;131(5):725–37. doi:10.1007/s00439-011-1116-4.

Liu Y, Zhou LY, Meng XW. Genetic polymorphism of two enzymes with alcoholic liver disease in Northeast China. Hepatogastroenterology. 2012;59(113):204–7. doi:10.5754/hge10136.

Namikawa C, Shu-Ping Z, Vyselaar JR, Nozaki Y, Nemoto Y, Ono M, Akisawa N, Saibara T, Hiroi M, Enzan H, Onishi S. Polymorphisms of microsomal triglyceride transfer protein gene and manganese superoxide dismutase gene in non-alcoholic steatohepatitis. J Hepatol. 2004;40(5):781–6. doi:10.1016/j.jhep.2004.01.028.

Nanji MS, Nguyen VT, Kawasoe JH, Inui K, Endo F, Nakajima T, Anezaki T, Cox DW. Haplotype and mutation analysis in Japanese patients with Wilson disease. Am J Hum Genet. 1997;60(6):1423–9.

Nischalke HD, Berger C, Lutz P, Langhans B, Wolter F, Eisenhardt M, Kramer B, Kokordelis P, Glassner A, Muller T, Rosendahl J, Fischer J, Berg T, Grunhage F, Leifeld L, Soyka M, Nattermann J, Sauerbruch T, Stickel F, Spengler U. Influence of the CXCL1 rs4074 A allele on alcohol induced cirrhosis and HCC in patients of European descent. PLoS One. 2013;8(11):e80848. doi:10.1371/journal.pone.0080848.

Oliveira CP, Stefano JT, Cavaleiro AM, Zanella Fortes MA, Vieira SM, Rodrigues Lima VM, Santos TE, Santos VN, de Azevedo Salgado AL, Parise ER, Ferreira Alves VA, Carrilho FJ, Correa-Giannella ML. Association of polymorphisms of glutamate-cystein ligase and microsomal triglyceride transfer protein genes in non-alcoholic fatty liver disease. J Gastroenterol Hepatol. 2010;25(2):357–61. doi:10.1111/j.1440-1746.2009.06001.x.

Panigrahi I, Ahmad F, Kapoor R, Sharma PK, Makharia G, Saxena R. Evidence for non-HFE linked hemochromatosis in Asian Indians. Indian J Med Sci. 2006;60(12):491–5.

Papanikolaou G, Samuels ME, Ludwig EH, MacDonald ML, Franchini PL, Dube MP, Andres L, MacFarlane J, Sakellaropoulos N, Politou M, Nemeth E, Thompson J, Risler JK, Zaborowska C, Babakaiff R, Radomski CC, Pape TD, Davidas O, Christakis J, Brissot P, Lockitch G, Ganz T, Hayden MR, Goldberg YP. Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. Nat Genet. 2004;36(1):77–82. doi:10.1038/ng1274.

Pelusi S, Rametta R, Della Corte C, Congia R, Dongiovanni P, Pulixi EA, Fargion S, Fracanzani AL, Nobili V, Valenti L. Juvenile hemochromatosis associated with heterozygosity for novel hemojuvelin mutations and with unknown cofactors. Ann Hepatol. 2014;13(5):568–71.

Potekhina ES, Lavrov AV, Samokhodskaya LM, Efimenko AY, Balatskiy AV, Baev AA, Litvinova MM, Nikitina LA, Shipulin GA, Bochkov NP, Tkachuk VA, Bochkov VN. Unique genetic profile of hereditary hemochromatosis in Russians: high frequency of C282Y mutation in population, but not in patients. Blood Cells Mol Dis. 2005;35(2):182–8. doi:10.1016/j.bcmd.2005.06.012.

Raszeja-Wyszomirska J, Kurzawski G, Zawada I, Suchy J, Lubinski J, Milkiewicz P. HFE gene mutations in patients with alcoholic liver disease. A prospective study from northwestern Poland. Pol Arch Med Wewn. 2010;120(4):127–31.

Roetto A, Papanikolaou G, Politou M, Alberti F, Girelli D, Christakis J, Loukopoulos D, Camaschella C. Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. Nat Genet. 2003;33(1):21–2. doi:10.1038/ng1053.

Rosmorduc O, Hermelin B, Boelle PY, Parc R, Taboury J, Poupon R. ABCB4 gene mutation-associated cholelithiasis in adults. Gastroenterology. 2003;125(2):452–9.

Roy N, Mukhopadhyay I, Das K, Pandit P, Majumder PP, Santra A, Datta S, Banerjee S, Chowdhury A. Genetic variants of TNFalpha, IL10, IL1beta, CTLA4 and TGFbeta1 modulate the indices of alcohol-induced liver injury in East Indian population. Gene. 2012;509(1):178–88. doi:10.1016/j.gene.2012.07.077.

Salamone F, Galvano F, Li Volti G. Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. N Engl J Med. 2010;363(2):194–5.

Santos PC, Cancado RD, Pereira AC, Schettert IT, Soares RA, Pagliusi RA, Hirata RD, Hirata MH, Teixeira AC, Figueiredo MS, Chiattone CS, Krieger JE, Guerra-Shinohara EM. Hereditary hemochromatosis: mutations in genes involved in iron homeostasis in Brazilian patients. Blood Cells Mol Dis. 2011;46(4):302–7. doi:10.1016/j.bcmd.2011.02.008.

Song J, da Costa KA, Fischer LM, Kohlmeier M, Kwock L, Wang S, Zeisel SH. Polymorphism of the PEMT gene and susceptibility to nonalcoholic fatty liver disease (NAFLD). FASEB J. 2005;19(10):1266–71. doi:10.1096/fj.04-3580com.

Sookoian S, Castano GO, Burgueno AL, Gianotti TF, Rosselli MS, Pirola CJ. The nuclear receptor PXR gene variants are associated with liver injury in nonalcoholic fatty liver disease. Pharmacogenet Genomics. 2010;20(1):1–8. doi:10.1097/FPC.0b013e328333a1dd.

Spinner NB, Colliton RP, Crosnier C, Krantz ID, Hadchouel M, Meunier-Rotival M. Jagged1 mutations in alagille syndrome. Hum Mutat. 2001;17(1):18–33. doi:10.1002/1098-1004(2001)17:1<18::AID-HUMU3>3.0.CO;2-T.

Strautnieks SS, Kagalwalla AF, Tanner MS, Knisely AS, Bull L, Freimer N, Kocoshis SA, Gardiner RM, Thompson RJ. Identification of a locus for progressive familial intrahepatic cholestasis PFIC2 on chromosome 2q24. Am J Hum Genet. 1997;61(3):630–3. doi:10.1086/515501.

Tanzi RE, Petrukhin K, Chernov I, Pellequer JL, Wasco W, Ross B, Romano DM, Parano E, Pavone L, Brzustowicz LM, et al. The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. Nat Genet. 1993;5(4):344–50. doi:10.1038/ng1293-344.

Thomas GR, Forbes JR, Roberts EA, Walshe JM, Cox DW. The Wilson disease gene: spectrum of mutations and their consequences. Nat Genet. 1995;9(2):210–7. doi:10.1038/ng0295-210.

Valenti L, Fracanzani AL, Rametta R, Fraquelli M, Soverini G, Pelusi S, Dongiovanni P, Conte D, Fargion S. Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis. J Hepatol. 2012a;57(6):1319–25. doi:10.1016/j.jhep.2012.07.041.

Valenti L, Maggioni P, Piperno A, Rametta R, Pelucchi S, Mariani R, Dongiovanni P, Fracanzani AL, Fargion S. Patatin-like phospholipase domain containing-3 gene I148M polymorphism, steatosis, and liver damage in hereditary hemochromatosis. World J Gastroenterol. 2012b;18(22):2813–20. doi:10.3748/wjg.v18.i22.2813.

van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, Bull LN, van den Berg IE, Berger R, Houwen RH, Klomp LW. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology. 2004;127(2):379–84.

Vujic M. Molecular basis of HFE-hemochromatosis. Front Pharmacol. 2014;5:42. doi:10.3389/fphar.2014.00042.

Wan YJ, Poland RE, Lin KM. Genetic polymorphism of CYP2E1, ADH2, and ALDH2 in Mexican-Americans. Genet Test. 1998;2(1):79–83. doi:10.1089/gte.1998.2.79.

Wang FS. Current status and prospects of studies on human genetic alleles associated with hepatitis B virus infection. World J Gastroenterol. 2003;9(4):641–4.

Wang J, Zhou X, Zhao J, Li Z, Li X. Screening for feature genes associated with hereditary hemochromatosis and functional analysis with DNA microarrays. Genet Mol Res. 2013;12(4):6240–8. doi:10.4238/2013.December.4.11.

Ward CJ, Hogan MC, Rossetti S, Walker D, Sneddon T, Wang X, Kubly V, Cunningham JM, Bacallao R, Ishibashi M, Milliner DS, Torres VE, Harris PC. The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. Nat Genet. 2002;30(3):259–69. doi:10.1038/ng833.

Wilfred de Alwis NM, Day CP. Genetics of alcoholic liver disease and nonalcoholic fatty liver disease. Semin Liver Dis. 2007;27(1):44–54. doi:10.1055/s-2006-960170.

Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, Landt CL, Harrison SA. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology. 2011;140(1):124–31. doi:10.1053/j.gastro.2010.09.038.

Yokoyama A, Mizukami T, Matsui T, Yokoyama T, Kimura M, Matsushita S, Higuchi S, Maruyama K. Genetic polymorphisms of alcohol dehydrogenase-1B and aldehyde dehydrogenase-2 and liver cirrhosis, chronic calcific pancreatitis, diabetes mellitus, and hypertension among Japanese alcoholic men. Alcohol Clin Exp Res. 2013;37(8):1391–401. doi:10.1111/acer.12108.

Zamani F, Bagheri Z, Bayat M, Fereshtehnejad SM, Basi A, Najmabadi H, Ajdarkosh H. Iranian hereditary hemochromatosis patients: baseline characteristics, laboratory data and gene mutations. Med Sci Monit. 2012;18(10):CR622–9.

Zeng T, Guo FF, Zhang CL, Song FY, Zhao XL, Xie KQ. Roles of cytochrome P4502E1 gene polymorphisms and the risks of alcoholic liver disease: a meta-analysis. PLoS One. 2013;8(1):e54188. doi:10.1371/journal.pone.0054188.

Zhang B, Xie W, Krasowski MD. PXR: a xenobiotic receptor of diverse function implicated in pharmacogenetics.Pharmacogenomics.2008;9(11):1695–709.doi:10.2217/14622416.9.11.1695.

Zhou J, Zhai Y, Mu Y, Gong H, Uppal H, Toma D, Ren S, Evans RM, Xie W. A novel pregnane X receptor-mediated and sterol regulatory element-binding protein-independent lipogenic pathway. J Biol Chem. 2006;281(21):15013–20. doi:10.1074/jbc.M511116200.

Dr. Nirupma Trehanpati Dr. Nirupma Trehanpati is an Associate Professor at Institute of Liver and Biliary Sciences. Her major research interests are to study the role of T cells and dendritic cells in type 1 diabetes mellitus (T1DM) and therapeutic role of specific CD4/CD8 responses during sequential therapy using tenofovir followed by PEG interferon to monotherapy with tenofovir in chronic hepatitis B patients and to decipher the role of notch signalling pathway in the pathogenesis of virus-related hepatocellular carcinoma.

Implication of Pre-replication Complex Proteins in Human Disease

4

Abid Khan, Arindam Chakraborty, and Supriya G. Prasanth

Proper DNA replication is essential for the maintenance of genomic integrity during cell division. The DNA replication process in eukaryotes is an immaculately concerted and complex process that has evolved several safeguards to ensure the prevention and propagation of erroneous DNA duplication. Hence any mistakes therein, either due to mutations in the components of the replication machinery or change in their expression pattern, could lead to significant genomic abnormalities, which could manifest in a variety of diseases. In this chapter, we will discuss some of the major disease phenotypes associated with DNA replication machinery, particularly focusing on the members of the pre-replication complex (Pre-RC).

DNA replication begins with the loading of the pre-RC onto replication origins during G1 phase of the cell cycle. The process of "replication licensing" ensures that the chromosomes are replicating only once per cell cycle. Licensing entails the sequential loading of ORC (origin recognition complex), the loading factors Cdc6 (cell division cycle 6) and Cdt1 (chromatin licensing and DNA replication factor 1) and the replication associated helicase MCM2-7 (minichromosome maintenance 2–7) (Bell 2002; Bell and Dutta 2002; Dutta and Bell 1997; Kelly and Brown 2000). Briefly, ORC and Cdc6 form a stable complex on DNA, which then recruits Cdt1. Cdt1 loading onto origins is essential for the recruitment of MCMs. Cdt1 levels are regulated by Geminin during the cell cycle. At the end of mitosis, cellular Geminin levels are low, and therefore it binds to Cdt1 at a lower stoichiometric ratio, allowing Cdt1 to be "active" and load onto the origins and associate with ORC and Cdc6. At the G1/S boundary, Geminin levels rise and stay stable through S-G2-M. During these phases, Geminin binds to Cdt1 at a higher stoichiometric ratio, suppresses its activity, and can titrate it off of the pre-RC. Alternatively, Cdt1 is also known to undergo ubiquitination-mediated proteolysis during early S phase (Nishitani et al.

A. Khan • A. Chakraborty • S.G. Prasanth (✉)
Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, 601S Goodwin Avenue, Urbana, IL 61801, USA
e-mail: supriyap@life.illinois.edu

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_4

2004, 2006). These mechanisms preclude MCMs from loading onto origins post licensing and therefore prevent relicensing of the origins and hence rereplication (Blow and Dutta 2005; Lau et al. 2006; Li and Blow 2005). Following the formation of a functional pre-RC, initiation of DNA replication requires the activity of Cdks (cyclin-dependent kinases) and Ddk (Dbf4-dependent kinase, Cdc7) during G1/S to phosphorylate pre-RC which then facilitates the recruitment of MCM10, Cdc45, and GINS leading to origin firing during S phase (Sheu and Stillman 2006; Stillman 1996; Zou and Stillman 1998).

4.1 Mutations of ORC in Meier–Gorlin Syndrome (MGS)

Origin recognition complex, a six subunit complex, was first identified in *Saccharomyces cerevisiae* as a sequence-specific DNA replication initiator complex (Bell and Stillman 1992). It binds to origins in an ATP-dependent manner and recruits other members of pre-RC. Although ORC proteins exhibit high sequence homology with higher eukaryotes, the mechanism of origin recognition and chromatin association is highly variable (Bell and Dutta 2002). Unlike *S. cerevisiae*, in higher organisms, ORCs do not exhibit DNA sequence-specific origin-binding activity; however, they have been shown to bind to AT-rich DNA elements (Vashee et al. 2003). In *Xenopus laevis*, *Homo sapiens*, and *Drosophila melanogaster*, one or more ORC subunits stay bound to chromatin throughout cell cycle. While the levels of Orc2-5 do not change during the cell cycle in human cells, Orc1 gets degraded in post G1 cells (Mendez et al. 2002). In order to understand the mechanism of ORC stability and activity, many new ORC-interacting partners have been identified (Shen and Prasanth 2012), including LRWD1/ORCA (ORC associated) (Bartke et al. 2010; Shen et al. 2010) which stabilizes ORC on chromatin. In addition to its canonical role in replication initiation, ORC components have been shown to exhibit a wide variety of replication-independent functions such as heterochromatin organization (Prasanth et al. 2010), centriole, and centrosome duplication (Huang et al. 1998; Pak et al. 1997; Prasanth et al. 2010), cytokinesis (Prasanth et al. 2002). Although ORC is involved in an extremely important and highly conserved cellular process, yet there aren't many reports implicating ORC subunits in human disease. Here we will discuss the two major diseases associated with ORC deregulation.

Recently, Meier–Gorlin syndrome (MGS), a primordial dwarfism syndrome, has been linked to the mutations in genes encoding the pre-RC components like Orc1, Orc4, Orc6, Cdc6, and Cdt1. This rare autosomal genetic recessive disorder was originally described by Meier et al. (1959) and subsequently by Gorlin et al. (1975). However, the term Meier–Gorlin syndrome was first time coined by Boles et al. (1994). It has also been known as ear, patellae, short stature syndrome (Cohen et al. 1991), characterized by bilateral microtia (small ears), hypoplastic or absence of patellae, short stature, craniofacial anomalies, and growth retardation (Bongers et al. 2001; Gorlin et al. 1975; Loeys et al. 1999). Genital anomalies and mammary hypoplasia have also been reported in some cases (Bongers et al.

2001; de Munnik et al. 2012a). More than 60 cases of MGS have been reported worldwide and 35 individuals out of 45 patients with MGS have been shown to carry mutations in five genes encoding pre-RC component proteins: ORC1, ORC4, ORC6, CDT1, and CDC6 (Bicknell et al. 2011a, b; de Munnik et al. 2012a, b; Guernsey et al. 2011).

Molecular genetic analysis in two sibs with microcephalic primordial dwarfism resembling Meier–Gorlin syndrome, from a Saudi–Arabian family, identified missense mutation in Orc1 gene (Bicknell et al. 2011b). The mutation was caused by a homozygous A to G transition (c.314A>G) resulting into a non-conservative amino acid substitution (E127G) in exon 4 of Orc1 gene (Bicknell et al. 2011b). Subsequent analysis in 204 additional individuals with microcephalic primordial dwarfism further identified biallelic missense mutations, including a recurrent mutation, R105Q, that changes a conserved amino acid at the N-terminal BAH domain of ORC1 (Bicknell et al. 2011b). Studies in cell lines established from MGS patients carrying the mutations E127G and R105Q revealed reduced chromatin binding of ORC1 and impaired pre-RC assembly. Further, cells carrying ORC1 mutation showed impaired licensing and replication origin activation and perturbed S phase entry and progression (Bicknell et al. 2011b). Interestingly, targeting of Orc1 in zebrafish morphant model significantly reduced the embryo size and caused dwarfism (Bicknell et al. 2011b). Similar growth defects and dwarf-like phenotype was also observed upon mcm5 depletion in zebrafish, suggesting the role of impaired origin licensing in manifesting the MGS-like growth retardation (Bicknell et al. 2011b). Recent study by Kuo et al. (2012) has further linked the ORC1 BAH domain mutation to Meier–Gorlin syndrome (MGS). The histone H4 dimethylated at lysine 20 (H4K20me2) is enriched at the site of replication in diverse metazoans. Orc1 binds to this methylated histone through its BAH domain and regulates ORC-chromatin binding (Kuo et al. 2012). Disruption of $ORC1_{BAH}$ to H4K20me2 recognition impairs ORC-chromatin association at replication origins (Kuo et al. 2012). Interestingly, the wild-type human Orc1 mRNA but not the H4K20me2-binding pocket mutants (hORC1-Y64A and hORC1-W88A) mRNA could rescue the dwarf phenotype of Orc1 depleted zebrafish, when co-injected with orc1-targeting morpholino oligonucleotides (Kuo et al. 2012). Additionally, zebrafish depleted of H4K20me2 or mice lacking H4K20me2 exhibited similar growth defect phenotype of orc1 morphants (Kuo et al. 2012; Schotta et al. 2008), indicating a possible connection of BAH domain mutation and MGS pathogenesis. Defects in primary cilia formation, cilia function, and chondro-induction have also been attributed to Orc1 MGS mutations (Stiff et al. 2013). Alteration of centrosome duplication associated with Orc1 mutations that disrupt Cyclin E-CDK2 kinase inhibition has also been implicated in some patients with Meier–Gorlin syndrome (Hossain and Stillman 2012). Subsequent sequencing study by Bicknell et al. (2011b) in patients with MGS, microcephaly, and profound growth retardation further identified compound heterozygosity for a splice acceptor site mutation (intron 9 splice acceptor site) and a frameshift mutation (p.Val667fsX24), respectively, in conjugation with the previously identified recurrent mutation R105Q (Bicknell et al. 2011a).

Sequencing of Orc4 gene from the individuals with MGS lead to the identification of a homozygous missense mutation, Y174C affecting the consensus AAA+ (ATPase associated with a wide range of cellular activities) domain of ORC4 (Bicknell et al. 2011a). Similar mutation was also independently identified by Guernsey et al. (2011) from additional patients. Compound heterozygosity for Y174 and another frameshift mutation causing premature protein truncation have also been reported in some cases (Bicknell et al. 2011a; Guernsey et al. 2011). The reported Tyr^{174} is highly conserved across the taxa, maps between the Walker B motif, and the sensor domain I and has been shown to be crucial for interacting with a conserved arginine residue on an adjacent helix structure (Bell 2002; Chuang and Kelly 1999; Guernsey et al. 2011; Iyer et al. 2004). Functional analysis in yeast *S. cerevisiae*, carrying the missense mutation Y174C ($orc4^{Y232C}$) revealed to be pathogenic and resulted in slower growth rate due to the defect in G1 to S phase transition (Guernsey et al. 2011; Ladha 2011).

Compound heterozygosity for a loss-of-function mutation caused by a 2 bp deletion and a missense mutation in the Orc6 gene have been reported from the three siblings of a Turkish family (Bicknell et al. 2011a). Though the Orc6 molecule is poorly conserved among metazoans, the amino acid tyrosine, which is substituted by serine (Y232S) as a result of the missense mutation, is highly conserved from yeast to human (Bicknell et al. 2011a; Bleichert et al. 2013). In *Drosophila*, biochemical studies have shown that the conserved amino acids including the tyrosine mutated in MGS patients at the C-terminal end of the Orc6 is essential for its assembly with ORC via its interaction with Orc3 (Bleichert et al. 2013). Functional study with corresponding human Orc6 MGS mutation in *Drosophila* (Y225S) reduced its recruitment in ORC and weakened its association with Orc3 (Bleichert et al. 2013). In human also, pulldown assay using Orc6 mutant carrying MGS mutation also affected its binding to ORC1-5 (Bleichert et al. 2013). Similar observation was also made in *S. cerevisiae*, suggesting that MGS mutation in Orc6 interferes with ORC function by reducing its association with Orc3 (Bleichert et al. 2013).

Mutations have also been reported in well-conserved residues of Cdt1 at the C-terminal domain of the protein, a potential site for MCM interaction during origin recognition (Bicknell et al. 2011a; Guernsey et al. 2011; Jee et al. 2010). Another homozygous missense mutation that substitutes a conserved threonine residue (T323R) in the Cdc6 gene has also been identified in a patient with MGS (Bicknell et al. 2011a). This mutation also lies within AAA domain, an ATP-binding domain known for its crucial function during DNA replication (Bicknell et al. 2011a; Herbig et al. 1999).

Molecular genetic analysis of patients with MGS clustered all the mutations in five pre-RC components. Biochemical characterizations further indicate the impaired functions of DNA replication initiation associated with these mutations. Therefore, it has been suggested that the defects in replication licensing leads to the clinical manifestation of growth failure in the individuals with Meier–Gorlin syndrome. In addition, recent work has demonstrated that Orc1 mutations in MGS patients are due to disruption of an Orc1 CDK inhibitory domain and centrosomal reduplication (Hossain and Stillman 2012). This is further supported by the

observations that a reduced efficiency in the formation of cilia due to centrosome defects contribute to the clinical features associated with MGS (Stiff et al. 2013). These results suggest that the multi-talented ORC proteins have diverse functions, and intact ORC is critical to maintaining genomic integrity.

4.2 Dysregulation of Pre-replication Complex Proteins in Cancer

Any perturbation in the process of origin licensing can have deleterious effects on the propagation of faithful replication and consequently could lead to drastic genomic instabilities, which are a hallmark of many cancers. Therefore, it is not surprising that many of the pre-RC components are upregulated in several cancers.

Functional depletion studies of pre-RC components such as ORC, Cdc7, and Cdt1 in mammalian cells lead to cell cycle arrest and/or cell death (Feng et al. 2003; Kim et al. 2002). Specifically, depletion of ORC subunits causes aberrant DNA replication, S phase progression, mitotic defects, and cell death (Prasanth et al. 2002, 2004). Other reports have demonstrated that pre-RC components are essential for S phase checkpoint signaling and genomic stability (Clay-Farrace et al. 2003) as well. These reports clearly demonstrate that normal cellular levels of pre-RC components are vital for faithful DNA replication and normal cell cycle progression. How then could insufficiency in pre-RC components manifest at a physiological level? One hypothesis is that insufficiency in pre-RC components could contribute to tumorigenesis by inducing genomic instability.

Licensing proteins have been found to be significantly upregulated in neoplastic cells but not in quiescent cells, making them highly useful diagnostic tools for detecting tumors. Cdc6 levels have been reported to correlate with the presence of neoplastic cells (Semple and Duncker 2004). Increased expression of Cdc6 and Cdt1 has been observed in cervical, lung, and brain cancers (Karakaidos et al. 2004). Furthermore, introduction of Cdt1 expressing NIH3T3 cells into nude mice can drive tumorigenesis (Arentson et al. 2002). Mice that overexpress Cdt1 have been shown to develop lymphoblastic lymphomas in a p53 null background (Seo et al. 2005).

Mcm proteins have been particularly effective as diagnostic and prognostic markers for many different tumors (Gonzalez et al. 2005). It is well established that MCM expression is lost during differentiation, and it is not present in the differentiated epithelial cells of the cervix, urinary tract, etc. (Freeman et al. 1999; Gonzalez et al. 2005). Interestingly, it is found to be upregulated in malignant tumors of these sites (Hiraiwa et al. 1997; Todorov et al. 1998; Tachibana et al. 2005). MCM7 levels were found to be elevated in keratinocytic tumors relative to the normal cells (Hiraiwa et al. 1997). MCM5 levels were shown to be upregulated in cervical and esophageal cancers (Williams et al. 1998). MCM8 disruption and an alternative splice form have been observed in hepatic carcinoma (Gozuacik et al. 2003) and choriocarcinoma (Johnson et al. 2003), respectively. Elegant work in mice carrying a hypomorphic MCM4 allele, *Chaos3*, exhibits higher incidence rates of mammary adenocarcinoma and increased

sensitivity to DNA breaks under replication stress (Shima et al. 2007). This report and several others suggest that partial loss-of-individual pre-RC components could lead to genomic instability and contribute to tumorigenesis. MCM family of proteins has the potential to be a diagnostic as well as a prognostic tumor marker for a variety of tumors including colorectal tumors, lung cancer, and oral and anal cancer (Giaginis et al. 2010, 2011; Hua et al. 2014). MCMs are routinely used in the clinics, including for early detection of cancer and because of its abundance and ease of detection of this nuclear protein, a high-throughput screening approach is possible.

Although these data suggest involvement of pre-RC components in tumorigenesis, they do not however illustrate pre-RC's importance in the induction of oncogenesis. As with loss of function, gain of pre-RC function can also have serious replication defects that promote oncogenesis. Overexpression of Cdc6, Cdt1, or MCM7 in mammalian cells results in abnormal DNA replication and S-G2-M checkpoint activation (Honeycutt et al. 2006; Karakaidos et al. 2004). An overabundance in Cdt1 results in rereplication and chromosome fragmentation in *Xenopus* (Davidson et al. 2006). Abnormal replication leading to genomic instability is hypothesized to promote tumorigenesis. Consistent with this notion, groups have shown that Cdt1 transgenic mice under a p53 null background develop lymphoblastic lymphoma and that Cdt1 overexpression in NIH3T3 cells can promote transformation (Seo et al. 2005). Although these reports clearly demonstrated the oncogenic potential of pre-RC, there are very few reports delineating the mechanism by which pre-RC components promote oncogenesis. One such study demonstrated a detailed mechanism of action for Cdc6 in oncogenesis via transcriptional repression of the tumor suppressor INK4/ARF locus. Gonzalez et al. showed that overexpression of Cdc6 causes hypermethylation of the locus thereby suppressing its transcriptional output. Cdc6 binds to a replication origin in the locus, which is also a transcriptional control element and recruits histone deacetylase complex resulting in hypermethylation and repression of p14ARF, p16INK4A, and p14INK4B expression (Gonzalez et al. 2006). The locus encodes three critical cell cycle inhibitors; therefore, repression of these products alleviates cell cycle control, and together with abundance in Cdt1, there is enhanced cellular proliferation, which contributes to tumorigenesis. Consistent with this they also show that overexpression of Cdt1 can cooperate with Ras to transform mouse embryonic fibroblasts (MEFs). Taken together these data suggest that proper pre-RC function impacts checkpoint signaling, genome stability, and tumorigenesis.

The role of pre-RC proteins in tumorigenesis came from understanding the biology of tumor viruses Epstein–Barr virus (EBV) or human herpes virus (HHV). EBV achieves replication of its genome by hijacking the host cellular replication machinery. EBV was first identified in a cell line derived from Burkitt's lymphoma patient by Epstein et al. in Barr group, and hence the name (Epstein et al. 1964). One of the most common human viruses is thought to infect more than 90% of human population. EBV is transmitted through saliva or genital secretion. EBV infection does not usually illicit strong symptoms in most cases. It can remain latent indefinitely inside the B-cells. Under immunodeficient conditions of the host, EBV gets activated and switch to its lytic replication mode, producing millions of virions. Reactivation of EBV can then result in a range of cancers including epithelial malignancies, lymphomas, and lymphoproliferative disorders (Maeda et al. 2009; Tao et al. 2006).

EBNA1 (Epstein–Barr nuclear antigen 1), a protein encoded by the EBV genome, binds to its origin of viral replication, oriP, and then recruits host cell pre-RC components to initiate replication (Dhar et al. 2001). An Orc2 hypomorphic allele abrogates EBNA1-mediated viral DNA replication but this can be rescued by overexpression of wt Orc2, suggesting that host Orc2 is critical for the viral genome replication. EBNA1 associates with Orc2 and ChIP experiments show that Orc2 binds to the oriP sequence (Chaudhuri et al. 2001; Dhar et al. 2001). Orc1 and MCM bind to the oriP sequence in a cell cycle-dependent manner, while other ORC subunits remain associated with oriP throughout cell cycle (Chaudhuri et al. 2001; Ritzi et al. 2003). Overexpression of Geminin inhibits EBNA1-ORC-mediated replication of EBV genome from oriP, and this can be rescued by overexpressing Cdt1 (Dhar et al. 2001). Taken together these data suggests that EBV hijacks not only ORC but the entire pre-RC complex to facilitate replication of its genome.

Conclusion

DNA replication is essential for normal development, and the central player in this field is the pre-RC that ensures faithful propagation of the genome through successive cellular divisions. ORCs play a pivotal role in the pre-RC. Detailed mapping and characterization of genetic mutations with disease phenotypes such as in the case of MGS have provided valuable information with regard to its role in normal development. On the other hand, EBV provides a simple and valuable tool to study origin licensing by pre-RC and for therapeutic intervention.

Although there is a growing body of literature implicating pre-RC in cancer, there still remain a number of unanswered questions with regard to the precise role of pre-RC in oncogenesis. For example, how is the expression of pre-RC components deregulated during cancer? Are mutations in pre-RC components alone sufficient to drive oncogenesis or do they require secondary mutations in key tumor suppressors? How do cancer cells evade replication checkpoint activation? Some of these questions warrant extensive investigation that will help elucidate the oncogenic potential of pre-RC proteins and improve our understanding of the importance of proper DNA replication in carcinogenesis.

Acknowledgments We thank members of the Prasanth laboratory for discussions and suggestions. Work in the Prasanth lab is supported by NSF career (1243372) and NIH (1RO1GM099669) awards to SGP.

References

Arentson E, Faloon P, Seo J, Moon E, Studts JM, Fremont DH, Choi K. Oncogenic potential of the DNA replication licensing protein CDT1. Oncogene. 2002;21:1150–8.

Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010;143:470–84.

Bell SP. The origin recognition complex: from simple origins to complex functions. Genes Dev. 2002;16:659–72.

Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem. 2002;71:333–74.

Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature. 1992;357:128–34.

Bicknell LS, Bongers EM, Leitch A, Brown S, Schoots J, Harley ME, Aftimos S, Al-Aama JY, Bober M, Brown PA, van Bokhoven H, Dean J, Edrees AY, Feingold M, Fryer A, Hoefsloot LH, Kau N, Knoers NV, Mackenzie J, Opitz JM, Sarda P, Ross A, Temple IK, Toutain A, Wise CA, Wright M, Jackson AP. Mutations in the pre-replication complex cause Meier–Gorlin syndrome. Nat Genet. 2011a;43:356–9.

Bicknell LS, Walker S, Klingseisen A, Stiff T, Leitch A, Kerzendorfer C, Martin CA, Yeyati P, Al Sanna N, Bober M, Johnson D, Wise C, Jackson AP, O'Driscoll M, Jeggo PA. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier–Gorlin syndrome. Nat Genet. 2011b;43:350–5.

Bleichert F, Balasov M, Chesnokov I, Nogales E, Botchan MR, Berger JM. A Meier–Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. elife. 2013;2:e00882.

Blow JJ, Dutta A. Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol. 2005;6:476–86.

Boles RG, Teebi AS, Schwartz D, Harper JF. Further delineation of the ear, patella, short stature syndrome (Meier–Gorlin syndrome). Clin Dysmorphol. 1994;3:207–14.

Bongers EM, Opitz JM, Fryer A, Sarda P, Hennekam RC, Hall BD, Superneau DW, Harbison M, Poss A, van Bokhoven H, Hamel BC, Knoers NV. Meier–Gorlin syndrome: report of eight additional cases and review. Am J Med Genet. 2001;102:115–24.

Chaudhuri B, Xu H, Todorov I, Dutta A, Yates JL. Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein–Barr virus. Proc Natl Acad Sci USA. 2001;98:10085–9.

Chuang RY, Kelly TJ. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proc Natl Acad Sci USA. 1999;96:2656–61.

Clay-Farrace L, Pelizon C, Santamaria D, Pines J, Laskey RA. Human replication protein Cdc6 prevents mitosis through a checkpoint mechanism that implicates Chk1. EMBO J. 2003;22:704–12.

Cohen B, Temple IK, Symons JC, Hall CM, Shaw DG, Bhamra M, Jackson AM, Pembrey ME. Microtia and short stature: a new syndrome. J Med Genet. 1991;28:786–90.

Davidson IF, Li A, Blow JJ. Deregulated replication licensing causes DNA fragmentation consistent with head-to-tail fork collision. Mol Cell. 2006;24:433–43.

de Munnik SA, Bicknell LS, Aftimos S, Al-Aama JY, van Bever Y, Bober MB, Clayton-Smith J, Edrees AY, Feingold M, Fryer A, van Hagen JM, Hennekam RC, Jansweijer MC, Johnson D, Kant SG, Opitz JM, Ramadevi AR, Reardon W, Ross A, Sarda P, Schrander-Stumpel CT, Schoots J, Temple IK, Terhal PA, Toutain A, Wise CA, Wright M, Skidmore DL, Samuels ME, Hoefsloot LH, Knoers NV, Brunner HG, Jackson AP, Bongers EM. Meier–Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. Eur J Hum Genet. 2012a;20:598–606.

de Munnik SA, Otten BJ, Schoots J, Bicknell LS, Aftimos S, Al-Aama JY, van Bever Y, Bober MB, Borm GF, Clayton-Smith J, Deal CL, Edrees AY, Feingold M, Fryer A, van Hagen JM, Hennekam RC, Jansweijer MC, Johnson D, Kant SG, Opitz JM, Ramadevi AR, Reardon W, Ross A, Sarda P, Schrander-Stumpel CT, Sluiter AE, Temple IK, Terhal PA, Toutain A, Wise CA, Wright M, Skidmore DL, Samuels ME, Hoefsloot LH, Knoers NV, Brunner HG, Jackson AP, Bongers EM. Meier–Gorlin syndrome: growth and secondary sexual development of a microcephalic primordial dwarfism disorder. Am J Med Genet A. 2012b;158A:2733–42.

Dhar SK, Yoshida K, Machida Y, Khaira P, Chaudhuri B, Wohlschlegel JA, Leffak M, Yates J, Dutta A. Replication from oriP of Epstein–Barr virus requires human ORC and is inhibited by geminin. Cell. 2001;106:287–96.

Dutta A, Bell SP. Initiation of DNA replication in eukaryotic cells. Annu Rev Cell Dev Biol. 1997;13:293–332.

Epstein MA, Achong BG, Barr YM. Virus particles in cultured lymphoblasts from Burkitt's lymphoma. Lancet. 1964;1:702–3.

Feng D, Tu Z, Wu W, Liang C. Inhibiting the expression of DNA replication-initiation proteins induces apoptosis in human cancer cells. Cancer Res. 2003;63:7356–64.

Freeman A, Morris LS, Mills AD, Stoeber K, Laskey RA, Williams GH, Coleman N. Minichromosome maintenance proteins as biological markers of dysplasia and malignancy. Clin Cancer Res. 1999;5:2121–32.

Giaginis C, Vgenopoulou S, Vielh P, Theocharis S. MCM proteins as diagnostic and prognostic tumor markers in the clinical setting. Histol Histopathol. 2010;25:351–70.

Giaginis C, Giagini A, Tsourouflis G, Gatzidou E, Agapitos E, Kouraklis G, Theocharis S. MCM-2 and MCM-5 expression in gastric adenocarcinoma: clinical significance and comparison with Ki-67 proliferative marker. Dig Dis Sci. 2011;56:777–85.

Gonzalez MA, Tachibana KE, Laskey RA, Coleman N. Control of DNA replication and its potential clinical exploitation. Nat Rev Cancer. 2005;5:135–41.

Gonzalez S, Klatt P, Delgado S, Conde E, Lopez-Rios F, Sanchez-Cespedes M, Mendez J, Antequera F, Serrano M. Oncogenic activity of Cdc6 through repression of the INK4/ARF locus. Nature. 2006;440:702–6.

Gorlin RJ, Cervenka J, Moller K, Horrobin M, Witkop Jr CJ. Malformation syndromes. A selected miscellany. Birth Defects Orig Artic Ser. 1975;11:39–50.

Gozuacik D, Chami M, Lagorce D, Faivre J, Murakami Y, Poch O, Biermann E, Knippers R, Brechot C, Paterlini-Brechot P. Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. Nucleic Acids Res. 2003;31:570–9.

Guernsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, Perry S, Ferguson M, LeBlanc M, Paquette J, Patry L, Rideout AL, Thomas A, Orr A, McMaster CR, Michaud JL, Deal C, Langlois S, Superneau DW, Parkash S, Ludman M, Skidmore DL, Samuels ME. Mutations in origin recognition complex gene ORC4 cause Meier–Gorlin syndrome. Nat Genet. 2011;43:360–4.

Herbig U, Marlar CA, Fanning E. The Cdc6 nucleotide-binding site regulates its activity in DNA replication in human cells. Mol Biol Cell. 1999;10:2631–45.

Hiraiwa A, Fujita M, Nagasaka T, Adachi A, Ohashi M, Ishibashi M. Immunolocalization of hCDC47 protein in normal and neoplastic human tissues and its relation to growth. Int J Cancer. 1997;74:180–4.

Honeycutt KA, Chen Z, Koster MI, Miers M, Nuchtern J, Hicks J, Roop DR, Shohet JM. Deregulated minichromosomal maintenance protein MCM7 contributes to oncogene driven tumorigenesis. Oncogene. 2006;25:4027–32.

Hossain M, Stillman B. Meier–Gorlin syndrome mutations disrupt an Orc1 CDK inhibitory domain and cause centrosome reduplication. Genes Dev. 2012;26:1797–810.

Hua C, Zhao G, Li Y, Bie L. Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas. BMC Cancer. 2014;14:526.

Huang DW, Fanti L, Pak DT, Botchan MR, Pimpinelli S, Kellum R. Distinct cytoplasmic and nuclear fractions of Drosophila heterochromatin protein 1: their phosphorylation levels and associations with origin recognition complex proteins. J Cell Biol. 1998;142:307–18.

Iyer LM, Leipe DD, Koonin EV, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol. 2004;146:11–31.

Jee J, Mizuno T, Kamada K, Tochio H, Chiba Y, Yanagi K, Yasuda G, Hiroaki H, Hanaoka F, Shirakawa M. Structure and mutagenesis studies of the C-terminal region of licensing factor Cdt1 enable the identification of key residues for binding to replicative helicase Mcm proteins. J Biol Chem. 2010;285:15931–40.

Johnson EM, Kinoshita Y, Daniel DC. A new member of the MCM protein family encoded by the human MCM8 gene, located contrapodal to GCD10 at chromosome band 20p12.3-13. Nucleic Acids Res. 2003;31:2915–25.

Karakaidos P, Taraviras S, Vassiliou LV, Zacharatos P, Kastrinakis NG, Kougiou D, Kouloukoussa M, Nishitani H, Papavassiliou AG, Lygerou Z, Gorgoulis VG. Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant p53 on tumor growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. Am J Pathol. 2004;165:1351–65.

Kelly TJ, Brown GW. Regulation of chromosome replication. Annu Rev Biochem. 2000;69:829–80.

Kim JM, Nakao K, Nakamura K, Saito I, Katsuki M, Arai K, Masai H. Inactivation of Cdc7 kinase in mouse ES cells results in S-phase arrest and p53-dependent cell death. EMBO J. 2002;21:2168–79.

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. Nature. 2012;484:115–9.

Ladha S. Of ORC and forks: the identification of mutations implicated in Meier–Gorlin syndrome. Clin Genet. 2011;80:506–7.

Lau E, Zhu C, Abraham RT, Jiang W. The functional role of Cdc6 in S-G2/M in mammalian cells. EMBO Rep. 2006;7:425–30.

Li A, Blow JJ. Cdt1 downregulation by proteolysis and geminin inhibition prevents DNA re-replication in Xenopus. EMBO J. 2005;24:395–404.

Loeys BL, Lemmerling MM, Van Mol CE, Leroy JG. The Meier–Gorlin syndrome, or ear-patella-short stature syndrome, in sibs. Am J Med Genet. 1999;84:61–7.

Maeda E, Akahane M, Kiryu S, Kato N, Yoshikawa T, Hayashi N, Aoki S, Minami M, Uozaki H, Fukayama M, Ohtomo K. Spectrum of Epstein–Barr virus-related diseases: a pictorial review. Jpn J Radiol. 2009;27:4–19.

Meier Z, Poschiavo N, Rothschild M. Case of arthrogryposis multiplex congenita with mandibulofacial dysostosis (Franceschetti syndrome). Helv Paediatr Acta. 1959;14:213–6.

Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B. Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol Cell. 2002;9:481–91.

Nishitani H, Lygerou Z, Nishimoto T. Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of geminin through its N-terminal region. J Biol Chem. 2004;279:30807–16.

Nishitani H, Sugimoto N, Roukos V, Nakanishi Y, Saijo M, Obuse C, Tsurimoto T, Nakayama KI, Nakayama K, Fujita M, Lygerou Z, Nishimoto T. Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. EMBO J. 2006;25:1126–36.

Pak DT, Pflumm M, Chesnokov I, Huang DW, Kellum R, Marr J, Romanowski P, Botchan MR. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell. 1997;91:311–23.

Prasanth SG, Prasanth KV, Stillman B. Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. Science. 2002;297:1026–31.

Prasanth SG, Prasanth KV, Siddiqui K, Spector DL, Stillman B. Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. EMBO J. 2004;23:2651–63.

Prasanth SG, Shen Z, Prasanth KV, Stillman B. Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proc Natl Acad Sci USA. 2010;107:15093–8.

Ritzi M, Tillack K, Gerhardt J, Ott E, Humme S, Kremmer E, Hammerschmidt W, Schepers A. Complex protein-DNA dynamics at the latent origin of DNA replication of Epstein–Barr virus. J Cell Sci. 2003;116:3971–84.

Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callen E, Celeste A, Pagani M, Opravil S, De La Rosa-Velazquez IA, Espejo A, Bedford MT, Nussenzweig A, Busslinger M, Jenuwein T. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev. 2008;22:2048–61.

Semple JW, Duncker BP. ORC-associated replication factors as biomarkers for cancer. Biotechnol Adv. 2004;22:621–31.

Seo J, Chung YS, Sharma GG, Moon E, Burack WR, Pandita TK, Choi K. Cdt1 transgenic mice develop lymphoblastic lymphoma in the absence of p53. Oncogene. 2005;24:8176–86.

Shen Z, Prasanth SG. Emerging players in the initiation of eukaryotic DNA replication. Cell Div. 2012;7:22.

Shen Z, Sathyan KM, Geng Y, Zheng R, Chakraborty A, Freeman B, Wang F, Prasanth KV, Prasanth SG. A WD-repeat protein stabilizes ORC binding to chromatin. Mol Cell. 2010;40:99–111.

Sheu YJ, Stillman B. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol Cell. 2006;24:101–13.

Shima N, Alcaraz A, Liachko I, Buske TR, Andrews CA, Munroe RJ, Hartford SA, Tye BK, Schimenti JC. A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. Nat Genet. 2007;39:93–8.

Stiff T, Alagoz M, Alcantara D, Outwin E, Brunner HG, Bongers EM, O'Driscoll M, Jeggo PA. Deficiency in origin licensing proteins impairs cilia formation: implications for the aetiology of Meier–Gorlin syndrome. PLoS Genet. 2013;9:e1003360.

Stillman B. Cell cycle control of DNA replication. Science. 1996;274:1659–64.

Tachibana K.-e.K, Gonzalez MA, and Coleman N. Cell-cycle-dependent regulation of DNA replication and its relevance to cancer pathology. J. Pathol. 2005;205:123–29.

Tao Q, Young LS, Woodman CB, Murray PG. Epstein–Barr virus (EBV) and its associated human cancers—genetics, epigenetics, pathobiology and novel therapeutics. Front Biosci. 2006;11:2672–713.

Todorov IT, Werness BA, Wang HQ, et al: HsMCM2/BM28: A novel proliferation marker for human tumors and normal tissues. Lab Invest 78:73,1998–78.

Vashee S, Cvetic C, Lu W, Simancek P, Kelly TJ, Walter JC. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev. 2003;17:1894–908.

Williams GH, Romanowski P, Morris L, Madine M, Mills AD, Stoeber K, Marr J, Laskey RA, Coleman N. Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. Proc Natl Acad Sci USA. 1998;95:14932–7.

Zou L, Stillman B. Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin. Science. 1998;280:593–6.

Dr. Supriya G. Prasanth Dr. Supriya G. Prasanth is a professor at the Department of Cell and Developmental Biology, University of Illinois, USA. The long-term interest of her lab is to understand how the initiation of DNA replication is coordinated with chromatin organization and cell cycle progression in mammalian cells. Her lab focuses on studying how origin recognition complex (ORC), and ORC-associated factors (ORCA/LRWD1) regulate DNA replication and heterochromatin organization.

Non-muscle Myosin II Motor Proteins in Human Health and Diseases

5

Venkaiah Betapudi

5.1 Introduction

Man-made machines are involved in mediating a diverse range of human activities in modern world, so are natural myosin motor proteins in driving multiple aspects of cellular life. Myosins belong to a special group of proteins called mechanochemical enzymes or colloquially molecular machines/motor proteins because of their ability to move on intracellular tracks and convert cellular free energy released from ATP into mechanical work (Betapudi 2014; Bustamante et al. 2004). Human body is like a complex machine that is made up of approximately 37.2 trillion cells (Bianconi et al. 2013), and each individual cell is equipped with a variety of molecular machines to perform specific mechanical function. Based on their involvement in mediating different mechanical functions in the cellular world, they are classified into polymerization (actin, microtubule, dynamin), cytoskeletal (myosin, kinesin, dynein), rotary (F0F1-ATP synthase), and nucleic acid (DNA and RNA polymerases, topoisomerases, helicase, remodels the structure of chromatin, SW1/SNF complex, structural maintenance of chromosomes, viral DNA packaging protein) motor proteins (Howard 2014; Kolomeisky 2013). In addition to these specialized motor proteins, cells express another unique type of motor protein called "prestin" that is essential for auditory processing; however, its expression is limited to mammalian cochlear outer hair cells in order to produce mechanical amplification in the auditory portion of the ear cochlea. Unlike the classical ATP-dependent motor proteins, this special membrane motor protein with piezoelectric properties directly converts voltage into mechanical work within microseconds and thus received its

V. Betapudi
Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH 44106, USA
e-mail: betapuv@ccf.org; vxb19@case.edu

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_5

name prestin to represent a musical notation "presto" which means extremely fast in Italian (Zheng et al. 2000). Another special type of motor protein myosin III with ATPase and kinase activity has been reported in the retina of several organisms. This unique motor protein with kinase activity is believed to play critical roles in mediating the visual phototransduction in rod, cone, and photosensitive ganglion cells of the retina of many organisms. Thus, cell-type-specific expression of these molecular machines with dedicated functions is probably a part of nature's strategy for the eukaryotic cell origin and diversification.

Given their ability to operate in a cellular world where Brownian motion and viscous forces dominate inertia, these biological machines are also called Brownian motors. Interestingly, these cellular motor proteins transduce the ATP-released cellular free energy into mechanical work more efficiently than man-made combustion engines (Kabir et al. 2011; van den Heuvel et al. 2007). Many modern cell biologists believe that the mechanical work performed by these cellular motor proteins intersects with every facet of cell biology. Indeed, the mechanical work performed by these motor proteins drive several cellular activities that are essential for mediating reproduction, childbirth, growth, development, immunity, and singing a courtship song in fruit flies as well as predisposing human beings to a certain degree of risk for various pathological conditions and diseases (Chakravorty et al. 2014; Maravillas-Montero and Santos-Argumedo 2012; Min et al. 2014; Pecci et al. 2014; Slonska et al. 2012; Stedman et al. 2004). It is largely believed that no biological cell can function and operate without the involvement of these multifunctional molecular machines. The present chapter is about the discovery, current understanding, and recent advances in various aspects of myosin motor proteins as well as their regulation and relevance to human health and diseases.

5.2 Discovery of Myosin Motor Proteins

More than 150 years ago in Heidelberg, a soluble protein that is responsible for muscle rigidity or contraction was identified by Willy Kuhne in the extracts of smooth and striated muscles and then named "myosin" (myo- + -ose + -in) or colloquially called "Kuhne's myosin" (Kühne 1864). Myosin means "within muscle," and the term "myo" was originated from a Greek word "mys" to describe muscle. Several decades later, Kuhne's myosin in solution was demonstrated to transmit light occasionally with different velocities (Muralt and Edsall 1930). These birefringence properties of asymmetric liquid crystals gave first clue about an unstable uniform shape and size of Kuhne's myosin particle. Within four years of the discovery of myosin birefringence properties, Lohmann demonstrated that the chemical energy released from ATP is required for muscle contraction (Lohmann 1934). This led to identify molecules with ATPase activities that are present in the extracts of smooth muscles. In less than a decade, the ATPase activity for Kuhne's myosin was reported by two independent laboratories (Engelhardt and Liubimova 1994). Later, protein purification and crystallization studies uncovered presence of yet another motor protein called actin in Kuhne's myosin particle (Szent-Gyorgyi 1943;

Straub 1942). Nearly after three decades of the revelation of ATPase activity for Kuhne's myosin, identification of another protein with ATPase activity in *Acanthamoeba castellanii* (Pollard and Korn 1973) has led to the discovery of a large number of divergent motor proteins in many organisms. Recent technological advancements in genome sequencing, molecular biological techniques, and bioinformatic approaches have made possible to identify more than 145 myosins in eukaryotes except in red algae and diplomonad protists to date (Vale 2003). This large number of myosin motor proteins with significant sequence similarities is placed under a single family called "myosin superfamily." Recent genome analysis studies uncovered 20, 687 protein-coding genes in the human genome (Pennisi 2012), and 40 of them are myosin motor proteins-coding genes. However, lower eukaryotes such as *Drosophila melanogaster* (fruit fly), *Oryza sativa* (rice), *Arabidopsis thaliana* (mouse-ear cress, a small flowering plant), *Saccharomyces cerevisiae* (yeast), and *Dictyostelium discoideum* (social amoeba) have 13, 14, 17, 5, and 13 myosin coding genes, respectively.

Recent survey of eukaryotic genomes and phylogenetic analyses of the myosin gene family reveal continuous evolution of new myosins with a significant expansion of their abundance throughout eukaryotic evolution predating to the origin of animal multicellularity (Sebe-Pedros et al. 2014). This may suggest complex and a wide variety of cellular roles for these molecular machines in higher organisms.

5.3 Cellular Processes Mediated by Myosin Motor Proteins

Similar to other groups of molecular machines that are expressed in different cell and tissue types, myosins are abundantly and ubiquitously expressed motor proteins in the human body; nonetheless, some of them display cell- and tissue-specific expressions perhaps due to their specialization in mediating certain cellular functions. Myosins are an essential component of the cell cytoskeleton, a complex network of small and microfilaments distributed throughout the cytoplasm. The cytoskeleton is made up of proteins encoded by approximately 441 genes in human. Myosins associate with actin filaments to form "actomyosin system," an important part of the cytoskeleton in the cells. The actomyosin system organization, adaptability, and dynamic state are critical for maintaining and changing cell morphology and polarity during a variety of cellular processes such as cell division, migration, endocytosis, intracellular trafficking, microparticles release, and apoptotic cell death (Betapudi et al. 2013; Lin et al. 2012). Thus, these molecular machines drive a large number of cellular processes that are essential for growth, development, maintaining normal physiology, and mediating the death of an organism. Myosins also play an important role in creating water-filled body cavities called "hydroskeleton" or "hydrostatic skeleton" that is often found in many ectothermic organisms and soft-bodied invertebrates such as starfish, sea urchins, and earthworms (Serwe et al. 1993). The actomyosin-mediated contraction of the surrounding circular, longitudinal, and helical muscles as well as fluid pressure helps these organisms change their shape and produce movement during burrowing and/or swimming (Kier 2012).

Interestingly, the non-helical hydroskeleton forms a functional basis for the mechanical behavior of mammalian penis (Kelly 2002). Unlike mammals, plants express only a few types of myosin motor proteins. Though the specific functions of these motor proteins are yet to be uncovered, some of them have been implicated in the formation and operation of plasmodesmata that are involved in intercellular transportation and communication in plants (Baluska et al. 2001; Wang and Pesacreta 2004). Myosins play mechanical roles in photosynthesis indirectly by mediating intracellular chloroplast distribution in response to external light conditions (Paves and Truve 2007). In addition, myosins involve in mediating intracellular trafficking of mitochondria and endoplasmic reticulum in mesophyll cells (Liebe and Menzel 1995).

5.4 Common Structural Features and Classification of Myosin Motor Proteins

Although each myosin motor protein plays a specific role in driving cellular processes in the biological world, majority of them display structural similarities. Most myosin motor proteins with a few exceptions have a distinct N-terminal head or motor domain followed by the neck, trunk, and a C-terminal tail domain. Myosin motor domain carries ATPase activity and high binding affinity for actin filaments. These special features allow myosin motor protein to operate on actin filaments that are known to spread like intracellular tracks across cytoplasm. Myosin motor domains amino acid sequence is conserved among all species; however, the sequence of their tail domain remains variable. The tail domain with variable amino acid sequence perhaps carries different binding affinity for other cellular proteins and/or cargo. The proteins that bind tail domain could be a part of cargo and/or regulators of myosin motor activity in the cells. Proteins that bind myosin tail domain may determine the fate of intracellular destinations of these motor proteins. All these molecular machines with identical motor domains operate on common and or closely related intracellular tracks with different destinations ascribed to their tail domains in a given cellular world.

The myosin superfamily that consists of more than 145 members is categorized into different classes based on phylogenic analysis of their conserved heads, domain architectures, specific amino acid polymorphisms, and organismal distributions (Foth et al. 2006; Odronitz et al. 2007; Richards and Cavalier-Smith 2005). Roman numerals are assigned to each class of myosins. Names are given in an alphabetical order according to their discovery when more than one myosin of the same class is expressed in a given organism. According to the classification of the myosin superfamily members, Kuhne's myosin that was discovered for the first time in muscle extracts was identified as a class II myosin and therefore called conventional myosin motor protein and/or the founding member of the myosin superfamily.

The present chapter is focused on the current understanding of class II myosin motor proteins and particularly about their regulation and relevance to human health and diseases.

5.5 Class II Myosins/Myosin II Motor Proteins

Class II myosins or myosin II motor proteins are expressed in all eukaryotes but not in plants. Nearly three dozen class II myosins are reported throughout eukaryotic kingdom to date (Bagshaw 1993). At least one class II myosin is believed to be expressed in every eukaryotic cell. Based on their cell-type expressions and motor or tail domain sequences, class II myosins are further divided into four different subclasses. They are (1) yeast myosins, (2) Dictyostelium or Acanthamoeba myosins, (3) skeletal or cardiac or sarcomeric myosins, and (4) vertebrate smooth muscle or non-muscle myosins. Class II myosins are believed to be originated in ancestral eukaryotes unikonts with or without a single flagellum like amoebozoans, fungi, and holozoans (Richards and Cavalier-Smith 2005). While simple unicellular organisms like social amoeba (Dictyostelium) adopted a single class II myosin gene, complex multicellular organisms except fruit fly (Drosophila) acquired multiples of them during evolution. Fifteen of the total 40 myosin genes present in the human genome encode class II myosins (MYH1, MYH2, MYH3, MYH4, MYH6, MYH7, MYH7B, MYH8, MYH9, MYH10, MYH11, MYH13, MYH14, MYH15, MYH16); however, not all of them are active (Berg et al. 2001). Skeletal myosin and cardiac myosin encoding genes are located on chromosome 17 and chromosome 14 in humans, respectively. While sarcomeric myosin encoding gene is located on human chromosome 7, genes that encode smooth muscle myosins are located on chromosome 3 and 16. MYH11 located on chromosome 16 undergoes alternative splicing and encodes four distinct myosin II isoforms in human smooth muscle cells (Matsuoka et al. 1993). MYH8 is an important paralog of MYH11. MYH9, MYH10, and MYH14 genes located on different chromosomes encode myosin IIA, myosin IIB, and myosin IIC motor proteins, respectively. MYH9 located on chromosome 22 undergoes alternate splicing to express two isoforms in the cochlea (Li et al. 2008). MYH10 and MYH14 are located on chromosome 17 and 19, respectively. Both of these motor proteins express two tissue-specific splice variants (Kim et al. 2005, 2008; Ma et al. 2006). Interestingly, myosin IIA, myosin IIB, and myosin IIC are expressed exclusively in non-muscle cells, hence non-muscle myosin II motor proteins (Golomb et al. 2004; Leal et al. 2003; Simons et al. 1991; Toothaker et al. 1991). Myosin IIA, myosin IIB, and myosin IIC are expressed in every human non-muscle cell with a few exceptions. However, the expression of these motor proteins depends on cell and tissue types (Golomb et al. 2004; Kawamoto and Adelstein 1991). No tissue or cell type appears to express all three non-muscle myosin II motor proteins; however, many cell types express at least one or two of them under normal physiological conditions. The relative expression of myosin II motor proteins does not remain the same in all cell or tissue types. For instance, non-muscle myosin IIA and myosin IIB are expressed in endothelial and epithelial cells at similar levels. However, myosin IIB and myosin IIC are expressed abundantly in nervous and lung tissue, respectively. Non-muscle myosin IIA is the only conventional myosin II motor protein expressed in the circulating platelets (Maupin et al. 1994). Thus, preferential expression of myosin II motor proteins in different cell types reflects their specialization in driving separate, dedicated, and probably

nonredundant cellular functions. Why doesn't a single cell or tissue type express myosin IIA, myosin IIB, and myosin IIC motor proteins not clearly understood? Perhaps, preferential expression of myosin II paralogs is necessary for maintaining different cell and tissue types. No significant amount of non-muscle myosin II motor protein is reported in the circulatory system or in any other body fluid under normal physiological conditions. However, presence of myosin IIA and myosin IIB is reported in the urine of patients with the hereditary Alport syndrome and other kidney diseases (Pohl et al. 2013).

The remaining part of the present chapter is focused on the recent developments in our understanding of non-muscle myosin IIA, myosin IIB, and myosin IIC motor proteins.

5.6 Subcellular Localization of Non-muscle Myosin II Motor Proteins

Subcellular localization of a protein provides clues about its cellular functions in a given tissue. Non-muscle myosin II motor proteins associate mainly with actin filaments to generate intracellular contractile forces that are required for mediating cellular functions. Therefore, the subcellular localization of these molecular machines is linked to actin cytoskeleton dynamics that depend upon the state of a given cell. For instance, in a nonmotile quiescent cell, non-muscle myosin II motor proteins localize mostly in the cytoplasm, and some remain associated with actin filaments. However, these motor proteins are also found in the nuclei of proliferating myoblasts (Rodgers 2005). Interestingly, these cytosolic myosin II motor proteins undergo transient localization to cleavage furrow during cytokinesis by unknown mechanisms. It is largely believed that the myosin II-generated contractile forces mediate the separation of two daughter cells during cytokinesis; however, their specific roles and underlying mechanisms are not clearly understood. Recent studies have shown that non-muscle myosin IIB is required for the completion of meiotic cytokinesis in male but not in female mice (Yang et al. 2012). Myosin IIB has been shown to prevent endomitosis or polyploidization during differentiation of megakaryocytes (Lordier et al. 2012). Myosin IIB also regulates enucleation process in erythrocytes that is akin to cytokinesis in other cells (Ubukawa et al. 2012).

Cell migration has been viewed as an index of cellular life because of its importance in maintaining growth, development, normal physiology, and immunity of a given organism. Cells during this complex process display frequent changes in their shapes and make attachment with matrix and neighboring cells. The cell cytoskeleton plays critical roles in driving this essential cellular process. The actomyosin system, an essential component of the cytoskeleton, is extensively studied in migrating cells. The actomyosin system generates contractile forces to mediate cell dynamics during migration. However, nature has limited this extraordinary ability of migration to animal and other eukaryotic cells but not to plant cells. Unlike animal and other eukaryotic cells, plant cells cannot change their shapes, interact with matrix, and extend lamellipodia probably due to their rigid cell wall and absence of

class II myosins. Both mechanical and functional roles of myosin IIA and myosin IIB are extensively investigated in migrating cells for the past two decades. Many laboratories reported that myosin IIA and myosin IIB play specific roles in mediating cell dynamics during migration. Cells initiate migration by extending their membrane in the form of lamellipodia perhaps as a part of strategy to probe environment for favorable conditions and proper directions. During migration, cells display periodic extension and retraction of their lamellipodia by unknown mechanism. Interestingly, the cytosolic myosin IIA and myosin IIB translocate to the lamellipodia during cell migration. Both myosin IIA and myosin IIB display distinct localization and drive lamellipodia extension in opposite direction. On one hand, myosin IIB promotes lamellipodia and growth cone extension, and on the other, myosin IIA drives retraction of the lamellipodia during cell migration (Betapudi 2010; Brown and Bridgman 2003; Rochlin et al. 1995). However, the underlying mechanisms of myosin IIA and IIB transient localization to lamellipodia and their interacting proteins during cell migration are not clearly understood. The intracellular localization and specific roles of myosin IIC in driving cell migration are not clearly understood. Myosin II activity is also necessary for keratinocytes' migration, a critical step in the re-epithelialization of human skin wound. Both myosin IIA and myosin IIB display transient localization to the lamellipodia during keratinocytes' migration (Betapudi et al. 2010). Keratinocytes do not express myosin IIC motor proteins.

In addition to their localization to lamellipodia, myosin II motor proteins display specific subcellular localization in quiescent cells. Both myosin IIA and myosin IIB are found in the Golgi complex. Localization of myosin II in the nucleus and to nuclear membrane is yet to be uncovered. Myosin II motor proteins are not transmembrane proteins; however, they localize to cell membrane in order to mediate internalization of the cell surface receptors including EGFR and CXCR4 (Kim et al. 2012; Rey et al. 2007). Also, myosin II localizes to contractile vacuoles in water living microorganisms for unknown reasons. Contractile vacuoles are presumed to be homologues of lysosomes that are known to carry enzymes necessary for breakdown of waste materials and cellular debris in higher organisms. Myosin II-mediated mechanical forces have been implicated in operating contractile vacuoles probably to expel additional water and toxic materials from amoeba in hypo-osmotic conditions (Betapudi and Egelhoff 2009). Myosin II motor proteins are also found on the outside of the cells infected with virus and implicated in mediating viral infection (Arii et al. 2010; van Leeuwen et al. 2002). Localization of myosin II is yet to be identified in extracellular microvesicles that are known to make intercellular communications; however, their activity is necessary for microvesicle secretion from endothelial cells treated with antiphospholipid syndrome autoantibodies (Betapudi et al. 2010). The intracellular localization and specific roles of myosin II in cells undergoing apoptotic death are not clearly understood; however, their mediation is presumed to be required for the execution of cell death (Flynn and Helfman 2010; Solinet and Vitale 2008; Tang et al. 2011).

Thus, myosin II motor proteins display different subcellular localization to play specific roles in mediating cellular processes that are necessary for growth, development, and death. For instance, lower eukaryote like amoeba can survive with certain

developmental defects in the absence of myosin II motor protein (Xu et al. 1996). However, the absence of either myosin IIA or myosin IIB or myosin IIC motor protein is lethal for mouse embryo growth and development (Conti and Adelstein 2008).

5.7 Assembly of Non-muscle Myosin II Motor Protein in the Cells

Unlike many other proteins with enzyme activity, no single polypeptide alone is known to exist and function as a molecular machine in any biological system. Every motor protein exists as a globular multiprotein complex and performs specific functions in the cells. Similar to multiple components used in building a man-made machine, many polypeptides are involved in the assembly of a functional biological molecular machine. For example, cells use six polypeptides that are non-covalently associated to assemble a functional non-muscle myosin II motor protein complex with an average molecular weight of 520 kDa. Though the underlying assembly mechanism is not clearly not known, two myosin heavy chain (MHC) polypeptides and four light chain polypeptides build a functional non-muscle myosin II motor protein in the cells. The MHC of myosin IIA, myosin IIB, and myosin IIC motor proteins are encoded by MYH9, MYH10, and MYH14 genes in humans, respectively. Each MHC with an average molecular weight of 220 kDa has isoelectric point of 6.8. Every myosin II motor complex comprises homodimer of a specific MHC; however, recent studies show the existence of heterodimers due to a significant sequence similarity. The light chain polypeptides encoded by different non-myosin genes are divided into essential light chains (ELC) and regulatory light chains (RLC) based on their specific functional roles in operating myosin II motor complexes in cells. Each myosin II motor complex comprises two essential light chains and two regulatory light chains. Based on their extraction methods, ELC and RLC are also called alkali and 5,5′-dithiobis/2-nitrobenzoate (DTNB) light chains, respectively. Comparing with the size of MHC polypeptide, both ELC and RLC proteins are very small with 16 and 22 kDa molecular weight, respectively. While MHC-homodimer forms backbone of the myosin II motor complex, light chains mostly involve in controlling motor activity. Myosin heavy chains are specific for each motor protein complex; however, both ELC and RLC are commonly found in all myosin II motor protein complexes. Although tissue-specific alternatively spliced MHC, ELC, and RLC polypeptides are expressed in higher eukaryotes, their specific roles are not clearly understood to date. In addition to MHC, ELC, and RLC, no other protein has been identified as a component of non-muscle myosin II motor protein complex in the cells.

The assembly process of myosin II motor protein complex occurs in the Golgi complex. This understudied assembly process is mediated by UCS (UNC-45/Cro1/She4) chaperone and many other cellular proteins. As a part of this complicated assembly process, both myosin heavy and light chains undergo proper folding in order to build a functional myosin II motor protein complexes in the Golgi

apparatus (Gazda et al. 2013; Hellerschmied and Clausen 2014). This assembly process which is common for building myosin IIA, myosin IIB, and myosin IIC motor proteins remains elusive. Though the transcriptional regulation of ELC and RLC is poorly understood, the MHC expression of myosin IIA, myosin IIB, and myosin IIC is under the control of house-keeping promoter having no TATA element, a core sequence commonly found in the promoters of 24% human genes (Kawamoto 1994; Weir and Chen 1996). Thus, cells keep these multifunctional myosin II motor proteins readily available for mediating cellular functions. However, differential expression of MHC occurs in pathophysiological conditions. Serum and mitotic stimulants are known to induce differential expression of MHC (Kawamoto and Adelstein 1991; Toothaker et al. 1991). Differential expression of myosin II motor proteins has been attributed for the aggressive growth and metastasis of cancer cells.

5.8 Operation of Myosin II Motor Protein in the Cells

In response to external and internal cues, both heavy and light chains play specific roles in operating myosin II motor complexes in the cells. The heavy chains that form the backbone of the myosin II motor complex can be subdivided into a distinct head, neck, and tail domains. Each domain plays a specific role in building and operating these molecular machines in the cells. The N-terminus of the heavy chain starts with a globular head domain followed by a small neck region that is linked to a long alpha-helical tail domain. The N-terminal head domain carries ATPase activity in order to release free energy in the cells. Thus, head domain is also called motor domain or functional engine of these molecular machines. In addition to ATPase activity, motor domain carries high binding affinity for actin filaments. Because of its high binding affinity for actin filaments, the operation of these biological machines is restricted to actin filaments only despite cells building a complex network of intracellular tracks for trafficking and many other functions. Motor domain upon hydrolysis of ATP undergoes conformational change. This affects motor domain interaction with actin filaments, a key element of the cell strategy to generate mechanotransduction for performing various functions. The heavy chains of non-muscle myosin IIA, myosin IIB, and myosin IIC display a significant protein sequence similarity in their motor domains. However, all of them carry different binding affinities for actin filaments. Thus, myosin IIA, myosin IIB, and myosin IIC motor proteins mediate mechanotransduction with different energetic efficiencies in the cells. This probably suggests that myosin IIA, myosin IIB, and myosin IIC motor proteins are meant for playing different not redundant roles in the ells.

The motor domain of class II myosins follows a short neck region that consists of two conserved IQ motifs (IQxxxRGxxxR); however, myosins of other classes may have more or less than two IQ motifs (Cheney and Mooseker 1992). The IQ motifs of the neck region form an amphiphilic uninterrupted seven-turn α-helix that has high binding affinity for myosin light chains and/or calmodulin in $Ca^{+}2$-independent manner. This high binding affinity allows ELC and RLC to occupy the

first and second IQ motifs of the neck region, respectively. Both ELC and RLC by binding IQ motifs give stability to MHC. RLC also plays an additional role by offering functional regulation of MHC. The IQ motifs of the neck region allow light chains to acquire either compact or extended conformation. Thus, the light chains attached neck region functions like a linker and lever arm for myosin II motor domain to amplify energy conversion into mechanical work. The length of the neck region is presumed to have direct impact on motor speed and energy transduction into mechanical work (Uyeda et al. 1996). Except class XIV Toxoplasma myosin A, the neck region of every myosin carries IQ motifs (Heintzelman and Schwartzman 1997). The IQ motifs with approximately 25 amino acids in length are present in many other myosin heavy chains. This allows ELC to bind heavy chains of other myosins of class V, VI, and VII carrying IQ motifs; however, RLC binds exclusively to myosins of class II and XVIII (Chen et al. 2007; Tan et al. 2008).

The neck region of the motor domain is followed by a long tail domain with variable amino acid sequences. This long tail domain with coiled-coil α-helices ends into a short non-helical tailpiece. Interestingly, the coiled-coil tail domains of two heavy chains undergo homodimerization in order to form a single rodlike structure. This allows myosin II motor complex to have two functional engines or motor domains with a single coiled-coil rodlike structure, hence double-headed myosin II motor protein or double-engined molecular machine. This double-engined myosin II motor protein exists in compact and linear forms in the cells. Myosin II attains a compact folded conformation due to a "proline-kink" at the junction of head and rod domains and attachment of its C-terminal tail domain to RLC (Craig et al. 1983; Onishi and Wakabayashi 1982; Trybus et al. 1982). Thus, depending upon cell requirement, myosin II can exist either in linear or compact conformations in the cells. Myosin II with compact folded structure sediments at 10S (Svedberg) hence 10S form. Myosin II in 10S form displays high binding affinity for ADP and inorganic phosphate (Pi) and no ATPase activity (Cross et al. 1986, 1988). However, myosin II in the linear elongated conformation attained upon its C-terminal tail end detachment from RLC becomes active with high binding affinity for ATP. The active myosin II motor complex is an elongated form sediment at 6S and therefore called 6S form (Trybus and Lowey 1984). Interestingly, myosin II motor proteins in the elongated form have the tendency to assemble into a highly ordered parallel and antiparallel thick filament due to intermolecular interactions between coiled-coil rod domains. Thus, myosin II activation and formation into a thick filament is one of the most important steps in the process of generating contractile forces in the cells. RLC plays critical roles in regulating filamentation by controlling linear and compact formation of myosin II motor proteins in the cells. Myosin II tail domains form large aggregates without proper filamentation in the absence of RLC (Pastra-Landis and Lowey 1986; Rottbauer et al. 2006). Thus, the RLC-controlled myosin rod filamentation and motor domain interaction with actin filaments are the most important aspects of cell strategy for converting ATP-released cellular free energy into force and mechanical work using non-muscle myosin II motor proteins.

5.9 Regulation of RLC Phosphorylation and Myosin II Motor Activity

Despite carrying 60–80% sequence similarity at the amino acid level and same quaternary structure, non-muscle myosin IIA, myosin IIB, and myosin IIC paralogs appear to be diverged from a common ancestor more than 600 million years ago. However, in response to internal and external cues, these molecular machines undergo different regulatory mechanisms in the cells (Jung et al. 2008). The role of RLC in regulating myosin II motor protein activity is extensively studied in a wide variety of biological systems since its discovery in rabbit skeletal muscle myosins more than three decades ago (Casadei et al. 1984). RLC peptide does not exist alone in the cells; however, when remains are associated with the IQ motif of the neck region, it undergoes phosphorylation and dephosphorylation on its S1, S2,T9,T18, and S19 amino acids in order to turn on and turn off myosin II motor activity in the cells. RLC phosphorylation on S19 alone or on both T18 and S19 amino acids turns on myosin II motor activity with increased ATPase activity and elongated 6S conformation that allows simultaneous assembly of myosin rods into thick filaments (Betapudi et al. 2006, 2010; Somlyo and Somlyo 2003; Wendt et al. 2001). However, phosphorylation of RLC has no effect on the affinity of myosin motor domain for actin filaments in the cells (Sellers et al. 1982). Dephosphorylation of T18 and S19 amino acids or RLC phosphorylation on S1, S2, and S9 induces myosin II motor proteins to acquire compact 10S conformation with no myosin rods available for thick filamentation in the cells. Such site-specific phosphorylation of RLC turns-off myosin II motor protein activity in the cells. Thus, RLC by undergoing reversible phosphorylation on its certain specific amino acids plays a major role in regulating the activity of myosin II motor protein complexes in a wide variety of cell and tissue types.

We now know enough about the regulatory mechanisms of RLC reversible phosphorylation in normal and abnormal physiological conditions. RLC reversible phosphorylation is tightly regulated by both myosin-specific phosphatase and a wide variety of protein kinases in the cells. In response to external and internal cues, a wide variety of protein kinases phosphorylate RLC in order to regulate the activity of myosin II motor proteins in the cells. Protein kinases including myosin light chain kinase (MLCK/MYLK), rho-associated coiled-coil-containing kinase (ROCK), citron kinase or citron rho-interactive kinase (CRIK) or serine/threonine-protein kinase 21 (STK21), leucine zipper interacting protein kinase (ZIPK) or death associated protein kinase 3 (DAPK3), and myotonic dystrophy kinase-related CDC42-binding kinase (MRCK/CDC42BP) directly phosphorylate T18 and S19 amino acids of RLC to activate myosin II in the cells. However, protein kinase C (PKC) phosphorylates S1, S2, and S3 amino acids of RLC to inactivate myosin II in cells undergoing cytokinesis (Nishikawa et al. 1984). Interestingly, these RLC phosphorylating kinases not only display specific subcellular localizations but also respond to a wide variety of signaling pathways

in many settings. For instance, MLCK in response to $Ca^{+}2$-calmodulin phosphorylates RLC to activate myosin II that is localized next to cell membrane (Totsukawa et al. 2004). This site-specific subcellular localization and MLCK activation are governed by many other upstream protein kinases such as p21 activated kinase 1 (PAK1), Abl tyrosine kinase, Src, and arrest defective 1 in different cell types (Dudek et al. 2004; Sanders et al. 1999; Shin et al. 2008). A small GTP-binding protein RhoA activates ROCK and citron kinase in the central part of the cell. Shroom3, an actin binding protein, regulates ROCK subcellular localization and RLC phosphorylation in neuroepithelial cells (Haigo et al. 2003; Hildebrand 2005). Cell death and survival regulating DAPK3 displays nuclear localization and phosphorylates RLC in apoptotic cells in a $Ca2^{+}$/calmodulin-independent manner (Murata-Hori et al. 1999). PKC regulates myosin II activity by phosphorylating RLC in the presence of $Ca^{+}2$ and DAG (diacylglycerol) and/or phorbol esters in mitotic cells (Varlamova et al. 2001). The subcellular site-specific RLC phosphorylation, dephosphorylation, and myosin II activation are tightly controlled by protein phosphatase 1 (PP1), a ubiquitously expressed myosin-specific phosphatase in the cells (Matsumura and Hartshorne 2008; Rai and Egelhoff 2011; Xia et al. 2005). Protein kinases and phosphatases that are involved in regulating RLC phosphorylation also phosphorylate other substrates in the cells. For instance, MLCK phosphorylates a proline-rich protein tyrosine kinase 2 (PYK2/PTK2B) or focal adhesion kinase 2 (FAK2) that are known to promote lung vascular endothelial cell permeability during sepsis (Xu et al. 2008). ROCK directly phosphorylates LIM kinase and myosin phosphatase MYPT1, a regulatory subunit of PP1 in many types of cells and tissues (Kimura et al. 1996; Leung et al. 1996). MYPT1 phosphorylation results into inactivation of PP1, and that leads to a significant increase in RLC phosphorylation and myosin II activation in the cells. ZIPK, MRCK, and PKC are implicated in regulating MYPT1 phosphorylation in many cell and tissue types. PKC is also involved in phosphorylating MHC to regulate myosin II activity in the cells. Thus, a wide variety of protein kinases and phosphatases are involved in regulating RLC phosphorylation in higher organisms. However, lower eukaryotes appear to have a few kinases and phosphatases to regulate RLC phosphorylation and myosin II activity. For instance, MLCK-A is the only RLC phosphorylating kinase identified in *Dictyostelium discoideum* to date (Tan and Spudich 1990). Unlike MLCK in higher organisms, MLCK-A was shown to phosphorylate S13 of RLC in $Ca^{+}2$-calmodulin-independent manner (Tan and Spudich 1990). RLC phosphorylation on S13 amino acid increases myosin II motor activity and regulates cell morphological changes without affecting normal growth and development of *Dictyostelium discoideum* (Chen et al. 1994; Griffith et al. 1987; Liu et al. 1998; Matsumura 2005; Uyeda et al. 1996). Although many cellular proteins are now known to undergo more than 200 distinct posttranslational modifications with structural and functional diversity, except reversible phosphorylation, no other posttranslational modification of RLC with a role either in regulating myosin II motor activity or filamentation is reported to date.

5.10 Regulation of MHC Phosphorylation and Myosin II Motor Activity

Myosin heavy chain phosphorylation was first identified in macrophages more than three decades ago, and then its role was linked to myosin II filamentation and localization in lower eukaryotes such as Acanthamoeba and *Dictyostelium discoideum* (Barylko et al. 1986; Collins and Korn 1980; Egelhoff et al. 1993; Kuczmarski and Spudich 1980; Kuznicki et al. 1983; Pasternak et al. 1989; Trotter 1982; Trotter et al. 1985). We now know that MHC phosphorylation plays a major role in regulating myosin II activity in a wide variety of cell and tissue types. Recent advanced phosphorylation prediction tools revealed multiple putative phosphorylation sites with their cognate protein kinases on the heavy chains of non-muscle myosin IIA, myosin IIB, and myosin IIC. Although bioinformatic tools predicted multiple putative phosphorylation sites in the motor, neck, and tail domains of MHC, only a few sites with their cognate protein kinases in the coiled-coil and non-helical tail regions are known to date. For instance, the MHC of non-muscle myosin IIA with 1960 amino acids was predicted to undergo phosphorylation on 132 different residues; however, phosphorylation on only a few residues such as T1800, S1803, and S1808 in the coiled coil and S1943 in the non-helical tail regions were reported to date. While the MHC of myosin IIB with 1976 amino acids was predicted to undergo phosphorylation on 122 different residues, more than 135 putative phosphorylation sites were predicted on myosin IIC heavy chain with 1995 amino acids; however, phosphorylation of a few sites in the coiled-coil and non-helical tail regions of their C-terminal ends is identified to date (Dulyaninova and Bresnick 2013). It appears that MHC phosphorylation is regulated by tyrosine-, serine-, and threonine-specific kinases in the cells. Interestingly, majority of the putative phosphorylated amino acids in the motor domains of all three myosin heavy chains are tyrosine residues; however, their tail domains are heavily targeted by serine- and threonine-specific protein kinases. For instance, serine- and threonine-specific kinases like casein kinase 2 (CK2), PKC members, and alpha-kinase family members are now known to phosphorylate C-terminal ends of all three myosin heavy chains in normal physiological and pathological conditions (Clark et al. 2008a, b; Dulyaninova et al. 2005; Murakami et al. 1998; Ronen and Ravid 2009). PKC members phosphorylate S1916 and S1937 residues of myosin IIA and myosin IIB, respectively (Conti et al. 1991; Even-Faitelson and Ravid 2006). PKC members also phosphorylate other multiple serine residues in myosin IIB and threonine residues in myosin IIC coiled-coil regions (Murakami et al. 1998; Ronen and Ravid 2009). CK2 is known to phosphorylate S1943 residue in the non-helical tail region of myosin IIA in vitro, and its role has been implicated in regulating myosin II assembly and localization in pathological conditions. However, recent in vitro studies including chemical inhibition and siRNA-mediated depletion of CK2 showed no significant change in S1943 phosphorylation and breast cancer cell migration on fibronectin-coated surfaces (Betapudi et al. 2011). CK2 is also known to phosphorylate multiple serine and threonine residues in the coiled-coil and non-helical tail regions of non-muscle

myosin IIB and myosin IIC heavy chains in many cell types (Murakami et al. 1998; Ronen and Ravid 2009; Rosenberg et al. 2013). Therefore, the role of CK2 in regulating non-muscle myosin II-mediated cellular functions in certain specific pathological conditions cannot be ruled out. Protein kinases that target tyrosine residues of non-muscle myosin II motor domain are not identified.

The role of alpha-kinase family members in regulating MHC phosphorylation is extensively studied in mammals and social amoeba. Alpha-kinase family members are serine-threonine-protein kinases that belong to a small and unique group of protein kinases with catalytic domains having no significant similarity at amino acid level with the catalytic domains of conventional protein kinases (De la Roche et al. 2002; Middelbeek et al. 2010; Ryazanov et al. 1999; Scheeff and Bourne 2005). Unlike conventional protein kinases that are known to phosphorylate amino acid residues present in β-turns, loops, and irregular structures of their substrates, eukaryotic translation initiation factor 2 kinase (eIF2 kinase), the first member of the alpha-kinase family, shows unusual propensity to phosphorylate amino acids residues located in the α-turns of its cellular substrates hence α-kinase (Luck-Vielmetter et al. 1990; Vaillancourt et al. 1988). However, recent in vitro phosphorylation studies demonstrated that other members of alpha-kinase family also target amino acids of the non-alpha helical structures of their cellular substrates (Clark et al. 2008a; Jorgensen et al. 2003). Members of the alpha-kinase family are identified only in human and amoeba to date (Ryazanov et al. 1999; Scheeff and Bourne 2005). The human genome carries six different genes that encode alpha-kinases. Among them, transient receptor potential melastatin 6 (TRPM6) and transient receptor potential melastatin 7 (TRPM7) kinases are extensively studied to date. TRPM6 and TRPM7 are bifunctional protein kinases that belong to a large protein family of transient receptor potential cation channels. These protein kinases play critical roles in sensing mechanical stress, pain, temperature, taste, touch, and osmolarity (Mene et al. 2013; Middelbeek et al. 2010; Ramsey et al. 2006; Runnels 2011; Su et al. 2010). Both TRPM6 and TRPM7 kinases phosphorylate T1800, S1803, and S1808 residues in the coiled-coil region of MHC to control myosin IIA filamentation and association with actin filaments (Clark et al. 2008a, b). In addition, these multifunctional alpha-kinases phosphorylate many amino acids in the non-helical tail regions of myosin IIB and myosin IIC to control their filamentation and cellular roles. The MHC of myosin II is the only known substrate for TRPM6; however, TRPM7 can also phosphorylate annexin I or lipocortin I, a calcium/phospholipid-binding protein that promotes membrane fusion and exocytosis. The members of alpha-kinase family are also extensively studied in *Dictyostelium discoideum*. Dictyostelium expresses MHCK-A, MHCK-B, MHCK-C, MHCK-D, and vWFA alpha-kinases. Except vWFA kinase, all of them prefer to phosphorylate T1823, T1833, and T2029 residues in the tail region of myosin II in Dictyostelium (De la Roche et al. 2002; Egelhoff et al. 2005; Underwood et al. 2010; Yumura et al. 2005). Alpha-kinases by phosphorylating these sites control myosin II filamentation and play critical roles in regulating growth and development of *Dictyostelium discoideum*. Although vWFA kinase fails to phosphorylate myosin II heavy chain in vitro, but this special alpha-kinase regulates the expression and filamentation of myosin II heavy chain by

unknown mechanism (Betapudi et al. 2005). Unlike other alpha-kinase family members, vWFA kinase localizes to contractile vacuoles that are known to expel toxic metals and excess water from the cytoplasm of *Dictyostelium discoideum*. This special alpha-kinase appears to play critical roles in controlling the myosin II-mediated mechanical work implicated in regulating the dynamics of contractile vacuoles and survival of *Dictyostelium discoideum* in abnormal osmotic conditions; however, the underlying mechanisms are not clearly understood (Betapudi and Egelhoff 2009). It has been viewed that vWFA kinase protects *Dictyostelium discoideum* from osmotic shock death by regulating myosin II heavy chain expression and filamentation (Betapudi and Egelhoff 2009).

No phosphatase that is specific to the heavy chains of non-muscle myosin II motor proteins is reported in mammals to date. However, the expression of myosin II heavy chain phosphatase has been reported in *Dictyostelium discoideum* (Murphy and Egelhoff 1999). Therefore, not much is known about the mechanism of dephosphorylation of myosin II heavy chains in the cells.

5.11 Non-muscle Myosin II-Interacting Proteins

In addition to protein kinases and phosphatases, several other cellular proteins are involved in regulating non-muscle myosin II motor proteins in the cells. Some of these myosin II regulating non-enzymatic cellular proteins are S100A4, lethal giant larvae (Lgl), myosin binding protein H, and S100P. These proteins interact directly with myosin II heavy chain to control phosphorylation and filament assembly in flies and mammals (Du et al. 2012; Ford et al. 1997; Hosono et al. 2012; Kriajevska et al. 1994; Vasioukhin 2006). The Lgl protein was initially identified as a tumor suppressor protein in fruit fly and then implicated in regulating myosin II activity to control epithelial cell polarization and asymmetric cell division in higher organisms. Lgl forms a complex with C-terminal ends of myosin II heavy chains in the cells. However, the stability of this Lgl-myosin II complex depends on MHC phosphorylation by PKC in the cells (Betschinger et al. 2005; Kalmes et al. 1996; Plant et al. 2003; Strand et al. 1994). Thus, the Lgl protein by interacting with coiled-coil regions of the MHC controls myosin II filamentation and localization in proliferating cells (Dahan et al. 2012; De Lorenzo et al. 1999). Deletion of the Lgl protein encoding gene located on the human chromosome 17 is implicated in the development of Smith-Magenis syndrome, a developmental disorder that is known affect many body parts, intellectual disability, and sleep disturbances (De Leernyder et al. 2001; Smith et al. 1986). However, the role of mutant Lgl protein in controlling MHC phosphorylation and non-muscle myosin II cellular functions is not clearly understood. The metastasis factor mts1 or S100A4 or calvasculin that belongs to S100 family of calcium-binding proteins interacts with C-terminal ends of the MHC of non-muscle myosin II in the cells. Interaction of mts1 with C-terminal ends of the MHC of myosin IIA promotes phosphorylation on S1943 and disassembly of myosin II filamentation; however, the underlying mechanisms remain elusive to date (Badyal et al. 2011; Kiss et al. 2012; Li et al. 2003; Mitsuhashi et al. 2011). S100-P

or migration-inducing gene 9 protein (MIG9), another member of S100 family of calcium-binding proteins and a novel therapeutic target for cancer, is known to interact with MHC of non-muscle myosin II motor proteins in the cells. MIG9 is overexpressed in many cancer cells and is known to interact with the MHC of myosin IIA to induce disassembly of myosin II filamentation in migrating cells (Du et al. 2012). Myosin-binding protein H (MYBPH) is associated with measles disease. MYBPH interacts with ROCK1 and MHC to control RLC phosphorylation and myosin II filamentation in migrating cells; however, the underlying mechanisms remain elusive to date (Hosono et al. 2012). It has been shown that RLC phosphorylation and myosin II filamentation are essential for myosin II activation and cell migration, but recent studies suggest that the unassembled myosin II with phosphorylated RLC regulates initiation of focal adhesion complexes formation and lamellipodia extension during cell migration (Shutova et al. 2012). This may suggest that non-muscle myosin II motor proteins perform cellular functions without undergoing filamentation. It would be interesting to know how cells coordinate regulation of RLC and MCH phosphorylation to control non-muscle myosin II filamentation and cellular functions. Tropomyosin with two-stranded alpha-helical coiled-coil protein is an integral part of the actin cytoskeleton system in the cells. This cellular protein interacts with myosin II filaments in muscle and non-muscle cells. Tropomyosin is implicated in regulating myosin II localization to plasma membrane and stress fiber formation (Bryce et al. 2003). Supervillin, an actin filament binding and cell membrane-associated scaffolding protein, has been implicated in regulating non-muscle myosin II motor activity. Supervillin by interacting with MLCK controls RLC phosphorylation and myosin II activity in the cells (Takizawa et al. 2007). Thus, many cellular proteins at different levels are involved in regulating the activity of non-muscle myosin II motor proteins to mediate a wide variety of cellular functions.

5.12 Non-muscle Myosin II Motor Proteins Predispose Humans to Diseases

Plants can grow, develop, and live normal life without class II myosin motor proteins; however, other eukaryotes including human require these multifunctional molecular machines for their proper growth, development, and survival. Now we know that the MYH9 germline-ablated mice without the expression of non-muscle myosin IIA motor protein die on 6.5 embryonic day (E) because of defective cell-cell interaction and lack of polarized visceral endoderm (Conti et al. 2004). The MYH10 germline-ablated mice without the expression of another non-muscle myosin II motor protein myosin IIB were able to survive until E14.5 due to brain and cardiac developmental defects (Tullio et al. 1997, 2001). Although the MYH14-ablated mice that express no myosin IIC motor protein were able to live with no obvious developmental defects till adulthood, however, these mice require the expression of another non-muscle myosin II motor protein myosin IIB (Ma et al. 2010). Now we have enough evidence for the occurrence of mutations, misregulation, deletion, and

alternative splicing of MYH9, MYH10, and MYH14. The consequences of these changes at genetic level are the onset and progression of a wide variety of pathological conditions in humans. Nearly four dozen mutations are reported in myosin IIA encoding gene MYH9 to date (Saposnik et al. 2014). Interestingly, some of these mutations in MYH9 have been implicated in the development of a large number of autosomal-dominant disorders such as May-Hegglin anomaly, Sebastian platelet syndrome, Bernard-Soulier syndrome, Fechtner syndrome, Epstein syndrome, and Alport syndrome. These myosin IIA-linked autosomal diseases that are often worsen in late life of a patient are collectively called MYH9-related diseases (MYH9RD) (Balduini et al. 2011; Burt et al. 2008; Kelley et al. 2000; Pecci et al. 2008). The MYH9RD patients that carry R702C/H and R1165C/L mutations in the motor domain of non-muscle myosin IIA develop deafness, cataract, Döhle-like inclusions, nephritis, and thrombocytopenia with enlarged platelets in their middle age (De Rocco et al. 2013; Pecci et al. 2008, 2014). In addition, most MYH9RD patients develop renal diseases in their early adulthood. The circulating white blood cells of the MYH9RD patients carry non-muscle myosin IIA clumps with no cellular functions. Patients carrying D1424H/N/Y, V1516M, E1841K, and R1933X mutations in the tail domain of non-muscle myosin IIA live normal life with no symptoms of clinical relevance (Pecci et al. 2010). Abnormal expression of non-muscle myosin II motor proteins has been implicated in rendering humans to diseases. For instance, overexpression of myosin IIA was thought to increase cancer cell migration and metastasis as well as lung and kidney tumor invasion (Derycke et al. 2011; Gupton and Waterman-Storer 2006; Xia et al. 2012); however, this hypothesis has lost appreciation because of recent reports of myosin IIA roles in the posttranscriptional stabilization of tumor suppressor protein p53 and repression of squamous cell carcinoma in mice (Schramek et al. 2014). More than a decade ago, a chimeric MYH9-Alk transcript formed by the fusion of MYH9 and ALK (anaplastic lymphoma kinase) was reported in anaplastic large cell lymphoma; however, its disease relevance is not clearly established to date (Lamant et al. 2003). Polymorphisms in the MYH9 and adjacent APOL1 (apolipoprotein L1) have been implicated in the development of nondiabetic chronic kidney disease in African-Americans (O'Seaghdha et al. 2011). Although mutations in MYH10 that have relevance to human diseases with any clinical symptom are not reported, recently an E908X de novo mutation was identified in patients with microcephaly, hydrocephalus, cerebral, and cerebellar atrophy. Polymorphisms in MYH10 and some other genes have been linked to the development of abnormal heart with enlarged left atrium which was reported in a Caribbean Hispanic patient (Wang et al. 2010a). No direct link has been established in between myosin IIB expression and disease development to date. However, we now have enough evidence for establishing an indirect link in between the expression of myosin IIB and progression of a wide variety of pathological conditions including megakaryopoiesis, myocardial infarction, scar tissue formation, demyelination, and juvenile-onset neuronal ceroid lipofuscinosis (JCNL) or Batten disease (Antony-Debre et al. 2012). JCNL or Batten disease is a lysosomal storage disorder caused by mutations in CLN3 that encode a lysosomal membrane-binding chaperone that is known to bind non-muscle myosin IIB motor protein in the cells. Mutations in CLN3

not only inhibit its interaction with myosin IIB motor protein but also affect retrograde and anterograde trafficking in the Golgi complexes (Getty et al. 2011). Although myosin IIB role is not clearly established, many patients with CLN3 mutations show symptoms of dementia, seizures, loss of vision, and psychomotor disturbances (Cotman and Staropoli 2012). Abnormal regulation of non-muscle myosin IIC motor protein is also implicated in the development and progression of several diseases. For instance, mutations in MYH14 have been linked to the development of hoarseness, hereditary blindness (DFNA4), myopathy, and peripheral neuropathy (Choi et al. 2011; Donaudy et al. 2004). Expression of aberrant splicing products of MYH14 has been implicated in the development and progression of myotonic dystrophy type 1 (DM1), a multisystem genetic disorder that is known to affect 1 in 8000 people worldwide (Kumar et al. 2013; Rinaldi et al. 2012).

Abnormal regulation of myosin II-interacting proteins is also implicated in the development and progression of many diseased conditions in humans. Overexpression of ROCK and Mts1 that are known to regulate myosin II phosphorylation and filamentation is implicated in causing enhanced cancer cell migration, an essential step in metastasis and invasion (Boye and Maelandsmo 2010; Kim and Adelstein 2011; Sandquist et al. 2006). Mutations in RLC are known to affect singing male courtship song in fruit flies (Chakravorty et al. 2014). Mutations in MYLK, another RLC phosphorylating kinase, have been implicated in cancer development and progression (Greenman et al. 2007). Mutations in MYLK are linked to the development of Marfan syndrome and Ehlers-Danlos syndrome. Patients with these syndromes develop familial aortic dissections (FAD) or actual aortic tearing openings that may cause sudden death (Wang et al. 2010b). In addition, race-specific single nucleotide polymorphism variants of MYLK are implicated in the development and progression of asthma, acute lung injury, and sepsis (Flores et al. 2007; Gao et al. 2006, 2007). Mutations in TRPM6 that encodes an ion-channel kinase involved in phosphorylating myosin II heavy chain are known to cause hypomagnesemia in patients with secondary hypocalcemia (Schlingmann et al. 2002; Walder et al. 2002). Abnormal regulation of another ion-channel kinase TRPM7 that is known to phosphorylate myosin II heavy chain has been linked to Guamanian amyotrophic lateral sclerosis and parkinsonian dementia (ALS/PD), various forms of neoplasia, hypertension, and delayed neuronal death following cerebral ischemia (Bates-Withers et al. 2011).

Many pathogens are believed to manipulate and hijack these multifunctional molecular machines in order to promote their infection and pathogenesis. For instance, herpes simplex virus type 1 is believed to hijack myosin II motor protein for promoting its egression (Arii et al. 2010; van Leeuwen et al. 2002), murine leukemia virus for efficient infection (Lehmann et al. 2005), and Salmonella bacteria for promoting growth in macrophages (Wasylnka et al. 2008); however, the underlying mechanisms are not clearly understood to date. Another pathogen Kaposi's sarcoma herpes simplex virus that is known to cause AIDS-related neoplasm is believed to manipulate non-muscle myosin II and E3-ubiquitin ligase c-Cbl-mediated signaling pathway to induce macropinocytosis as a part of mechanism to infect blood vessels (Sharma-Walia et al. 2010). Interestingly, certain pathogens like HIV-1 that is known to cause renal disease is believed to inactivate non-muscle myosin IIA motor protein

selectively in the kidney in order to escape clearance through urine (Hays et al. 2012). Dengue virus type 2, a mosquito-borne single positive-stranded RNA virus, stimulates Rac1- and Cdc42-mediated signaling pathway to activate myosin II motor proteins for successful infection of host cells (Zamudio-Meza et al. 2009). Respiratory syncytial virus (RSV) that is known to cause runny nose, cough, headache, and severe respiratory tract infections is believed to activate non-muscle myosin II motor protein and other cytoskeletal proteins for rapid and efficient internalization during infection (Krzyzaniak et al. 2013). Non-muscle myosin II motor proteins play essential roles in maintaining host defense system. However, certain pathogens including hepatitis C virus weakens host defense system by inducing development of autoantibodies against non-muscle myosin IIA motor protein perhaps as a part of escape strategy from host defense network (von Muhlen et al. 1995).

5.13 Conclusions and Future Perspectives

If it is not surprising to say the transfer of energy from molecule to molecule and one object to another is the key to life, efficiency undoubtedly is the essence of life. Here, we have natural molecular machines as the most efficient convertors of cellular free energy into biological work that is essential for the sustenance of life. Among all the biological motor proteins known to date, perhaps class II myosins, especially non-muscle myosin IIA, myosin IIB, and myosin IIC motor proteins, have emerged as the main mechanotransducers of cellular free energy into work that is necessary for performing multiple biological processes ranging from birth to death in mammals' life. Although the discovery of the first motor protein dates back to the nineteenth century, research done during the past 30 years has led us to understand much about the underlying mechanisms of several myosin II-mediated cellular processes in many biological systems. We now know enough about these molecular machines and have proven beyond doubt that murine life does not exist without the expression of non-muscle myosin II motor proteins (Conti and Adelstein 2008). Interestingly, many patients with abnormally regulated and mutated non-muscle myosin IIA, myosin IIB, and myosin IIC motor proteins are reported; however, no patient lacks these biological molecular machines. Extrapolation of such findings with caution may suggest that life in human and other mammals cannot exist without the expression of non-muscle myosin II motor proteins. Therefore, the emergence of myosin II encoding genes is probably a turning point in the evolution of mammals. Although nature chose not to give myosin II genes to plants during evolution, mammals have acquired Myh10, Myh11, and Myh14 genes with a significant homology in nucleotide sequence. It is largely accepted that the expression of all three functional non-muscle myosin II motor proteins in humans is required to maintain proper growth, development, and immunity. Each cell and tissue type in humans is known to display differential expression of these myosin II paralogs; however, the reasons behind this differential expression remain elusive to date. Part of the reasons could be due to their specialization in mediating dedicated cellular functions that are specific to each cell and tissue type. However, this hypothesis will benefit from further understanding of

structural and posttranslational modifications of these molecular machines. We have made progress in identifying dozens of mutations in myosin II motors proteins and their regulating proteins; however, not much is known about many myosin II mutations and their clinical relevance to date. Therefore, development of novel strategies for the management and diagnosis of MYH9RD patients are necessary (Althaus and Greinacher 2010). We need to improve our current understanding of MYH9RD patients in order to develop myosin II-based novel therapeutic approaches in future. Although it is not unequivocally proven yet, many modern cancer cell biologists believe that non-muscle myosin II motor proteins that are known to drive cell migration and cytokinesis go awry in cancer and other diseased conditions. Overexpression of a particular myosin II motor protein has been implicated in mediating cancer progression and metastasis; however, further understanding of the expression profile of each motor protein in every cancer type is necessary to design and develop myosin II-based therapeutics in future. Also, we need to increase our limited knowledge on the expression of chimeric and alternate splicing products of myosin II motor proteins in pathological conditions in order to develop treatment options. During the past 30 years, we made a significant advancement in understanding of myosin II motor proteins in normal physiological conditions; however, we made very limited progress on understanding how pathogens hijack myosin II motor proteins for their efficient infection and propagation. Therefore, understanding what made these dedicated biological molecular machines work for the interests of pathogens is no less than a challenge to modern cell biologists. We still have limited knowledge on how myosin II motor proteins mediate release of extracellular microvesicles that are known to make intercellular communications and promote progression of many human diseases. Non-muscle myosin II-mediated mechanotransduction has been implicated in stem cell proliferation and differentiation (Chen et al. 2014); however, very little is known about the mechanical roles of myosin II paralogs. Therefore, further understanding of these biological machines will have a significant impact on stem cell-based tissue engineering, synthetic bioengineering, and therapeutic development.

Acknowledgements The present chapter was written with a particular theme on non-muscle myosin II motor proteins. The author apologizes for the omission of any study that has direct relevance with the main theme of the present chapter. The present work would not have been possible without the mentorship of author's previous mentors Professors Anil Kumar, Seyed E. Hasnain, and Thomas T. Egelhoff as well as private and public funding agencies including the US National Institute of Health.

Competing Interests No competing interests to be disclosed.

References

Althaus K, Greinacher A. MYH-9 related platelet disorders: strategies for management and diagnosis. Transfus Med Hemother. 2010;37(5):260–7.

Antony-Debre I, et al. MYH10 protein expression in platelets as a biomarker of RUNX1 and FLI1 alterations. Blood. 2012;120(13):2719–22.

Arii J, et al. Non-muscle myosin IIA is a functional entry receptor for herpes simplex virus-1. Nature. 2010;467(7317):859–62.

Badyal SK, et al. Mechanism of the Ca(2)+-dependent interaction between S100A4 and tail fragments of nonmuscle myosin heavy chain IIA. J Mol Biol. 2011;405(4):1004–26.
Bagshaw CR. Muscle contraction. 2nd ed. London: Chapman & Hall; 1993.
Balduini CL, Pecci A, Savoia A. Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias. Br J Haematol. 2011;154(2):161–74.
Baluska F, Cvrckova F, Kendrick-Jones J, Volkmann D. Sink plasmodesmata as gateways for phloem unloading. Myosin VIII and calreticulin as molecular determinants of sink strength? Plant Physiol. 2001;126(1):39–46.
Barylko B, Tooth P, Kendrick-Jones J. Proteolytic fragmentation of brain myosin and localisation of the heavy-chain phosphorylation site. Eur J Biochem. 1986;158(2):271–82.
Bates-Withers C, Sah R, Clapham DE. TRPM7, the Mg(2+) inhibited channel and kinase. Adv Exp Med Biol. 2011;704:173–83.
Berg JS, Powell BC, Cheney RE. A millennial myosin census. Mol Biol Cell. 2001;12(4): 780–94.
Betapudi V. Myosin II motor proteins with different functions determine the fate of lamellipodia extension during cell spreading. PLoS One. 2010;5(1):e8560.
Betapudi V. Life without double-headed non-muscle myosin II motor proteins. Front Chem. 2014;2:45.
Betapudi V, Egelhoff TT. Roles of an unconventional protein kinase and myosin II in amoeba osmotic shock responses. Traffic. 2009;10(12):1773–84.
Betapudi V, Mason C, Licate L, Egelhoff TT. Identification and characterization of a novel alpha-kinase with a von Willebrand factor A-like motif localized to the contractile vacuole and Golgi complex in Dictyostelium discoideum. Mol Biol Cell. 2005;16(5):2248–62.
Betapudi V, Licate LS, Egelhoff TT. Distinct roles of nonmuscle myosin II isoforms in the regulation of MDA-MB-231 breast cancer cell spreading and migration. Cancer Res. 2006;66(9): 4725–33.
Betapudi V, Rai V, Beach JR, Egelhoff T. Novel regulation and dynamics of myosin II activation during epidermal wound responses. Exp Cell Res. 2010;316(6):980–91.
Betapudi V, Gokulrangan G, Chance MR, Egelhoff TT. A proteomic study of myosin II motor proteins during tumor cell migration. J Mol Biol. 2011;407(5):673–86.
Betapudi V, Lominadze G, Hsi L, Willard B, Wu M, McCrae KR. Anti-beta2GPI antibodies stimulate endothelial cell microparticle release via a nonmuscle myosin II motor protein-dependent pathway. Blood. 2013;122(23):3808–17.
Betschinger J, Eisenhaber F, Knoblich JA. Phosphorylation-induced autoinhibition regulates the cytoskeletal protein Lethal (2) giant larvae. Curr Biol. 2005;15(3):276–82.
Bianconi E, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40(6):463–71.
Boye K, Maelandsmo GM. S100A4 and metastasis: a small actor playing many roles. Am J Pathol. 2010;176(2):528–35.
Brown ME, Bridgman PC. Retrograde flow rate is increased in growth cones from myosin IIB knockout mice. J Cell Sci. 2003;116(Pt 6):1087–94.
Bryce NS, et al. Specification of actin filament function and molecular composition by tropomyosin isoforms. Mol Biol Cell. 2003;14(3):1002–16.
Burt RA, Joseph JE, Milliken S, Collinge JE, Kile BT. Description of a novel mutation leading to MYH9-related disease. Thromb Res. 2008;122(6):861–3.
Bustamante C, Chemla YR, Forde NR, Izhaky D. Mechanical processes in biochemistry. Annu Rev Biochem. 2004;73:705–48.
Casadei JM, Gordon RD, Lampson LA, Schotland DL, Barchi RL. Monoclonal antibodies against the voltage-sensitive Na+ channel from mammalian skeletal muscle. Proc Natl Acad Sci USA. 1984;81(19):6227–31.
Chakravorty S, Vu H, Foelber V, Vigoreaux JO. Mutations of the Drosophila myosin regulatory light chain affect courtship song and reduce reproductive success. PLoS One. 2014;9(2):e90077.
Chen P, Ostrow BD, Tafuri SR, Chisholm RL. Targeted disruption of the Dictyostelium RMLC gene produces cells defective in cytokinesis and development. J Cell Biol. 1994;127(6 Pt 2): 1933–44.

Chen Z, et al. The May-Hegglin anomaly gene MYH9 is a negative regulator of platelet biogenesis modulated by the Rho-ROCK pathway. Blood. 2007;110(1):171–9.
Chen AK, Chen X, Lim YM, Reuveny S, Oh SK. Inhibition of ROCK-myosin II signaling pathway enables culturing of human pluripotent stem cells on microcarriers without extracellular matrix coating. Tissue Eng Part C Methods. 2014;20(3):227–38.
Cheney RE, Mooseker MS. Unconventional myosins. Curr Opin Cell Biol. 1992;4(1):27–35.
Choi BO, et al. A complex phenotype of peripheral neuropathy, myopathy, hoarseness, and hearing loss is linked to an autosomal dominant mutation in MYH14. Hum Mutat. 2011;32(6): 669–77.
Clark K, et al. The alpha-kinases TRPM6 and TRPM7, but not eEF-2 kinase, phosphorylate the assembly domain of myosin IIA, IIB and IIC. FEBS Lett. 2008a;582(20):2993–7.
Clark K, et al. TRPM7 regulates myosin IIA filament stability and protein localization by heavy chain phosphorylation. J Mol Biol. 2008b;378(4):790–803.
Collins JH, Korn ED. Actin activation of Ca2+-sensitive Mg2+-ATPase activity of Acanthamoeba myosin II is enhanced by dephosphorylation of its heavy chains. J Biol Chem. 1980;255(17):8011–4.
Conti MA, Adelstein RS. Nonmuscle myosin II moves in new directions. J Cell Sci. 2008;121(Pt 1):11–8.
Conti MA, Sellers JR, Adelstein RS, Elzinga M. Identification of the serine residue phosphorylated by protein kinase C in vertebrate nonmuscle myosin heavy chains. Biochemistry. 1991; 30(4):966–70.
Conti MA, Even-Ram S, Liu C, Yamada KM, Adelstein RS. Defects in cell adhesion and the visceral endoderm following ablation of nonmuscle myosin heavy chain II-A in mice. J Biol Chem. 2004;279(40):41263–6.
Cotman SL, Staropoli JF. The juvenile Batten disease protein, CLN3, and its role in regulating anterograde and retrograde post-Golgi trafficking. Clin Lipidol. 2012;7(1):79–91.
Craig R, Smith R, Kendrick-Jones J. Light-chain phosphorylation controls the conformation of vertebrate non-muscle and smooth muscle myosin molecules. Nature. 1983;302(5907):436–9.
Cross RA, Cross KE, Sobieszek A. ATP-linked monomer-polymer equilibrium of smooth muscle myosin: the free folded monomer traps ADP.Pi. EMBO J. 1986;5(10):2637–41.
Cross RA, Jackson AP, Citi S, Kendrick-Jones J, Bagshaw CR. Active site trapping of nucleotide by smooth and non-muscle myosins. J Mol Biol. 1988;203(1):173–81.
Dahan I, Yearim A, Touboul Y, Ravid S. The tumor suppressor Lgl1 regulates NMII-A cellular distribution and focal adhesion morphology to optimize cell migration. Mol Biol Cell. 2012;23(4):591–601.
De la Roche MA, Smith JL, Betapudi V, Egelhoff TT, Cote GP. Signaling pathways regulating Dictyostelium myosin II. J Muscle Res Cell Motil. 2002;23(7–8):703–18.
De Leersnyder H, de Blois MC, Vekemans M, Sidi D, Villain E, Kindermans C, Munnich A. beta(1)-adrenergic antagonists improve sleep and behavioural disturbances in a circadian disorder, Smith-Magenis syndrome. J Med Genet. 2001;38(9):586–90.
De Lorenzo C, Mechler BM, Bryant PJ. What is Drosophila telling us about cancer? Cancer Metastasis Rev. 1999;18(2):295–311.
De Rocco D, et al. MYH9-related disease: five novel mutations expanding the spectrum of causative mutations and confirming genotype/phenotype correlations. Eur J Med Genet. 2013;56(1):7–12.
Derycke L, Stove C, Vercoutter-Edouart AS, De Wever O, Dolle L, Colpaert N, Depypere H, Michalski JC, Bracke M. The role of non-muscle myosin IIA in aggregation and invasion of human MCF-7 breast cancer cells. Int J Dev Biol. 2011;55(7–9):835–40.
Donaudy F, et al. Nonmuscle myosin heavy-chain gene MYH14 is expressed in cochlea and mutated in patients affected by autosomal dominant hearing impairment (DFNA4). Am J Hum Genet. 2004;74(4):770–6.
Du M, et al. S100P dissociates myosin IIA filaments and focal adhesion sites to reduce cell adhesion and enhance cell migration. J Biol Chem. 2012;287(19):15330–44.
Dudek SM, et al. Pulmonary endothelial cell barrier enhancement by sphingosine 1-phosphate: roles for cortactin and myosin light chain kinase. J Biol Chem. 2004;279(23):24692–700.
Dulyaninova NG, Bresnick AR. The heavy chain has its day: regulation of myosin-II assembly. Bioarchitecture. 2013;3(4):77–85.

Dulyaninova NG, Malashkevich VN, Almo SC, Bresnick AR. Regulation of myosin-IIA assembly and Mts1 binding by heavy chain phosphorylation. Biochemistry. 2005;44(18):6867–76.
Egelhoff TT, Lee RJ, Spudich JA. Dictyostelium myosin heavy chain phosphorylation sites regulate myosin filament assembly and localization in vivo. Cell. 1993;75(2):363–71.
Egelhoff TT, Croft D, Steimle PA. Actin activation of myosin heavy chain kinase A in Dictyostelium: a biochemical mechanism for the spatial regulation of myosin II filament disassembly. J Biol Chem. 2005;280(4):2879–87.
Engelhardt WA, Liubimova MN. [Myosin and adenosine triphosphatase (Nature, 144, 688, Oct. 14, 1939)]. Mol Biol (Mosk). 1994;28(6):1229–30.
Even-Faitelson L, Ravid S. PAK1 and aPKCzeta regulate myosin II-B phosphorylation: a novel signaling pathway regulating filament assembly. Mol Biol Cell. 2006;17(7):2869–81.
Flores C, Ma SF, Maresso K, Ober C, Garcia JG. A variant of the myosin light chain kinase gene is associated with severe asthma in African Americans. Genet Epidemiol. 2007;31(4):296–305.
Flynn PG, Helfman DM. Non-muscle myosin IIB helps mediate TNF cell death signaling independent of actomyosin contractility (AMC). J Cell Biochem. 2010;110(6):1365–75.
Ford HL, Silver DL, Kachar B, Sellers JR, Zain SB. Effect of Mts1 on the structure and activity of nonmuscle myosin II. Biochemistry. 1997;36(51):16321–7.
Foth BJ, Goedecke MC, Soldati D. New insights into myosin evolution and classification. Proc Natl Acad Sci USA. 2006;103(10):3681–6.
Gao L, et al. Novel polymorphisms in the myosin light chain kinase gene confer risk for acute lung injury. Am J Respir Cell Mol Biol. 2006;34(4):487–95.
Gao L, et al. Macrophage migration inhibitory factor in acute lung injury: expression, biomarker, and associations. Transl Res. 2007;150(1):18–29.
Gazda L, Pokrzywa W, Hellerschmied D, Lowe T, Forne I, Mueller-Planitz F, Hoppe T, Clausen T. The myosin chaperone UNC-45 is organized in tandem modules to support myofilament formation in C. elegans. Cell. 2013;152(1–2):183–95.
Getty AL, Benedict JW, Pearce DA. A novel interaction of CLN3 with nonmuscle myosin-IIB and defects in cell motility of Cln3(−/−) cells. Exp Cell Res. 2011;317(1):51–69.
Golomb E, Ma X, Jana SS, Preston YA, Kawamoto S, Shoham NG, Goldin E, Conti MA, Sellers JR, Adelstein RS. Identification and characterization of nonmuscle myosin II-C, a new member of the myosin II family. J Biol Chem. 2004;279(4):2800–8.
Greenman C, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446(7132):153–8.
Griffith LM, Downs SM, Spudich JA. Myosin light chain kinase and myosin light chain phosphatase from Dictyostelium: effects of reversible phosphorylation on myosin structure and function. J Cell Biol. 1987;104(5):1309–23.
Gupton SL, Waterman-Storer CM. Spatiotemporal feedback between actomyosin and focal-adhesion systems optimizes rapid cell migration. Cell. 2006;125(7):1361–74.
Haigo SL, Hildebrand JD, Harland RM, Wallingford JB. Shroom induces apical constriction and is required for hingepoint formation during neural tube closure. Curr Biol. 2003;13(24): 2125–37.
Hays T, D'Agati VD, Garellek JA, Warren T, Trubin ME, Hyink DP, He JC, Klotman PE. Glomerular MYH9 expression is reduced by HIV-1. AIDS. 2012;26(7):797–803.
Heintzelman MB, Schwartzman JD. A novel class of unconventional myosins from Toxoplasma gondii. J Mol Biol. 1997;271(1):139–46.
Hellerschmied D, Clausen T. Myosin chaperones. Curr Opin Struct Biol. 2014;25:9–15.
van den Heuvel MG, Bolhuis S, Dekker C. Persistence length measurements from stochastic single-microtubule trajectories. Nano Lett. 2007;7(10):3138–44.
Hildebrand JD. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. J Cell Sci. 2005;118(Pt 22):5191–203.
Hosono Y, Usukura J, Yamaguchi T, Yanagisawa K, Suzuki M, Takahashi T. MYBPH inhibits NM IIA assembly via direct interaction with NMHC IIA and reduces cell motility. Biochem Biophys Res Commun. 2012;428(1):173–8.
Howard J. Jonathon Howard: motor proteins go walkabout. Interview by Caitlin Sedwick. J Cell Biol. 2014;204(2):150–1.

Jorgensen R, Ortiz PA, Carr-Schmid A, Nissen P, Kinzy TG, Andersen GR. Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. Nat Struct Biol. 2003;10(5):379–85.

Jung HS, Burgess SA, Billington N, Colegrave M, Patel H, Chalovich JM, Chantler PD, Knight PJ. Conservation of the regulated structure of folded myosin 2 in species separated by at least 600 million years of independent evolution. Proc Natl Acad Sci USA. 2008;105(16):6022–6.

Kabir AM, Kakugo A, Gong JP, Osada Y. How to integrate biological motors towards bio-actuators fueled by ATP. Macromol Biosci. 2011;11(10):1314–24.

Kalmes A, Merdes G, Neumann B, Strand D, Mechler BM. A serine-kinase associated with the p127-l(2)gl tumour suppressor of Drosophila may regulate the binding of p127 to nonmuscle myosin II heavy chain and the attachment of p127 to the plasma membrane. J Cell Sci. 1996;109(Pt 6):1359–68.

Kawamoto S. Evidence for an internal regulatory region in a human nonmuscle myosin heavy chain gene. J Biol Chem. 1994;269(21):15101–10.

Kawamoto S, Adelstein RS. Chicken nonmuscle myosin heavy chains: differential expression of two mRNAs and evidence for two different polypeptides. J Cell Biol. 1991;112(5):915–24.

Kelley MJ, Jawien W, Ortel TL, Korczak JF. Mutation of MYH9, encoding non-muscle myosin heavy chain A, in May-Hegglin anomaly. Nat Genet. 2000;26(1):106–8.

Kelly DA. The functional morphology of penile erection: tissue designs for increasing and maintaining stiffness. Integr Comp Biol. 2002;42(2):216–21.

Kier WM. The diversity of hydrostatic skeletons. J Exp Biol. 2012;215(Pt 8):1247–57.

Kim JH, Adelstein RS. LPA(1)-induced migration requires nonmuscle myosin II light chain phosphorylation in breast cancer cells. J Cell Physiol. 2011;226(11):2881–93.

Kim KY, Kovacs M, Kawamoto S, Sellers JR, Adelstein RS. Disease-associated mutations and alternative splicing alter the enzymatic and motile activity of nonmuscle myosins II-B and II-C. J Biol Chem. 2005;280(24):22769–75.

Kim KY, Kawamoto S, Bao J, Sellers JR, Adelstein RS. The B2 alternatively spliced isoform of nonmuscle myosin II-B lacks actin-activated MgATPase activity and in vitro motility. Biochem Biophys Res Commun. 2008;369(1):124–34.

Kim JH, Wang A, Conti MA, Adelstein RS. Nonmuscle myosin II is required for internalization of the epidermal growth factor receptor and modulation of downstream signaling. J Biol Chem. 2012;287(33):27345–58.

Kimura K, et al. Regulation of myosin phosphatase by Rho and Rho-associated kinase (Rho-kinase). Science. 1996;273(5272):245–8.

Kiss B, Duelli A, Radnai L, Kekesi KA, Katona G, Nyitray L. Crystal structure of the S100A4-nonmuscle myosin IIA tail fragment complex reveals an asymmetric target binding mechanism. Proc Natl Acad Sci USA. 2012;109(16):6048–53.

Kolomeisky AB. Motor proteins and molecular motors: how to operate machines at the nanoscale. J Phys Condens Matter. 2013;25(46):463101.

Kriajevska MV, et al. Non-muscle myosin heavy chain as a possible target for protein encoded by metastasis-related mts-1 gene. J Biol Chem. 1994;269(31):19679–82.

Krzyzaniak MA, Zumstein MT, Gerez JA, Picotti P, Helenius A. Host cell entry of respiratory syncytial virus involves macropinocytosis followed by proteolytic activation of the F protein. PLoS Pathog. 2013;9(4):e1003309.

Kuczmarski ER, Spudich JA. Regulation of myosin self-assembly: phosphorylation of Dictyostelium heavy chain inhibits formation of thick filaments. Proc Natl Acad Sci USA. 1980;77(12):7292–6.

Kühne W. Untersuchungen über das Protoplasma und die Contractilitat. Leipzig: W. Engelmann; 1864.

Kumar A, Agarwal S, Agarwal D, Phadke SR. Myotonic dystrophy type 1 (DM1): a triplet repeat expansion disorder. Gene. 2013;522(2):226–30.

Kuznicki J, Albanesi JP, Cote GP, Korn ED. Supramolecular regulation of the actin-activated ATPase activity of filaments of Acanthamoeba Myosin II. J Biol Chem. 1983;258(10):6011–4.

Lamant L, et al. Non-muscle myosin heavy chain (MYH9): a new partner fused to ALK in anaplastic large cell lymphoma. Genes Chromosomes Cancer. 2003;37(4):427–32.

Leal A, et al. A novel myosin heavy chain gene in human chromosome 19q13.3. Gene. 2003;312:165–71.
van Leeuwen H, Elliott G, O'Hare P. Evidence of a role for nonmuscle myosin II in herpes simplex virus type 1 egress. J Virol. 2002;76(7):3471–81.
Lehmann MJ, Sherer NM, Marks CB, Pypaert M, Mothes W. Actin- and myosin-driven movement of viruses along filopodia precedes their entry into cells. J Cell Biol. 2005;170(2):317–25.
Leung T, Chen XQ, Manser E, Lim L. The p160 RhoA-binding kinase ROK alpha is a member of a kinase family and is involved in the reorganization of the cytoskeleton. Mol Cell Biol. 1996;16(10):5313–27.
Li ZH, Spektor A, Varlamova O, Bresnick AR. Mts1 regulates the assembly of nonmuscle myosin-IIA. Biochemistry. 2003;42(48):14258–66.
Li Y, Lalwani AK, Mhatre AN. Alternative splice variants of MYH9. DNA Cell Biol. 2008;27(3):117–25.
Liebe S, Menzel D. Actomyosin-based motility of endoplasmic reticulum and chloroplasts in Vallisneria mesophy ll cells. Biol Cell. 1995;85(2–3):207–22.
Lin P, et al. Nonmuscle myosin IIA facilitates vesicle trafficking for MG53-mediated cell membrane repair. FASEB J. 2012;26(5):1875–83.
Liu X, et al. Filament structure as an essential factor for regulation of Dictyostelium myosin by regulatory light chain phosphorylation. Proc Natl Acad Sci USA. 1998;95(24):14124–9.
Lohmann K. Über die enzymatische ausspaltung der kreatinephosphorsaure; zugleich ein beitrag zum chemismus der muskelkontraktion. Biochem. 1934;Z(271):264–77.
Lordier L, et al. RUNX1-induced silencing of non-muscle myosin heavy chain IIB contributes to megakaryocyte polyploidization. Nat Commun. 2012;3:717.
Luck-Vielmetter D, Schleicher M, Grabatin B, Wippler J, Gerisch G. Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II. FEBS Lett. 1990;269(1):239–43.
Ma X, Kawamoto S, Uribe J, Adelstein RS. Function of the neuron-specific alternatively spliced isoforms of nonmuscle myosin II-B during mouse brain development. Mol Biol Cell. 2006;17(5):2138–49.
Ma X, et al. Ablation of nonmuscle myosin II-B and II-C reveals a role for nonmuscle myosin II in cardiac myocyte karyokinesis. Mol Biol Cell. 2010;21(22):3952–62.
Maravillas-Montero JL, Santos-Argumedo L. The myosin family: unconventional roles of actin-dependent molecular motors in immune cells. J Leukoc Biol. 2012;91(1):35–46.
Matsumura F. Regulation of myosin II during cytokinesis in higher eukaryotes. Trends Cell Biol. 2005;15(7):371–7.
Matsumura F, Hartshorne DJ. Myosin phosphatase target subunit: Many roles in cell function. Biochem Biophys Res Commun. 2008;369(1):149–56.
Matsuoka R, Yoshida MC, Furutani Y, Imamura S, Kanda N, Yanagisawa M, Masaki T, Takao A. Human smooth muscle myosin heavy chain gene mapped to chromosomal region 16q12. Am J Med Genet. 1993;46(1):61–7.
Maupin P, Phillips CL, Adelstein RS, Pollard TD. Differential localization of myosin-II isozymes in human cultured cells and blood cells. J Cell Sci. 1994;107(Pt 11):3077–90.
Mene P, Punzo G, Pirozzi N. TRP channels as therapeutic targets in kidney disease and hypertension. Curr Top Med Chem. 2013;13(3):386–97.
Middelbeek J, Clark K, Venselaar H, Huynen MA, van Leeuwen FN. The alpha-kinase family: an exceptional branch on the protein kinase tree. Cell Mol Life Sci. 2010;67(6):875–90.
Min SY, Ahn HJ, Park WS, Kim JW. Successful renal transplantation in MYH9-related disorder with severe macrothrombocytopenia: first report in Korea. Transplant Proc. 2014;46(2):654–6.
Mitsuhashi M, Sakata H, Kinjo M, Yazawa M, Takahashi M. Dynamic assembly properties of nonmuscle myosin II isoforms revealed by combination of fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy. J Biochem. 2011;149(3):253–63.
von Muhlen CA, et al. Non-muscle myosin as target antigen for human autoantibodies in patients with hepatitis C virus-associated chronic liver diseases. Clin Exp Immunol. 1995;100(1):67–74.

Murakami N, Chauhan VP, Elzinga M. Two nonmuscle myosin II heavy chain isoforms expressed in rabbit brains: filament forming properties, the effects of phosphorylation by protein kinase C and casein kinase II, and location of the phosphorylation sites. Biochemistry. 1998;37(7):1989–2003.

Muralt A, Edsall JT. Studies in the physical chemistry of muscle globulin. IV. The anisotropy of myosin and double refraction of flow. J Biol Chem. 1930;89:351–86.

Murata-Hori M, Suizu F, Iwasaki T, Kikuchi A, Hosoya H. ZIP kinase identified as a novel myosin regulatory light chain kinase in HeLa cells. FEBS Lett. 1999;451(1):81–4.

Murphy MB, Egelhoff TT. Biochemical characterization of a Dictyostelium myosin II heavy-chain phosphatase that promotes filament assembly. Eur J Biochem. 1999;264(2):582–90.

Nishikawa M, Sellers JR, Adelstein RS, Hidaka H. Protein kinase C modulates in vitro phosphorylation of the smooth muscle heavy meromyosin by myosin light chain kinase. J Biol Chem. 1984;259(14):8808–14.

O'Seaghdha CM, Parekh RS, Hwang SJ, Li M, Kottgen A, Coresh J, Yang Q, Fox CS, Kao WH. The MYH9/APOL1 region and chronic kidney disease in European-Americans. Hum Mol Genet. 2011;20(12):2450–6.

Odronitz F, Hellkamp M, Kollmar M. diArk—a resource for eukaryotic genome research. BMC Genomics. 2007;8:103.

Onishi H, Wakabayashi T. Electron microscopic studies of myosin molecules from chicken gizzard muscle I: the formation of the intramolecular loop in the myosin tail. J Biochem. 1982;92(3):871–9.

Pasternak C, Flicker PF, Ravid S, Spudich JA. Intermolecular versus intramolecular interactions of Dictyostelium myosin: possible regulation by heavy chain phosphorylation. J Cell Biol. 1989;109(1):203–10.

Pastra-Landis SC, Lowey S. Myosin subunit interactions. Properties of the 19,000-dalton light chain-deficient myosin. J Biol Chem. 1986;261(31):14811–6.

Paves H, Truve E. Myosin inhibitors block accumulation movement of chloroplasts in Arabidopsis thaliana leaf cells. Protoplasma. 2007;230(3–4):165–9.

Pecci A, et al. Position of nonmuscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease. Hum Mutat. 2008;29(3):409–17.

Pecci A, et al. MYH9 related disease: four novel mutations of the tail domain of myosin-9 correlating with a mild clinical phenotype. Eur J Haematol. 2010;84(4):291–7.

Pecci A, et al. MYH9-related disease: a novel prognostic model to predict the clinical evolution of the disease based on genotype-phenotype correlations. Hum Mutat. 2014;35(2):236–47.

Pennisi E. Profile: Ewan Birney. Genomics' big talker. Science. 2012;337(6099):1167–9.

Plant PJ, Fawcett JP, Lin DC, Holdorf AD, Binns K, Kulkarni S, Pawson T. A polarity complex of mPar-6 and atypical PKC binds, phosphorylates and regulates mammalian Lgl. Nat Cell Biol. 2003;5(4):301–8.

Pohl M, et al. Diagnosis of Alport syndrome—search for proteomic biomarkers in body fluids. Pediatr Nephrol. 2013;28(11):2117–23.

Pollard TD, Korn ED. Acanthamoeba myosin. II. Interaction with actin and with a new cofactor protein required for actin activation of Mg^{2+} adenosine triphosphatase activity. J Biol Chem. 1973;248(13):4691–7.

Rai V, Egelhoff TT. Role of B regulatory subunits of protein phosphatase type 2A in myosin II assembly control in Dictyostelium discoideum. Eukaryot Cell. 2011;10(4):604–10.

Ramsey IS, Delling M, Clapham DE. An introduction to TRP channels. Annu Rev Physiol. 2006;68:619–47.

Rey M, Valenzuela-Fernandez A, Urzainqui A, Yanez-Mo M, Perez-Martinez M, Penela P, Mayor Jr F, Sanchez-Madrid F. Myosin IIA is involved in the endocytosis of CXCR4 induced by SDF-1alpha. J Cell Sci. 2007;120(Pt 6):1126–33.

Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature. 2005;436(7054):1113–8.

Rinaldi F, et al. Aberrant splicing and expression of the non muscle myosin heavy-chain gene MYH14 in DM1 muscle tissues. Neurobiol Dis. 2012;45(1):264–71.

Rochlin MW, Itoh K, Adelstein RS, Bridgman PC. Localization of myosin II A and B isoforms in cultured neurons. J Cell Sci. 1995;108(Pt 12):3661–70.
Rodgers BD. Insulin-like growth factor-I downregulates embryonic myosin heavy chain (eMyHC) in myoblast nuclei. Growth Horm IGF Res. 2005;15(6):377–83.
Ronen D, Ravid S. Myosin II tailpiece determines its paracrystal structure, filament assembly properties, and cellular localization. J Biol Chem. 2009;284(37):24948–57.
Rosenberg MM, et al. High resolution characterization of myosin IIC protein tailpiece and its effect on filament assembly. J Biol Chem. 2013;288(14):9779–89.
Rottbauer W, et al. Cardiac myosin light chain-2: a novel essential component of thick-myofilament assembly and contractility of the heart. Circ Res. 2006;99(3):323–31.
Runnels LW. TRPM6 and TRPM7: a Mul-TRP-PLIK-cation of channel functions. Curr Pharm Biotechnol. 2011;12(1):42–53.
Ryazanov AG, Pavur KS, Dorovkov MV. Alpha-kinases: a new class of protein kinases with a novel catalytic domain. Curr Biol. 1999;9(2):R43–5.
Sanders LC, Matsumura F, Bokoch GM, de Lanerolle P. Inhibition of myosin light chain kinase by p21-activated kinase. Science. 1999;283(5410):2083–5.
Sandquist JC, Swenson KI, Demali KA, Burridge K, Means AR. Rho kinase differentially regulates phosphorylation of nonmuscle myosin II isoforms A and B during cell rounding and migration. J Biol Chem. 2006;281(47):35873–83.
Saposnik B, et al. Mutation spectrum and genotype-phenotype correlations in a large French cohort of MYH9-Related Disorders. Mol Genet Genomic Med. 2014;2(4):297–312.
Scheeff ED, Bourne PE. Structural evolution of the protein kinase-like superfamily. PLoS Comput Biol. 2005;1(5):e49.
Schlingmann KP, et al. Hypomagnesemia with secondary hypocalcemia is caused by mutations in TRPM6, a new member of the TRPM gene family. Nat Genet. 2002;31(2):166–70.
Schramek D, et al. Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. Science. 2014;343(6168):309–13.
Sebe-Pedros A, Grau-Bove X, Richards TA, Ruiz-Trillo I. Evolution and classification of myosins, a paneukaryotic whole-genome approach. Genome Biol Evol. 2014;6(2):290–305.
Sellers JR, Eisenberg E, Adelstein RS. The binding of smooth muscle heavy meromyosin to actin in the presence of ATP. Effect of phosphorylation. J Biol Chem. 1982;257(23):13880–3.
Serwe M, Meyer HE, Craig AG, Carlhoff D, D'Haese J. Complete amino acid sequence of the regulatory light chain of obliquely striated muscle myosin from earthworm, Lumbricus terrestris. Eur J Biochem. 1993;211(1–2):341–6.
Sharma-Walia N, et al. Kaposi's sarcoma associated herpes virus (KSHV) induced COX-2: a key factor in latency, inflammation, angiogenesis, cell survival and invasion. PLoS Pathog. 2010;6(2):e1000777.
Shin DH, Chun YS, Lee DS, Huang LE, Park JW. Bortezomib inhibits tumor adaptation to hypoxia by stimulating the FIH-mediated repression of hypoxia-inducible factor-1. Blood. 2008; 111(6):3131–6.
Shutova M, Yang C, Vasiliev JM, Svitkina T. Functions of nonmuscle myosin II in assembly of the cellular contractile system. PLoS One. 2012;7(7):e40814.
Simons M, et al. Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes. Circ Res. 1991;69(2):530–9.
Slonska A, Polowy R, Golke A, Cymerys J. Role of cytoskeletal motor proteins in viral infection. Postepy Hig Med Dosw (Online). 2012;66:810–7.
Smith AC, et al. Interstitial deletion of (17)(p11.2p11.2) in nine patients. Am J Med Genet. 1986;24(3):393–414.
Solinet S, Vitale ML. Isoform B of myosin II heavy chain mediates actomyosin contractility during TNFalpha-induced apoptosis. J Cell Sci. 2008;121(Pt 10):1681–92.
Somlyo AP, Somlyo AV. Ca2+ sensitivity of smooth muscle and nonmuscle myosin II: modulated by G proteins, kinases, and myosin phosphatase. Physiol Rev. 2003;83(4):1325–58.
Stedman HH, et al. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature. 2004;428(6981):415–8.

Strand D, et al. The Drosophila lethal(2)giant larvae tumor suppressor protein forms homo-oligomers and is associated with nonmuscle myosin II heavy chain. J Cell Biol. 1994;127(5):1361–73.

Straub FB. Actin. Stud Inst Med Chem Univ Szeged. 1942;II:3–15.

Su LT, et al. TRPM7 activates m-calpain by stress-dependent stimulation of p38 MAPK and c-Jun N-terminal kinase. J Mol Biol. 2010;396(4):858–69.

Szent-Györgyi A. The crystallization of myosin and some of its properties and reactions. Stud Inst Med Chem Univ Szeged. 1943;III:76–85.

Takizawa N, Ikebe R, Ikebe M, Luna EJ. Supervillin slows cell spreading by facilitating myosin II activation at the cell periphery. J Cell Sci. 2007;120(Pt 21):3792–803.

Tan JL, Spudich JA. Dictyostelium myosin light chain kinase. Purification and characterization. J Biol Chem. 1990;265(23):13818–24.

Tan I, Yong J, Dong JM, Lim L, Leung T. A tripartite complex containing MRCK modulates lamellar actomyosin retrograde flow. Cell. 2008;135(1):123–36.

Tang HW, et al. Atg1-mediated myosin II activation regulates autophagosome formation during starvation-induced autophagy. EMBO J. 2011;30(4):636–51.

Toothaker LE, et al. Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, characterization of the protein, chromosomal localization, and upregulation during myeloid differentiation. Blood. 1991;78(7):1826–33.

Totsukawa G, et al. Distinct roles of MLCK and ROCK in the regulation of membrane protrusions and focal adhesion dynamics during cell migration of fibroblasts. J Cell Biol. 2004;164(3):427–39.

Trotter JA. Living macrophages phosphorylate the 20,000 Dalton light chains and heavy chains of myosin. Biochem Biophys Res Commun. 1982;106(3):1071–7.

Trotter JA, Nixon CS, Johnson MA. The heavy chain of macrophage myosin is phosphorylated at the tip of the tail. J Biol Chem. 1985;260(26):14374–8.

Trybus KM, Lowey S. Conformational states of smooth muscle myosin. Effects of light chain phosphorylation and ionic strength. J Biol Chem. 1984;259(13):8564–71.

Trybus KM, Huiatt TW, Lowey S. A bent monomeric conformation of myosin from smooth muscle. Proc Natl Acad Sci USA. 1982;79(20):6151–5.

Tullio AN, et al. Nonmuscle myosin II-B is required for normal development of the mouse heart. Proc Natl Acad Sci USA. 1997;94(23):12407–12.

Tullio AN, et al. Structural abnormalities develop in the brain after ablation of the gene encoding nonmuscle myosin II-B heavy chain. J Comp Neurol. 2001;433(1):62–74.

Ubukawa K, et al. Enucleation of human erythroblasts involves non-muscle myosin IIB. Blood. 2012;119(4):1036–44.

Underwood J, Greene J, Steimle PA. Identification of a new mechanism for targeting myosin II heavy chain phosphorylation by Dictyostelium myosin heavy chain kinase B. BMC Res Notes. 2010;3:56.

Uyeda TQ, Abramson PD, Spudich JA. The neck region of the myosin motor domain acts as a lever arm to generate movement. Proc Natl Acad Sci USA. 1996;93(9):4459–64.

Vaillancourt JP, Lyons C, Cote GP. Identification of two phosphorylated threonines in the tail region of Dictyostelium myosin II. J Biol Chem. 1988;263(21):10082–7.

Vale RD. The molecular motor toolbox for intracellular transport. Cell. 2003;112(4):467–80.

Varlamova O, Spektor A, Bresnick AR. Protein kinase C mediates phosphorylation of the regulatory light chain of myosin-II during mitosis. J Muscle Res Cell Motil. 2001;22(3):243–50.

Vasioukhin V. Lethal giant puzzle of Lgl. Dev Neurosci. 2006;28(1–2):13–24.

Walder RY, et al. Mutation of TRPM6 causes familial hypomagnesemia with secondary hypocalcemia. Nat Genet. 2002;31(2):171–4.

Wang Z, Pesacreta TC. A subclass of myosin XI is associated with mitochondria, plastids, and the molecular chaperone subunit TCP-1alpha in maize. Cell Motil Cytoskeleton. 2004;57(4):218–32.

Wang L, et al. A comprehensive genetic study on left atrium size in Caribbean Hispanics identifies potential candidate genes in 17p10. Circ Cardiovasc Genet. 2010a;3(4):386–92.

Wang L, et al. Mutations in myosin light chain kinase cause familial aortic dissections. Am J Hum Genet. 2010b;87(5):701–7.
Wasylnka JA, Bakowski MA, Szeto J, Ohlson MB, Trimble WS, Miller SI, Brumell JH. Role for myosin II in regulating positioning of Salmonella-containing vacuoles and intracellular replication. Infect Immun. 2008;76(6):2722–35.
Weir L, Chen D. Characterization of the nonmuscle myosin heavy chain IIB promoter: regulation by E2F. Gene Expr. 1996;6(1):45–57.
Wendt T, Taylor D, Trybus KM, Taylor K. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. Proc Natl Acad Sci USA. 2001;98(8):4361–6.
Xia D, Stull JT, Kamm KE. Myosin phosphatase targeting subunit 1 affects cell migration by regulating myosin phosphorylation and actin assembly. Exp Cell Res. 2005;304(2):506–17.
Xia ZK, et al. Nonmuscle myosin IIA is associated with poor prognosis of esophageal squamous cancer. Dis Esophagus. 2012;25(5):427–36.
Xu XS, Kuspa A, Fuller D, Loomis WF, Knecht DA. Cell-cell adhesion prevents mutant cells lacking myosin II from penetrating aggregation streams of Dictyostelium. Dev Biol. 1996;175(2):218–26.
Xu J, et al. Nonmuscle myosin light-chain kinase mediates neutrophil transmigration in sepsis-induced lung inflammation by activating beta2 integrins. Nat Immunol. 2008;9(8):880–6.
Yang F, Wei Q, Adelstein RS, Wang PJ. Non-muscle myosin IIB is essential for cytokinesis during male meiotic cell divisions. Dev Biol. 2012;369(2):356–61.
Yumura S, et al. Multiple myosin II heavy chain kinases: roles in filament assembly control and proper cytokinesis in Dictyostelium. Mol Biol Cell. 2005;16(9):4256–66.
Zamudio-Meza H, Castillo-Alvarez A, Gonzalez-Bonilla C, Meza I. Cross-talk between Rac1 and Cdc42 GTPases regulates formation of filopodia required for dengue virus type-2 entry into HMEC-1 cells. J Gen Virol. 2009;90(Pt 12):2902–11.
Zheng J, et al. Prestin is the motor protein of cochlear outer hair cells. Nature. 2000;405(6783):149–55.

Dr. Venkaiah Betapudi Dr. Venkaiah Betapudi is an Assistant Professor of Molecular Medicine, Cleveland Clinic, Ohio, USA. His major research interests are cancer and endothelial cell biology, angiogenesis, mouse tumor models, lentivirus-mediated gene expression, inflammation, mechanism of endothelial cell activation, autoantibodies, microparticle secretion, myosin II roles and regulation, cell migration, wound healing, and metastasis, nanoparticles, and bleeding disorders.

Bioinformatics Databases: Implications in Human Health

6

Leena Rawal, Deepak Panwar, and Sher Ali

6.1 Introduction

Understanding genetic variations in human genome and its contribution to phenotypic change is one of the paramount goals in biology and medicine. The completion of the human genome sequencing has made information readily available in large public domains, thereby allowing the researchers to identify and characterize naturally occurring variations in the human DNA sequence across individuals. Genome-wide association studies (GWAS) are one of the most widely used analyses for investigating variants. The advent of progress in genome sequencing technologies such as high-throughput genotyping, next-generation sequencing, RNA expression, exome sequencing, and massively parallel sequencing has indeed accelerated the exploration of genetic variations involved in human diseases.

Advancements in these technologies have made it possible to record both moderate and proficient variations in the genomic architecture and to assess the events of transcripts in disease or control populations. The analysis of the data is usually

L. Rawal (✉)
Department of Cytogenetics, National Reference Laboratory, Dr Lal PathLabs Ltd, Block E, Sector 18, Rohini, New Delhi 110085, India

Molecular Genetics Division, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India
e-mail: leena.rawal7@gmail.com

D. Panwar
Molecular Genetics Division, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

S. Ali
Molecular Genetics Division, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_6

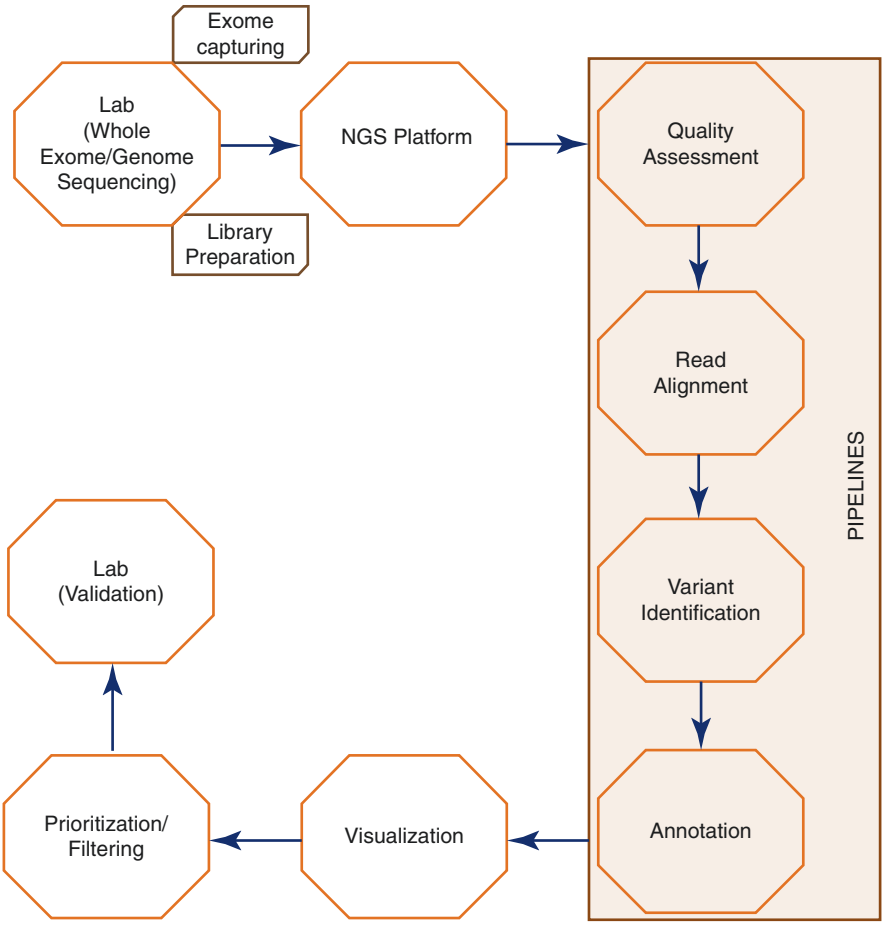

Fig. 6.1 A roadmap for whole-exome and whole-genome sequencing projects. Post library preparation, samples are sequenced on a certain platform. The following steps include quality assessment, read alignment against a reference genome, and variant identification. The annotation of the identified mutations is done to infer the biological relevance, and results can be displayed using dedicated tools. The detected mutations can be prioritized and filtered, followed by validation of the generated results in the lab (Adapted from Pabinger et al. 2014)

decomposed into five distinct steps: (1) quality assessment of the raw data, (2) read alignment to a reference genome, (3) variant identification, (4) annotation of the variants, and (5) data visualization (Pabinger et al. 2014) (Fig. 6.1).

As the genome evolves, the newly found genes with novel characteristics create phenotypic and genetic diversity in species. To ascertain the functions of the new genes, integrations of gene–gene interaction (GGI) networks across their homologues and ancestral genes are used to acquire their corresponding biological roles. Direct gene–gene and protein–protein interactions (PPIs) are one of the strongest

manifestations for the functional relation between genes and their interacting partners. In-depth understanding of the complete network of protein–protein interactions, i.e., the number, type, and distribution including the occurrence of key nodes in these networks, would open new avenues into the structures and properties of biological systems. Thus, bioinformatics and advanced computational tools/software have become important for the analysis of large amount of data excavated from such interactions. The purpose of this chapter is to make the readers aware of the bioinformatic tools and related softwares that indentify interactions within and between the genome (G), transcriptome (T), and phenome (P) that eventually have an impact on human health.

6.2 Essence of Bioinformatics

In the early 1990s, two new technological developments, high-throughput DNA sequencing and the Internet, allowed for an overwhelming explosion of biological data and its global dissemination. As a result of the former, whole-genome sequencing was made feasible. In quick succession, the genomes of bacteria, *Haemophilus influenza* (Fleischmann et al. 1995) and *Mycoplasma genitalium* (Fraser et al. 1995), in 1995; an archaeon, *Methanococcus jannaschii* (Bult et al. 1996), and a yeast, *Saccharomyces cerevisiae* (Goffeau et al. 1996), in 1996; a nematode, *Caenorhabditis elegans* (C. elegans Sequencing Consortium 1998), in 1998; the fruit fly, *Drosophila melanogaster* (Adams et al. 2000), in 2000; and finally a human, *Homo sapiens* (International Human Genome Sequencing Consortium 2004; Lander et al. 2001; Venter et al. 2001), in 2001 were sequenced.

Since then thousands of genomes have been sequenced; the central databases such as GenBank at National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), DNA database of Japan (DDBJ; http://www.ddbj.nig.ac.jp/), and European Molecular Biology Laboratory (EMBL; http://www.ebi.ac.uk/embl/) are simultaneously updated. Based on the sequence and corresponding organism information, data were subjected to the specific sequence data repositories from where information on human genome can be retrieved and stored into the Genome browser (http://genome.ucsc.edu), Ensembl (http://www.ensembl.org), and Golden Path server (http://genome.ucsc.edu/). Additionally, the information on novel genes can be salvaged from other resources such as UniGene and RefSeq (accessible at NCBI) (Mount and Pandey 2005). Thus, as the sequencing technology blossomed, so did the field of bioinformatics.

Bioinformatics, often cited as computational biology, is the application of computer science, statistics, applied mathematics, and information technology to the study of biology and biological problems. Its interdisciplinary approach provides unique solutions to extract novel biological information from sequenced data analysis. The ability to process large data has lent bioinformatics use in different biological fields, including comparative sequence analysis, genomics, biological literature analysis, macromolecular sequence analysis, metagenomics, phylogenetic studies,

sequence motif analysis, and transcriptional regulation. Therefore, currently with the wide availability of sequence data, we can gain virtually impossible insights into evolution of genome, protein world, domain–domain cross talk, and interacting network.

6.3 Genetic Variations and Databases

With time and ever-changing environmental factors, the panorama of human genetics has evolved to target complex diseases. Although variants in the genome take many forms, majority of them ascend mainly from two types of mutation events. The most common variant type occurred from a single-base mutation, also termed as single-nucleotide polymorphisms (SNPs). Experiential studies identified the common SNPs (>20% minor allele frequency) at 0.3–1 kb average intervals at chromosomes between any two individuals, which scales up to 5–10 million SNPs across the genome (Altshuler et al. 2000). It has been estimated that 50,000–200,000 SNPs may be biologically important (Chanock 2001; de Bakker et al. 2005). SNPs occurring in the exons of genes that do not alter protein primary structure are called "synonymous". SNPs in introns, regulatory, and gene-distant regions can also be functionally important, primarily by affecting gene regulation. A relatively common variant (MAF of 1–2%), G21210A, in the 3′ UTR of the prothrombin gene, *F2*, increases its expression, and carriers of the minor allele are at significantly increased risk for venous thrombosis (Poort et al. 1996).

The other variant type occurs as the result of deletion or insertion of a nucleotides stretch, so-called insertion/deletion (INDEL) polymorphisms. The most common insertion/deletion events occur in repetitive sequence elements, namely, variable number tandem repeat (VNTR) and microsatellites. Nucleotide substitutions in the genome have the potential to directly contribute to disease pathogenesis, depending upon their occurrence. Large expansions of trinucleotide repeats can lead to genomic instability, the classic example being fragile X syndrome. A dinucleotide repeat (DG8S737) on chromosome 8 has shown to be strongly associated with prostate cancer in African-Americans (Cheng et al. 2008; Freedman et al. 2006), though its functional importance is yet to be established. Although modest variations in STR and VNTR length impact on disease remain to be determined, evidence suggests that some may act as binding sites for nuclear proteins (Richards et al. 1993). Nucleotide substitutions in the protein-coding portions of genes sometimes result in the premature insertions of codons causing the termination of protein translation. These often become alleles that are effectively null as their transcribed mRNA is rapidly degraded by nonsense-mediated decay (Lykke-Andersen 2001). Additionally, gross chromosomal aberrations, for instance, deletions, inversions, or translocations of large segments of DNA, have been associated with several clinically characterized genomic syndromes. Few polymorphisms show direct impact by creating deleterious phenotypes. However, large deletions or duplications can be quantified only through intensive cytogenetic methods (Gratacos et al. 2001).

In bioinformatics research, databases are primary resources for researchers to retrieve the sequence data, which may collaboratively pool the data from other

Table 6.1 Databases for mining different human genetic variations

Databases	Web links
Mutation databases	
OMIM	http://www.ncbi.nlm.nih.gov/Omim/
HGMD	http://www.hgmd.org
GDB mutation way station	http://www.centralmutations.org/
HUGO mutation database initiative	http://www.genomic.unimelb.edu.au/mdi/
Central databases (SNPs and mutations)	
HGV base	http://hgvbase.cgb.ki.se/
Sequence variation database (SRS)	http://srs.ebi.ac.uk/
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/
The SNP consortium (TSC)	http://snp.cshl.org/
Genetic marker maps (microsatellites, STSs other markers)	
Marshfield maps	http://research.marshfieldclinic.org/genetics/
Genome database (GDB)	http://www.gdb.org
dbSTS	http://www.ncbi.nlm.nih.gov/STS/
UniSTS	http://www.ncbi.nlm.nih.gov/genome/sts/
Somatic and nonnuclear mutation databases	
MitoMap	http://www.gen.emory.edu/mitomap.html
Mitelman Map	http://cgap.nci.nih.gov/Chromosomes/Mitelman
Gene-orientated SNP and mutation visualization	
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/
PicSNP	http://picsnp.org
Protein mutation database	http://www.genome.ad.jp/htbin/ , http://pmd.ddbj.nig.ac.jp/~pmd/
Go!Poly	http://61.139.84.5/gopoly/
GeneLynx	http://www.genelynx.org
SNPper	http://bio.chip.org:8080/bio/snpper-enter
GeneSNPs	http://www.genome.utah.edu/genesnps/
GAP SNP database	http://lpgws.nci.nih.gov/

repositories into single available databank. The significant databases to mine the human genomic variations are enlisted in Table 6.1. These repositories are large reservoirs of information encompassing the genetic variation-causing diseases in the human genome. Concomitant analysis of the information retrieved through these databases could lead to a better understanding of disease phenotypes/genotype and their relationship with mutations.

6.4 Genome-Wide Association Studies (GWAS)

In 2007, the modern complex genetic era of GWAS began, and since then it has been tremendously successful in implicating several novel disease associations in many common complex diseases (Wellcome Trust Case Control Consortium 2007).

The completion of the human genome (Lander et al. 2001) and HapMap projects (The International HapMap Consortium 2003) coupled with advances in high-throughput genotyping and innovative sequencing technology has resulted in an explosion of genome-wide association studies (GWAS). According to the NIH (National Institute of Health, USA), GWAS have allowed researchers to study the genetic variations across the human genome and their association between different genetic markers, phenotypes, and different disease conditions.

In accordance with the common disease-common variant (CD-CV) hypothesis, GWAS is based on the fact that common variants (SNPs) within alleles in the common population will decipher much of the heritability of common diseases (Reich and Lander 2001; Schork et al. 2009). Validation and analyses of the CD-CV hypothesis provide an insight into the genetic makeup of common diseases, e.g., rheumatoid arthritis, type 2 diabetes, or hypertension, that may have been contributed from multiple alleles. If common variants show a minor effect but common diseases have strong inheritance in families, then disease must be inclined by multiple genetic factors. However, the frequency of each allele can vary between groups of individuals, and here is a possibility that the common variations can increase an individual's susceptibility to a disease. GWAS is a pioneering study that took advantage of favorable genetic and technological advances to uncover several novel disease associations with SNPs. As of November 16, 2013, GWAS catalog maintained by the National Human Genome Research Institute (NHGRI) documented 11,907 SNPs and 940 traits with 15,052 disease associations. Thus, identification of many novel genes has uncovered pathways that are implicated with disease etiology, leading to deeper insights on disease mechanisms. To help the research community in finding relevant publications and to further explore the reported variants, NHGRI has established and maintained the NHGRI GWAS Catalog (http://www.genome.gov/26525384), an online, regularly updated database of single-nucleotide polymorphism (SNP)-trait associations from GWAS. A bioinformatics tool, GWAS Integrator, offers a robust search capacity and a set of data mining functions by integrating information from the NHGRI GWAS Catalog, with data from other established bioinformatics resources including HapMap (http://hapmap.ncbi.nlm.nih.gov/), the Human Genome Epidemiology (HuGE) Navigator (http://www.hugenavigator.net/), SNP Annotation and Proxy Search (SNAP) (http://www.broadinstitute.org/mpg/snap/ldsearch.php), and University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

Among many achievements of GWAS, the most significant ones are the usage of data mining software to model complex genotype–phenotype relationships and expand the networking among biologists and bioinformaticians and enable the biological database to pave the way for genetic association studies. A roadmap suggesting the development of information analysis of GWAS has been shown in Fig. 6.2. Although GWAS contributed to many SNPs and variant discoveries, genetic basis of most common diseases still remains unexplored. One possible

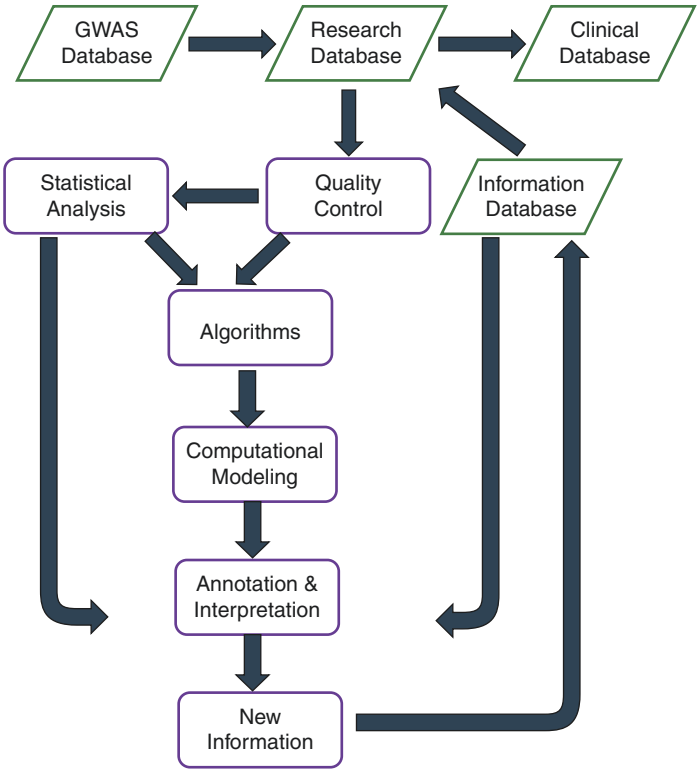

Fig. 6.2 Bioinformatics analyses of GWAS data. The use of filter and wrapper algorithms along with computational modeling approaches is recommended in addition to parametric statistical methods. Biological knowledge in public databases has a very important role to play at all levels of the analysis and interpretation (Adapted from Moore et al. 2010)

clarification for this "missing hereditability" is that old GWAS methods have focused on one SNP at a time and failed to detect the heterogenetic complexity of many genotype–phenotype relationships and gene–gene and gene–environment interactions. GWAS has been used in many complex disease association, for instance, Crohn's disease (Duerr et al. 2006; Libioulle et al. 2007), T1D and rheumatoid arthritis (RA) (Wellcome Trust Case Control Consortium 2007), and multiple sclerosis (MS) (De Jager et al. 2009). An overview of some GWAS-specific tools used for detection of nucleotide variant identification/annotation and data visualization is given in Table 6.2. Thus, advancements in GWAS will aid in identification of new genetic associations that will eventually help the researchers to use the information to develop better strategies to detect, treat, and prevent the diseases.

Table 6.2 Tools for nucleotide variant identification and annotation

Tool	Descriptive feature	Web link
Tools for variant identification		
Germline and somatic variations		
CRISP	Comprehensive Read analysis for Identification of Single-Nucleotide Polymorphisms (SNPs) from Pooled sequencing (CRISP) that is able to identify both rare and common variants by using two approaches: comparing the distribution of allele counts across multiple pools using contingency tables and evaluating the probability of observing multiple non-reference base calls due to sequencing errors alone	http://polymorphism.scripps.edu/%E2%88%BCvbansal/software/CRISP/
Dindel	A Bayesian approach for calling small (<50 nucleotides) insertions and deletions from short read data	
GATK	Genome Analysis Toolkit (GATK), a structured programming framework designed to ease the development of efficient and robust analysis tools for next-generation DNA sequencers using the functional programming philosophy of MapReduce	http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit
SAM tools	The Sequence Alignment/Map (SAM) format is a generic alignment format for storing read alignments against reference sequences, supporting short and long reads (up to 128 Mbp) produced by different sequencing platforms. It is flexible in style, compact in size, and efficient in random access	http://samtools.sourceforge.net
SNVer	SNVer (single-nucleotide variant caller/seeker), a statistical tool for detecting variants in analysis of NGS data, employing a binomial–binomial model to test the significance of observed allele frequency against sequencing error	http://snver.sourceforge.net/
VarScan	Detection of somatic mutations and copy number alterations (CNAs) in exome data	http://varscan.sourceforge.net
CVN identification		
CNVnator	Tool for identifying, genotyping, and characterizing CNVs	http://sv.gersteinlab.org/
CONTRA	*Co*py *n*umber *t*argeted *r*esequencing *a*nalysis (CONTRA), for targeted resequencing data such as those from whole-exome capture data	http://contra-cnv.sourceforge.net/
ExomeCNV	ExomeCNV is an R package tailored to detection of copy number variants (CNV) and loss of heterozygosity (LOH) from exome sequencing data. It exploits the unique discrete feature of exon definitions and incredible cross sample consistency of depth of coverage. ExomeCNV is most suitable when paired samples (e.g., tumor–normal pair) are available	

Table 6.2 (continued)

Tool	Descriptive feature	Web link
RDXplorer	Tool for copy number variant (CNV) detection in whole human genome sequence data using read depth (RD) coverage	http://rdxplorer.sourceforge.net/
Sequence variation identification		
BreakDancer	Tool can detect deletions, insertions, inversions, and intra- and inter-chromosomal translocations and computes the copy number	
BreakPointer	A fast algorithm to locate breakpoints of structural variants (SVs) from single-end reads produced by next-generation sequencing	https://github.com/ruping/Breakpointer
CLEVER	Authors provide structured documentation	http://clever-sv.googlecode.com
GASVPro	Software to detect SVs from paired-end mapping data	http://compbio.cs.brown.edu/projects/gasv/
SVMerge	SVMerge integrates calls from several existing SV callers: BreakDancerMax, Pindel, RDXplorer, CnD, and SECluster	http://svmerge.sourceforge.net/
Tools for variant annotation		
ANNOVAR	Integrated tool providing gene annotation at single-nucleotide variants (SNVs) and insertions/deletions, such as examining their functional consequence on genes, inferring cytogenetic bands, reporting functional importance scores, and finding variants in conserved regions	http://www.openbioinformatics.org/annovar/
AnnTools	Provides a set of helper tools for custom annotation	http://anntools.sourceforge.net/
NGS-SNP	Provides rich annotations for SNPs identified by the sequencing of whole genomes from any organism with reference sequences in Ensembl	http://stothard.afns.ualberta.ca/downloads/NGS-SNP/
SeattleSeq Annotation	The SeattleSeq Annotation server provides annotation of single-nucleotide variants (SNVs) and small indels, both known and novel. This annotation includes dbSNP rs ID, gene names and accession numbers, variation functions (e.g., missense), protein positions and amino acid changes, conservation scores, HapMap frequencies, PolyPhen predictions, and clinical association	http://snp.gs.washington.edu/SeattleSeqAnnotation138/
SNPeffect 4.0	SNPeffect primarily focuses on the molecular characterization and annotation of disease and polymorphism variants in the human proteome	http://snpeffect.switchlab.org/
VARIANT	VARIant ANalysis Tool (VARIANT) reports information on the variants found that include consequence type and annotations taken from different databases and repositories	http://variant.bioinfo.cipf.es

6.5 Gene–Gene Interactions

Complex diseases have adversely affected human health throughout the world. Several studies have reported that the complex diseases are caused by multiple loci (Kraft and Cox 2008; Seng and Seng 2008). Following the identification of disease-susceptibility polymorphisms by GWAS by using the standard single-locus analyses, the process is shifting towards the detection of gene interactions with another gene or other cellular factors.

The gene–environment interactions (G × E) are likely to affect complex phenotypes. The individuals with predisposing genetics are more likely to develop a disease when exposed to an altered environment than individuals exposed to the same environment, without predisposing genetics (Cambien et al. 1992; Jacques et al. 1996). In addition to interactions among genes and environment, interactions among different genetic loci (G × G) can also influence disease risk. The G × G are defined as epistatic when the allelic variations of one gene alter the effect of variations of another gene (Musani et al. 2007). Epistasis has been identified in human diseases and its role in public health has been highlighted (Small et al. 2002; Howard et al. 2002). On the other hand, if a genetic factor functions primarily through a complex mechanism that involves multiple genes and environmental factors, the effect might be missed when the gene is examined in isolation without allowing for its potential interactions with other unknown factors. Therefore, it is important to explore the gene–gene and/or gene–environment interactions in order to recognize the gene etiology of complex diseases.

A number of methods and several computational tools have been developed for the gene prioritization based on sequence-based features, gene-expression data, and functional annotation (Table 6.3). Also, many other approaches have been anticipated in the identifications of gene–gene and/or gene–environment interactions. Much of the known information on gene–phenotype association is distributed in various databases that explore variant-filtering strategies (Table 6.4). The outcomes from these analyses could lead to new genetic findings that account for the heritability of human diseases as well as provide novel insights about underlying genetic etiology through bench science research and clinical applications.

Table 6.3 Web resources and algorithms for gene–gene interaction and gene prioritization

Tools	Description	Web link
CAESAR	CAndidatE Search And Rank (CAESAR) is a tool for prioritizing candidate genes for complex traits. CAESAR exploits the knowledge of complex traits in literature by using ontologies to semantically map the trait information to gene and protein centric information from several different public data sources	http://visionlab.bio.unc.edu/caesar/
CANDID	CANDID is a genome-wide candidate identification and prioritization algorithm that uses a several heterogeneous data sources, some of them chosen to overcome bias due to previous knowledge of the user or against poorly characterized genes	https://dsgweb.wustl.edu/hutz/candid.html

Table 6.3 (continued)

Tools	Description	Web link
DiseaseNet	DiseaseNet is a platform for analyzing disease-associated protein–protein interaction (PPI) networks. This tool can be used to obtain disease-associated gene information and the cross talk with other diseases through PPI networks. Disease–gene network is defined as the network that is constructed from the known interactions between the genes implicated in different diseases	http://bioschool.iitd.ac.in/DiseaseNet/
ENDEAVOR	ENDEAVOUR is a web resource for the prioritization of candidate genes that uses a training set of genes known to be involved in a biological process of interest. The approach consists of (1) inferring several models (based on various genomic data sources), (2) applying each model to the candidate genes to rank those candidates against the profile of the known genes, and (3) merging the several rankings into a global ranking of the candidate genes	http://www.esat.kuleuven.be/endeavour
G2D	G2D (genes to diseases) is a web resource for prioritizing candidates genes for inherited diseases. It uses three algorithms based on different prioritization strategies. Candidate genes are prioritized according to their possible relation to an inherited disease using a combination of data mining on biomedical databases and gene sequence analysis. The input to the server is the genomic region where the user is looking for the disease-causing mutation, plus an additional piece of information depending on the algorithm used	http://www.ogic.ca/projects/g2d_2/
GeneDistiller	GeneDistiller provides knowledge-driven, fully interactive, and intuitive access to multiple data sources. It uses information from various data sources such as gene–phenotype associations, gene-expression patterns, and protein–protein interactions	http://www.genedistiller.org/
Gene Prospector	Gene Prospector is a web-based application that selects and prioritizes potential disease-related genes by using a highly curated and updated literature database of genetic association studies. Gene Prospector provides an online gateway for searching for evidence about human genes in relation to diseases and other phenotypes	http://www.hugenavigator.net/HuGENavigator/geneProspectorStartPage.do
GeneWanderer	GeneWanderer is a candidate disease–gene prioritization algorithm based on protein–protein interaction	http://compbio.charite.de/genewanderer

(continued)

Table 6.3 (continued)

Tools	Description	Web link
Gentrepid	Public candidate disease–gene prediction system that associates genes with specified phenotypes using genetic and bimolecular data. Gentrepid draws on two gene clustering methods to make candidate gene predictions: the Common Pathway Scanning (CPS) and Common Module Profiling (CMP) approaches	http://www.gentrepid.org/
MedSim	MedSim ranks candidate genes for a particular disease based on functional comparisons involving the gene ontology. MedSim uses functional annotations of known disease genes for assessing the similarity of diseases as well as the disease relevance of candidate genes	http://funsimmat.bioinf.mpi-inf.mpg.de/qf.php
MimMiner	MimMiner is a system for text-mining analysis of the human phenome that classifies human disease phenotypes from OMIM and phenotype similarities for similar human disease phenotypes at multiple levels of gene annotations	http://www.cmbi.ru.nl/MimMiner/cgi-bin/main.pl
MORPHIN	Prioritizes the most relevant human diseases for a given set of model organism genes, potentially highlighting new model systems for human diseases and providing context to model organism studies	http://www.inetbio.org/morphin
PGMapper	PGMapper is a software tool for automatically matching phenotype to genes from a defined genome region or a group of given genes by combining gene function information from the OMIM and PubMed databases. PGMapper is currently available for candidate gene search independently for human, mouse, rat, zebrafish, and 12 other species	http://www.genediscovery.org/pgmapper/index.jsp
PhenoPred	PhenoPred is an algorithm for detecting gene–disease associations based on a protein–protein interaction network, known gene–disease associations, protein sequences, and protein functional information at the molecular level. PhenoPred is supervised meaning that first each protein is mapped onto the spaces of disease and functional terms. In a second step, a support vector machine model is trained and used to detect gene–disease associations	http://www.phenopred.org
PRINCE	PRIoritizatioN and Complex Elucidation (PRINCE) is a network-based approach for predicting causal genes and protein complexes that are involved in a disease of interest. PRINCE generalizes the standard network-based approaches by both considering the network signal in a global manner and going beyond single genes to the modules that are affected in a given disease.	

Table 6.3 (continued)

Tools	Description	Web link
SNPs3D	SNPs3D is a web resource, coupled to a database that provides and integrates as much information as possible on disease–gene relationships at the molecular level. The SNPs3D resource has three primary modules. One of them identifies which genes are promising candidates for involvement in a specified disease	http://www.SNPs3D.org
ToppGene	ToppGene is a gene prioritization method that combines mouse phenotype data with human gene annotations and literature. It ranks candidate genes based on a similarity score for each annotation of each candidate by comparing to the enriched terms in a given set of training genes	http://toppgene.cchmc.org/

Table 6.4 Data sources and gene prioritization tools depending upon the data type

Data type	Data content	Possible sources	Tools
Experiment, observation	Linkage, association, pedigree, relevant texts, and other data	User provided	CAESAR, CANDID, ENDEAVOR, G2D, Gentrepid, GeneDistiller, PGMapper, PRINCE, ToppGene
Sequence, Structure, metadata	Sequence conservation, exon number, coding region length, known structural domains and sequence motifs, chromosomal location, protein localization, and other gene-centered information and predictions	SCOP, Pfam, PROSITE, UniProt, Entrez Gene, ENSEMBL, InterPro, LocDB, GeneCards, PredictProtein	CAESAR, CANDID, ENDEAVOR, G2D,Gentrepid, GeneDistiller, Gene Prospector, MedSim, MimMiner, PGMapper, PhenoPred, SNPs3D, ToppGene
Pathway, PPI, genetic linkage, expression	Disease–gene associations, pathways and gene–gene/protein–protein interactions predictions, and gene-expression data	KEGG, STRING, Reactome, DIP, BioGRID, GEO, ArrayExpress, ReLiance	CAESAR, CANDID, DiseaseNet, ENDEAVOR, G2D, Gentrepid, GeneDistiller, GeneWanderer, MedSim, PGMapper, PhenoPred, PRINCE, SNPs3D, ToppGene
Nonhuman data	Information about related genes and phenotypes in other species	OrthoDisease, OrthoMCL, MGD, Pathbase	CAESAR, CANDID, ENDEAVOR, GeneDistiller, Gene Prospector, GeneWanderer, MedSim, SNPs3D, ToppGene

(continued)

Table 6.4 (continued)

Data type	Data content	Possible sources	Tools
Ontologies	Gene, disease, phenotype, and anatomic ontologies	GO, DO, MPO, HPO, eVOC	CAESAR, ENDEAVOR, G2D, GeneDistiller, MedSim, PhenoPred, Prioritizer, SNPs3D, ToppGene
Mutation associations and effects	Information about existing mutations, their functional and structural effects and their association with diseases, and predictions of functional or structural effects for the mutations in the gene in question	dbSNP, PMD, GAD,DMDM, SNAP, PolyDoms, SNPdbe, SNPselector, RAVEN, SNPeffect, PHD-SNP, Mutation@A Glance, PromoLign, SIFT, PolyPhen, PupaSNP finder, FASTSNP	CAESAR, CANDID, Gene Prospector, GeneWanderer, SNPs3D, SUSPECTS

6.6 Protein–Protein Interactions

In the recent era of science, the need for exploring the structure and function of a protein has been an important field in biological research. Proteomics, a large-scale analysis of proteins, includes identification, expression, and functional characterization of proteins, their interactions, and other pre-, co-, and posttranslational modifications, if any. The main objective of proteomics is to assess the biological processes and analyze the protein–protein interaction networks (Blackstock and Weir 1999). All biological processes are orchestrated and regulated by proteins, and they accomplish their functions by interacting with other proteins and forming new protein–protein interactions (PPIs). Alterations in the PPI network can result in the development of a disease phenotype.

With the advent of high-throughput proteomics and emergence of computational biology, the potential focus of research has moved from analysis of individual protein to monitoring protein–protein interactions at an organism level. Current technology for characterization and identification for proteins from a cell encompasses two-dimensional gel electrophoresis for separation and tandem mass spectrometry for identification of the protein (Dove 1999). Other experimental methods such as Co-immunoprecipitation (CO-IP), phage display, and yeast two-hybrid system are often employed to detect interactions between cellular proteins (Fields and Song 1989; Bartel and Fields 1997; Uetz et al. 2000). Unfortunately, these experimental approaches are tedious and potentially inaccurate and provide information only about interaction and nothing about its structural conformation (Enright et al. 1999). To overcome this, advanced computational algorithms in structural bioinformatics are required. Computational predictions of protein–protein interactions combine bioinformatics approach and structural biological studies, to ascertain interactions at different levels. Such methods begin with a structural representation of each of the constituent proteins (either experimentally solved structures or comparative

models) and in the process attempts are made to predict whether or not two proteins will interact.

In PPI networks, nodes represent proteins (and, by extension, genes) and interactions represent physical protein–protein interactions. This includes a range of interaction types, such as hormone–receptor interactions, kinase–substrate interactions, or the stable bond between proteins in the same complex. The human PPI network consists mostly of proteins with very few interactions and a few proteins with a high number of interactions. By comparison, in a randomly connected network, most nodes have an average number of interactions, while some have many and some have few. Although researchers can infer about the interaction by analyzing the sequence of a gene/protein (Marcotte et al. 1999; Lu et al. 2002; Valencia and Pazos 2002; Salwinski and Eisenberg 2003), the structural information is needed to explore the possible function and interaction network building among the proteins. Elucidating the structural complexes of proteins may provide interaction details that are critical for understanding the molecular processes (LoConte et al. 1999; Chakrabarti and Janin 2002; Salwinski and Eisenberg 2003 Kortemme and Baker 2004). According to the study by Aloy and Russell (2004), about 10,000 different types of PPIs are stored in Protein Data Bank (PDB), while the number is approximately 2000 for nonredundant interacting proteins. Based on these calculations, current frequency of structure determination would take 20 years for complete elucidation of PPI network (Aloy and Russell 2004). Neither the interactions have been interpreted in the context of genotype–phenotype correlation nor there is information on the resultant diseased phenotype due to faulty protein interaction. Attempts on this line would surely add much-needed dimension toward the understanding of the role of faulty proteins in disease development. The majority of available public protein–protein databases are HPRD, IntAct, MINT, STRING, and BioGRID, and many others of these have their own unique feature with a large variation in architectural design and annotation (Table 6.5). Such databases simplify the identification of biological networks and formulate hypothesis based on the protein functions and cellular mechanism taking into account the rapidly growing PPI data.

Identification of protein 3D structure is important for many areas like drug designing and protein modeling. The 3D structure of a protein determines function and can infer the possible interacting partners; therefore, advanced techniques in drug development and analysis can make wide use of computer-aided program for protein structure visualization. Some of the software and/or servers used for visualizing the 3D protein structure are enlisted in Table 6.6. The protein complexes formed by physical interaction between proteins are the main elements responsible for cellular functions within the cell. Thus, the identification of complex formation is necessary to understand the structural organization of the cell. Computational methods such as protein–protein docking (PPD) are required to study protein complexes at the structural level. Such methods begin with some structural representation of each of the constituent proteins and attempt to produce an accurate 3D model of the complete complex. Thus, PPD methods give clear understanding for the (1) nature of interactions between interacting proteins, (2) three-dimensional conformation adopted by interacting proteins, and (3) atomic strength between the interacting proteins.

Table 6.5 Available databases USEFUL for PPI

Database	Description	URL
BIND	Peer-reviewed biomolecular interaction database containing published interactions and complexes	http://bind.ca/
BioGRID	Protein and genetic interactions from major model organism species	http://www.thebiogrid.org/
COGs	Orthology data and phylogenetic profiles	http://www.ncbi.nlm.nih.gov/COG
DIP	Experimentally determined interactions between proteins	http://dip.doe-mbi.ucla.edu/
HPRD	Human protein functions, PPIs, posttranslational modifications, enzyme–substrate relationships, and disease associations	http://www.hprd.org/
IntAct	Interaction data abstracted from literature or from direct data depositions by expert curators	http://www.ebi.ac.uk/intact/
iPfam	Physical interactions between those Pfam domains that have a representative structure in the Protein Data Bank (PDB)	http://ipfam.sanger.ac.uk/
MINT	Experimentally verified PPI mined from the scientific literature by expert curators	http://mint.bio.uniroma2.it/mint/
Predictome	Experimentally derived and computationally predicted functional linkages	http://visant.bu.edu/
ProLinks	The ProLinks database is a collection of inferences of functional linkages between proteins using four methods. These methods include the phylogenetic profile method which uses the presence and absence of proteins across multiple genomes to detect functional linkages; the gene cluster method, which uses genome proximity to predict functional linkage; Rosetta Stone, which uses a gene fusion event in a second organism to infer functional relatedness; and the gene neighbor method, which uses both gene proximity and phylogenetic distribution to infer linkage	http://prl.mbi.ucla.edu/prlbeta/
SCOPPI	Domain–domain interactions and their interfaces derived from PDB structure files and SCOP domain definitions	http://www.scoppi.org/
STRING	Protein functional linkages from experimental data and computational predictions	http://string.embl.de/
PINA	Protein Interaction Network Analysis (PINA) platform is an integrated platform for protein interaction network construction, filtering, analysis, visualization, and management. It integrates protein–protein interaction data from six public curated databases (IntAct, BioGRID, MINT, DIP, HPRD, MIPS/MPact) and builds a complete, nonredundant protein interaction dataset for six model organisms. Moreover, it provides a variety of built-in tools to filter and analyze the network for gaining insight into the network	http://cbg.garvan.unsw.edu.au/pina/
MIPS/Mpact	MIPS Mammalian Protein–Protein Interaction Database is a collection of manually curated high-quality PPI data collected from the scientific literature by expert curators	http://mips.helmholtz-muenchen.de/proj/ppi/

Table 6.5 (continued)

Database	Description	URL
MiMI	Michigan Molecular Interactions (MiMI) database comprehensively includes protein interaction information that has been integrated and merged from diverse protein interaction databases and other biological sources	http://mimi.ncibi.org/MimiWeb/main-page.jsp
UniHI	Unified Human Interactome (UniHI) integrates human protein–protein and transcriptional regulatory interactions from 15 distinct resources. The UniHI database includes tools (1) to search for molecular interaction partners of query genes or proteins in the integrated dataset; (2) to inspect the origin, evidence, and functional annotation of retrieved proteins and interactions; (3) to visualize and adjust the resulting interaction network; (4) to filter interactions based on method of derivation, evidence, and type of experiment as well as based on gene-expression data or gene lists; and (5) to analyze the functional composition of interaction networks	http://www.unihi.org/

Table 6.6 Tools used for PPIs and their networks

Tool	Distinctive features	Webpage
BioLayout express 3D	Facilitates microarray data analysis	http://www.biolayout.org/
Cytoscape	Versatile; implements many visualization algorithms; many plug-ins available	http://www.cytoscape.org/
Large graph layout	Especially useful for dynamic visualization of large graphs, force-directed layout algorithm	http://sourceforge.net/projects/lgl
Osprey	Provides network, connectivity filters, many layouts, and facilitates dataset superimposing	http://biodata.mshri.on.ca/osprey/servlet/Index
Pajek	Especially useful for the analysis of very large networks	http://vlado.fmf.uni-lj.si/pub/networks/pajek/
VisANT	Especially facilitates analysis of gene ontologies	http://visant.bu.edu/
yED	General purpose graph editor	http://www.yworks.com/products/yed/
Arena3D	3D view of the network	http://arena3d.org
MEDUSA	It was specially designed and optimized for accessing protein interaction data from STRING database	http://coot.embl.de/medusa

Different algorithms/software for automatic protein–protein docking are available (Table 6.7). In a recent study, we assessed the PPI between the human testis-specific protein, Y-encoded (TSPY) and eukaryotic translation elongation factor 1 alpha 2 (eEF1A2) docked complex and mapped their interacting interface (Fig. 6.3). PPD does not forecast which protein could interact, but it can predict how the proteins interact. Although docked complex with lowest interaction energy is considered as best solution given by PPD, other interaction energies are also involved in large surface displacements in docked conformation to finally form the protein complex.

Table 6.7 Details of protein–protein docking servers

Tools	Description	Web link
3D-Dock Suite	Integrated approach to protein docking with FTDock, RPScore, and MultiDock	www.sbg.bio.ic.ac.uk/docking
3D-Garden	System for modeling protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm	www.sbg.bio.ic.ac.uk/~3dgarden
Bielefeld protein docking	It detects geometrical and chemical complementarities between surface of proteins and estimates docking positions	www.techfak.unibielefeld/~posch/DOCKING/install.html
BiGGER	Protein docking algorithm integrated in Chimera, a molecular graphics and modeling program for studying protein structures and interactions	www.cqfb.fct.unl.py/bioin/chemera
ClusPro	Integrated approach to protein docking with DOT and ZDOCK and PIPER	www.cluspro.bu.edu/
DOT	It computes the electrostatic potential energy between two given proteins or other charged molecules	www.sdsc.edu/CCMS/dot/
ZDOCK	Performs a full rigid-body search of docking orientations between two proteins including performance optimization and a novel pairwise statistical energy potential	www.zdock.umassmed.edu/software
PIPER	FFT-based docking with pairwise potentials	www.structure.bu.edu/content/protein-protein-docking
Eschler NG	Enhanced version of the original ESCHER protein–protein automatic docking system developed in 1997	www.ddl.unimi,it/escherng/index.htm
HADDOCK	High-ambiguity-driven biomolecular docking that employs biochemical and/or biophysical interaction data	www.nmr.chem.uu.nl/haddock
Hex	Protein docking and molecular superposition program	www.hex.loria.fr
RosettaDock	Predicts the structure of protein complexes given the structures of the individual components and an approximate binding orientation	www.rosettadock.graylab.jhu.edu
FireDock	Fast interaction refinement in molecular docking is a web server for flexible refinement and scoring of protein–protein docking solutions	www.bioinfo3d.cs.tau.ac.il/FireDock/
GRAMM-X	It is a public web server for proteinx96 protein docking	www.vakser.bioinformatics.ku.edu/resources/gramm/grammx

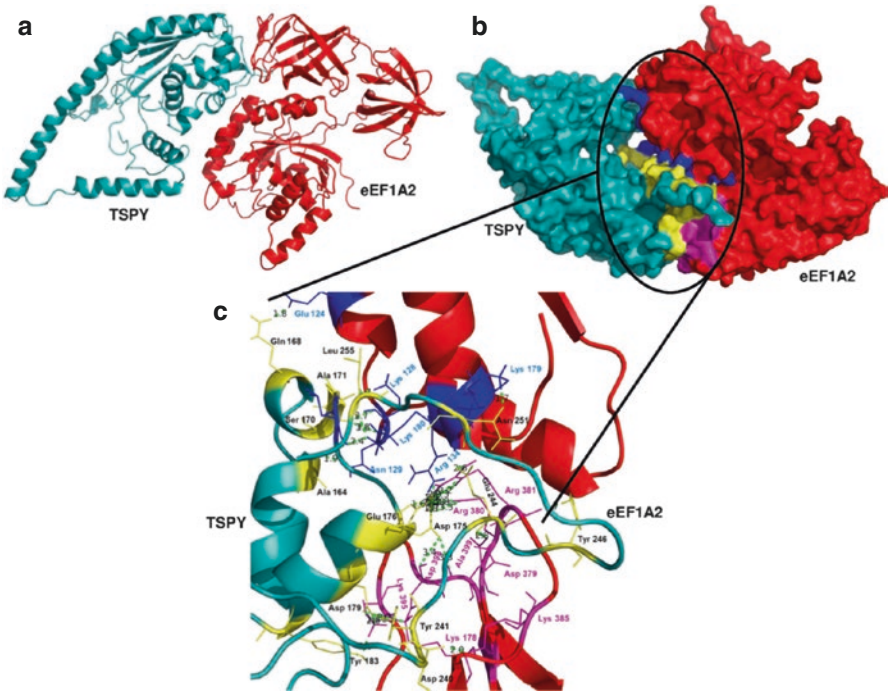

Fig. 6.3 Docked model of the TSPY–eEF1A2 complex and its structural mapping of intermolecular interfaces. (**a**) Cartoon representation of TSPY–eEF1A2 docked complex with low energy score determined by HADDOCK docking. TSPY displayed in teal and eEF1A2 in red. (**b**) A close-up of interacting interface TSPY–eEF1A2 complex is represented in surface representation. Interacting residues from TSPY (*yellow*) and domain I and domain III of eEF1A2 are shown in *blue* and *magenta*, respectively. (**c**) Interacting residues between TSPY and eEF1A2 obtained through hydrogen bonding are shown in line representation, TSPY residues (*yellow*), and eEF1A2 (domain I *blue* and domain III *magenta*) with the atomic distances (*green dotted lines*) with labeled residue name (Adapted from Panwar et al. 2015)

6.7 Domain–Domain Interactions

As protein–protein interactions mostly occur via domains instead of the whole protein surface, identification of domain–domain interaction (DDI) is an imperative step toward PPI prediction. Protein domains are compact regions within the protein's structure that possess a distinct function. On an average, observed protein domains have more than 200 residues, with an average domain size of about 100 residues.

A protein can have the multiple or a single domain, each one characteristically assigned with a precise function (Teichmann 2002), and combination of these domains can define the function and interaction of protein (Ingolfsson and Yona 2008). Protein domains from *homologous* might exist conservation in the interaction patterns across the domain superfamily as they preserve the same three-dimensional structures. Each domain also forms a three-dimensional structure that

is independently stable and folded. As 3D structure of proteins is critical for prediction of PPIs; the interacting interfaces (domains) must be folded into specific conformations so that they interact with other proteins (physically and energetically) (Fig. 6.4). Therefore, identifying the PPI at domain level can lead to an important step for identification of PPI (Deng et al. 2002; Lee et al. 2006; Ng et al. 2003; Guimaraes et al. 2006; Ta and Holm 2009).

Several methods have been used in prediction of domain interactions from PPI data graphs, but the first DDIs were identified based on 3D structures of protein complexes from Protein Data Bank. Databases such as iPfam, 3DID, and PInS extract DDIs from the interacting interfaces of known protein 3D structures. In *Saccharomyces cerevisiae*, *Escherichia coli*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens*, DDIs covered are less than 20% of their PPIs in databases. To complement DDIs, various computational methods have been proposed to predict DDIs in recent years (Table 6.8). Although 3D structures are the

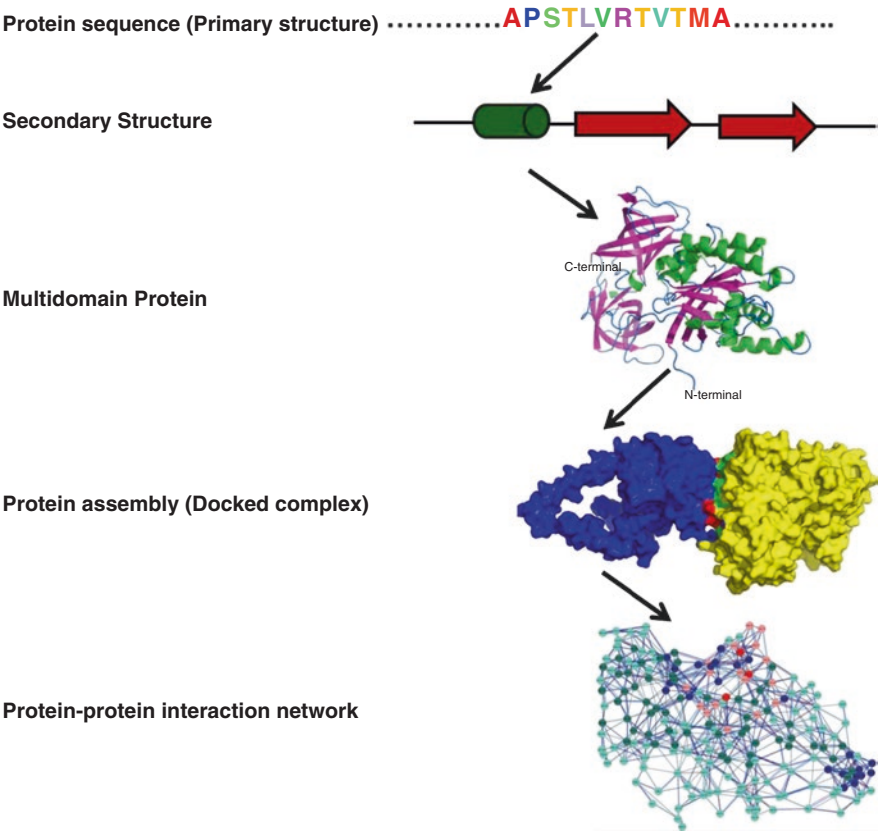

Fig. 6.4 A workflow of protein structure organization. Three basic levels are assigned to protein: primary, secondary, and multi-domain (tertiary and quaternary structure). Then multi-domain proteins form structural networks by interacting with each other at the residue level forming a complex network

Table 6.8 Databases and tools for domain–domain interaction

Tools	Description	Web link
3D interacting domains (3DID)	Search for domain–domain interactions in proteins for which high-resolution three-dimensional structures are known	www.3did.irbbarcelona.org
PFAM	A semiautomated definition of families (HMM based)	www.pfam.sanger.ac.uk/
SMART	Smaller number of domain families (manually curated)	www.smart.embl-heidelberg.de/
InterPro	Perhaps the most comprehensive tool	www.ebi.ac.uk/interpro/
Database of Domain Interactions and Bindings (DDIB)	Search for documented information on biomolecule interaction, especially protein domain–domain interactions	www.ddib.org/
DOMINE—a database of protein domain interactions	Database of known and predicted protein domain (domain–domain) interactions	www.domine.utdallas.edu/
DOMINO—a database of domainx96peptide interactions	Search for annotated experiments describing interactions mediated by protein interaction domains	www.mint.bio.uniroma2.it/domino/
CSDBase	Cold Shock Domain database	www.chemie.uni-marburg.de/~csdbase/
DIMA	Domain interaction map: experimental and predicted protein domain interactions	www.mips.gsf.de/genre/proj/dima2
DomIns	Database of Domain Insertions: Domain insertions in known protein structures	www.domins.org/
InterDom—a database of putative interacting protein domains for validating predicted protein interactions and complexes	Find evidence for the detected protein interactions based on putative protein domain interactions	www.InterDom.lit.org.sg
PDZBase—protein–protein interactions involving PDZ domains	Search for information on protein–protein interaction involving PDZ domains	www.icb.med.cornell.edu/services/pdz/start
PepCyber:P~PEP—a database of human protein–protein interactions mediated by phosphoprotein-binding domains	Database specialized in documenting human PPBD-containing proteins and PPBD-mediated interactions	www.pepcyber.org/PPEP/
PROCOGNATE—a cognate ligand domain mapping for enzymes	Database of cognate ligands for the domains of enzyme structures in CATH, SCOP, and Pfam	www.ebi.ac.uk/thornton-srv/databases/procognate/
The Homeodomain Resource	Search for curated information for the homeodomain protein family	www.genome.nhgri.nih.gov/homeodomain

base of predicting DDIs, many other computational methods like Pfam and CDD (Conserved Domain Database) begin annotating the domains from primary sequences. These databases share the protein domain annotation data, as each database has the unique annotations format. Some databases, e.g., InterPro (Hunter et al. 2009) and CDD (Marchler-Bauer et al. 2011), provide protein domain annotation information collected from several databases. Thus, domains are considered the fundamental units of protein structure, folding, function, and evolution.

Conclusions

The genomics and proteomics approaches coupled with bioinformatics tools have a symbiotic relationship; new experimental methods require newly adapted tools besides well-established techniques. The bioinformatics tools enable high-throughput production of experimental results with quality control, transformation of these results into protein networks, and their exploitation through visualization and analysis tools. Hence, bioinformatics-based approaches serve as an additional tool for data mining, as well as for gene identification and disease prediction.

Competing Interests No competing interests to be disclosed.

References

Adams MD, et al. The genome sequence of Drosophila melanogaster. Science. 2000;287(5461):2185–95.

Aloy P, Russell RB. Ten thousand domain-domain interactions for the molecular biologist. Nat Biotechnol. 2004;22(10):1317–21.

Altshuler D, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;407(6803):513–6.

Bartel PL, Fields S (eds). The yeast two-hybrid system. In: Bartel PL, Fields S, editors, Advances in molecular biology. New York: Oxford University Press; 1997.

Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. 1999;17(3):121–7.

Bult CJ, et al. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science. 1996;273(5278):1058–73.

C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. 1998;282(5396):2012–8.

Cambien F, et al. Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature. 1992;359(6396):641–4.

Chakrabarti P, Janin J. Dissecting protein–protein recognition sites. Proteins. 2002;47(3):334–43.

Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. Dis Markers. 2001;17(2):89–98.

Cheng I, et al. 8q24 and prostate cancer: association with advanced disease and meta-analysis. Eur J Hum Genet. 2008;16(4):496–505.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005;37(11):1217–23.

De Jager PL, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009;41(7):776–82.

Deng M, Mehta S, Sun F, Chen T. Inferring domain-domain interactions from protein-protein interactions. Genome Res. 2002;12:1540–8.

Dove A. Proteomics: translating genes into products? Nat Biotechnol. 1999;17(3):233–6.

Duerr RH, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science. 2006;314(5804):1461–3.
Enright AJ, Ililopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature. 1999;402(6757):86–90.
Fields S, Song O. A novel genetic system to detect protein–protein interactions. Nature. 1989;340(6230):245–6.
Fleischmann RD, et al. Whole genome random sequencing and assembly of haemophilus-influenzae Rd. Science. 1995;269:496–512.
Fraser CM, et al. The minimal gene complement of mycoplasma-genitalium. Science. 1995;270(5235):397–403.
Freedman ML, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African–American men. Proc Natl Acad Sci USA. 2006;103(38):14068–73.
Goffeau A, et al. Life with 6000 genes. Science. 1996;274(5287):546, 563–7.
Gratacòs M, et al. A polymorphic genomic duplication on human chromosome 15 is a susceptibility factor for panic and phobic disorders. Cell. 2001;106(3):367–79.
Guimaraes KS, Jothi R, Zotenko E, Przytycka TM. Predicting domain-domain interactions using a parsimony approach. Genome Biol. 2006;7:R104.
Howard TD, et al. Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma. Am J Hum Genet. 2002;70(1):230–6.
Hunter S, et al. InterPro: the integrative protein signature database. Nucleic Acids Res. 2009;37(D):D211–5.
Ingolfsson H, Yona G. Protein domain prediction. Methods Mol Biol. 2008;426:117–43.
International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–45.
Jacques PF, et al. Relation between folate status, a common mutation in methylene tetrahydrofolate reductase, and plasma homocysteine concentrations. Circulation. 1996;93(1):7–9.
Kortemme T, Baker D. Computational design of protein–protein interactions. Curr Opin Struct Biol. 2004;8(1):91–7.
Kraft P, Cox DG. Study designs for genome-wide association studies. Adv Genet. 2008;60:465–504.
Lander ES, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
Lee H, Deng M, Sun F, Chen T. An integrated approach to the prediction of domain-domain interactions. BMC Bioinformatics. 2006;7:269–27.
Libioulle C, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet 2007, 3(4):e58.
LoConte L, et al. The atomic structure of protein–protein recognition sites. J Mol Biol. 1999;285(5):2177–98.
Lu L, Lu H, Skolnick J. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. Proteins. 2002;49(3):350–64.
Lykke-Andersen J. mRNA quality control: marking the message for life or death. Curr Biol. 2001;11(3):R88–91.
Marchler-Bauer A, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 2011;39(D):D225–9.
Marcotte EM, et al. Detecting protein function and protein–protein interactions from genome sequences. Science. 1999;285(5428):751–3.
Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. Bioinformatics. 2010;26(4):445–55.
Mount DW, Pandey R. Using bioinformatics and genome analysis for new therapeutic interventions. Mol Cancer Ther. 2005;4(10):1636–43.
Musani SK, et al. Detection of gene x gene interactions in genome-wide association studies of human population data. Hum Hered. 2007;63(2):67–84.
Ng SK, Zhang Z, Tan SH, Lin K. InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. Nucleic Acids Res. 2003;31(1):251–4.

Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, et al. A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform. 2014;15(2):256–78.
Panwar D, Rawal L, Ali S. Molecular docking uncovers TSPY binds more efficiently with eEF1A2 compared to eEF1A1. J Biomol Struct Dyn. 2015;33(7):1412–23.
Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood. 1996;88(10):3698–703.
Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001;17(9):502–10.
Richards RI, Holman K, Yu S, Sutherland GR. Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. Hum Mol Genet. 1993;2(9):1429–35.
Salwinski L, Eisenberg D. Computational methods of analysis of protein–protein interactions. Curr Opin Struct Biol. 2003;13(3):377–82.
Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19(3):212–9.
Seng KC, Seng CK. The success of the genome-wide association approach: a brief story of a long struggle. Eur J Hum Genet. 2008;16:554–64.
Small KM, Wagoner LE, Levin AM, Kardia SL, Liggett SB. Synergistic polymorphisms of beta1- and alpha2C-adrenergic receptors and the risk of congestive heart failure. N Engl J Med. 2002;347(15):1135–42.
Ta HX, Holm L. Evaluation of different domain-based methods in protein interaction prediction. Biochem Biophys Res Commun. 2009;390(3):357–62.
Teichmann SA. Principles of protein-protein interactions. Bioinformatics. 2002;18(Suppl 2):S249.
The International HapMap Consortium. The International HapMap Project. Nature. 2003;426(6968):789–96.
Uetz P, et al. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. Nature. 2000;403(6770):623–7.
Valencia A, Pazos F. Computational methods for the prediction of protein interactions. Curr Opin Struct Biol. 2002;12(3):368–73.
Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291:1304–51.
Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–78.

Dr. Leena Rawal Dr. Leena Rawal is currently working as a scientist at the Department of Cytogenetics, Dr Lal PathLabs Limited, New Delhi. She holds a doctoral degree in molecular genetics from the National Institute of Immunology, New Delhi, and has significant years of experience in the field. Her expertise lies in the area of human and animal genetics, proteomics, gene regulation, cytogenetics, and bioinformatics. Earlier, she has headed the Molecular Diagnostics and Research Division of a renowned pharmacogenomics-based organization that provides state-of-the-art in vitro personalized diagnostic services to medical healthcare and allied communities.

Genomics of the Human Y Chromosome: Applications and Implications

7

Sanjay Premi, Jyoti Srivastava, and Sher Ali

7.1 Introduction

Mammalian sex chromosomes are often uniquely different from each other in terms of their structural organization and genes related to sex determination and differentiation. For instance, ~78 genes identified on the male-specific region of human Y chromosome (MSY) express mostly in testis and code for ~27 distinct proteins. According to a largely believed hypothesis, human Y chromosome lost most of the genes during its evolution except the ones essential for male sex determination. This evolutionary degeneration of the Y chromosome is commonly linked to its inability to undergo homologous recombination with the X chromosome or any of the autosomes. Due to its "gene-poor" landscape and continuously decreasing size, Y chromosome was hypothesized to disappear in ~10 million years. However, abundant literature from modern day research provides evidence on its continual sustenance. First is the MSY which is a large portion of the Y chromosome, and owing to which

S. Premi (✉)
Therapeutic Radiology, Yale Medical School, Yale Cancer Center, Yale University, New Haven, CT 06511, USA

Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India
e-mail: sanjay.premi@yale.edu

J. Srivastava (✉)
Department of Pharmacology, Yale Medical School, Yale University, New Haven, CT 06511, USA

Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India
e-mail: jyoti.srivastava@yale.edu

S. Ali
Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_7

Y does not participate in X-Y recombination. Any further reduction in its size would be a serious threat to human existence. Secondly, MSY is a result of segmental duplications (Hurles and Jobling 2003). These duplications lead to gene conversions and thus protect the human Y chromosome. Also, the Y chromosome is dominant as is witnessed by a male phenotype in patients with multiple X but only a single Y chromosome. Moreover, the highly palindromic and repetitive landscape of the Y chromosome leads to enhanced mutation rate which fuels higher levels of polymorphisms (Jobling et al. 2007). The MSY was described in detail by Jobling and Smith in 2003 (Fig. 7.1). Surprisingly, several Y chromosome haplotypes maintain fertility even without essential Y-linked genes. This highlights two facts:

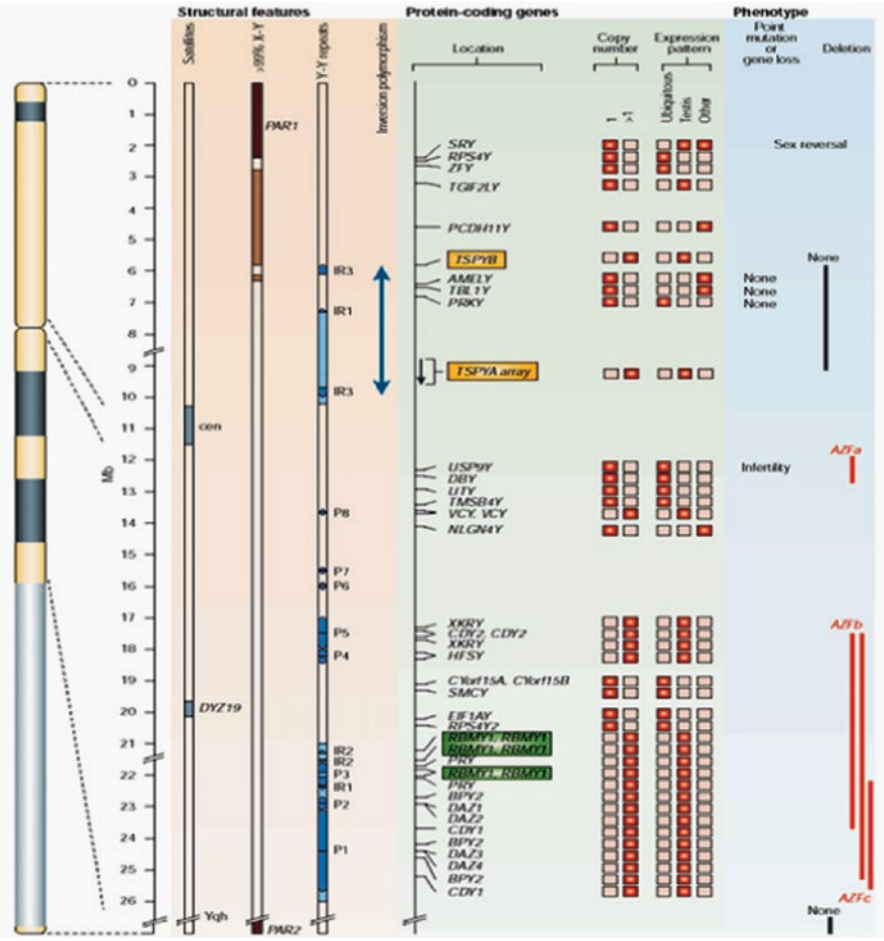

Fig. 7.1 Representative geography of the human Y chromosome. Locations of various loci, genes, and cytogenetic features are shown from *left* to *right* with numbering originating from short arm telomere. IR is inverted repeats and P for palindromes. The 27 known Y-linked genes and their estimated expression are also shown. Some of the well-studied clinical phenotypes associated with Y chromosome are designated on the *right*. This figure is taken from Jobling et al. (2003) *Nature Reviews* 4, 598–612

repetitive landscape of the Y acts as a buffer for loss of its genes and that Y chromosome to autosomal interactions might be essential for sustenance of male fertility. This chapter describes organizational complexities of the Y chromosome under various normal and disease phenotypes and effect of exogenous/environmental factors in augmenting these complexities. Due to its unique structural organization and nonhomologous nature, the Y chromosome is a unique tool for DNA-based diagnosis under normal and abnormal conditions related to male fertility.

7.2 Y Chromosome Evolution

Y chromosome is humans and the equivalent W chromosome in several other animals evolved independently. However, both of these sex chromosomes are gene poor, heterogametic, and often smaller than the autosomes (Hurst and Randerson 1999). Evolutionarily, four different processes have been hypothesized for the emergence of Y and W. First is *asexual degeneration*, which is defined as degeneration of a particular chromosome owing loss of its ability to recombine. This leads to loss of genes/DNA from the male sex chromosomes Y and W (Charlesworth and Charlesworth 2000; Steinemann 2000). Second is *sexual antagonism*, which is augmented fitness of a single or pool of genes on particular chromosomes. This masculinized the Y chromosome and feminized the X chromosomes (Brooks 2000; Charlesworth et al. 1987; Gavrilets et al. 2001; Wolfenbarger and Wilkinson 2001). Third is *constant selection*, which is sustained inheritance of the Y chromosome in human males leading to uniparental inheritance (Ting 1998; Wu et al. 2000). And the last hypothesis is *hemizygous exposure*, which is functional promotion of a particular chromosome (Y in humans) by fixing recessive mutations on the partner X chromosome. This process predicts the masculinization of the X chromosome. This is seen largely in reptiles where only Y or W chromosomes do not decide the sex, but the sex determination relies on temperature. This maintains the flexibility and sex ratios.

Since the Y and X chromosomes have distinctly different sizes, the Y chromosome is suggested to be a degenerated autosome, whereas the X behaves just like the autosomes. The pseudoautosomal region (PAR), a small region of Y homology with X chromosomes, consists of X-transposed and X-degenerate sequences (Skaletsky et al. 2003). Thus the X and Y probably originated from an autosomal chromosome pair. It is hypothesized that degeneration and differentiation of the Y started when it is accumulated by sex determination genes (Waters et al. 2001, 2005). Some evolutionary traces of the human Y chromosomes can still be seen in placental mammals, marsupials, and monotremes (Delbridge et al. 1997; Glas 1999; Pask et al. 2000). Although the expressional profile of Y-linked genes is still being explored, the human Y chromosome is considered as the most evolved sex chromosome. Further detailed investigations are envisaged to locate expressional variability of the Y-linked genes under abnormal conditions. Such information is critical for diagnosis, prognosis, comparative genomics, and probably ascertaining the causes of male infertility in addition to an important tool in better understanding of the Y chromosome.

7.3 Y-Linked Genes

Compared to functionally assorted genes or gene families on autosomes, Y-linked genes have distinct expressional profiles (Jobling and Tyler-Smith 2003). The non-recombining region on Y (NRY) or male-specific region (MSY) contains three distinct classes of genes. Class 1 includes "X-degenerate" single copy genes expressed ubiquitously having X-linked homologues with similar functions. Class 2 includes "X-transposed" genes supposed to have transposed to Y from X chromosome. These multi-copy genes express specifically in testis. Class 3 includes single copy Y-linked genes such as *SRY* which expresses specifically in embryonic bipotential gonad and adult testis (Harley et al. 2003). Others genes belonging to this category are *AMELY* and *PCDHY* which, with their active X-linked homologues, are expressed in developing tooth buds and brain, respectively (Crow 2000; Nachman 2001; Oota et al. 2001). Recently, another ampliconic region of MSY having more than 99.9% homology among them has been reported (Skaletsky et al. 2003). The ampliconic sequences harbor highest density of coding and noncoding Y-linked genes with variable copy numbers.

In addition to the MSY, several Y chromosomal regions do not abide by general rules of genetics. For example, most genes on the PAR maintain dosage compensation by eluding X-inactivation, whereas synaptobrevin like 1 (*SYBL1*) and sprouty homologue 3 (*HSPRY3*) on the long arm PAR, respectively, undergo X and Y inactivation in females and males. This epitomizes a complex evolutionary trail or gene translocations across X and Y (Jobling and Tyler-Smith 2003). Heterochromatin on Y chromosome is not supposed to transcribe; however, several chimeric and testis-specific transcripts have been reported recently from the long arm Y heterochromatin (Jehan et al. 2007). Genes on Y chromosome are envisaged to originate on several autosomes. One such example is the RNA-binding motif (*RBMY*) which is a spermatogenesis gene from the proto-XY pair. Nearly identical homologue of RBMY is called *HNRPG* or *RBMX* which is present on the human X chromosome (Delbridge et al. 1999). RBM is also present on X and Y chromosomes in rodent genome (Mazeyrat et al. 1999). Thus, it can be construed that the *RBM*/*RBMY* evolved from very old and common ancestors like proto-XY pair.

Two genes *ZFY* and *DFFRY* (*USP9Y*) are the examples added recently to the Y chromosomes. Both of these genes have copies on the X chromosome which probably evolved into Y homologues. But the marsupial *ZFY* and *USP9Y* are autosomal suggesting their recent addition to the eutherian X and Y chromosomes. The *ZFY* and *ZFX* express ubiquitously in human whereas *ZFY* is testis specific in mouse (Koopman et al. 1989). Similarly, expression of *DFFRY* is ubiquitous in humans and testis specific in mouse. *DFFRY* mutations are commonly detected in infertile human males (Saxena et al. 1996, 2000). Further, this gene seems to affect ovarian functions in *Drosophila*. It can be hypothesized that this gene acquired gonadal functions and then translocated to the Y chromosome.

Another example is four copies of the *DAZ* genes which code for RNA-binding motifs but lack an X chromosome homologue. Autosomal homologue *DAZLA* has been reported in mice and marsupials but not the *DAZ* (Delbridge et al. 1997). These

DAZ genes probably originated from transposition (Saxena et al. 1996, 2000). Another example is multicity human *CDY* gene which is completely absent in mouse genome. Instead, the autosomal *CDYL* performs both ubiquitous and testis-specific functions in mouse. Human *CDYL* contains introns, but *CDY* on the Y is intronless indicating that the latter must have translocated from chromosome 6 to Y through retroposition (Dorus 2003; Lahn 1997; Mazeyrat et al. 1999).

Interestingly, the most celebrated sex determination gene, *SRY*, also has a homologue *SOX3* on the X chromosome. The *SOX3* demonstrates dose-based sex differentiation signals in marsupials (Pask et al. 2000). In higher mammals like mouse and humans, *SOX3* specifically expresses in central nervous system (CNS) and genital ridge (Shen and Ingraham 2002). Sequence comparisons suggest *SRY* as a surviving relic of *SOX3* and possibly arose from a gene on the proto-sex chromosomes. It is likely that *SRY* and *SOX3* have different modes of actions in different species and possible brain functions in both sexes (Graves 2002).

7.4 Y Chromosome and Male Sex Determination

Sex determination and male (in)fertility is strongly linked to Y chromosome. Mouse models with two X chromosomes but transgenic for *SRY* develop male gonads suggesting *SRY* to be the testis-determining factor (TDF). Also, mutations in the *SRY* gene frequently lead to dysgenesis of male gonads in XY females which further supports its essential role in male sex determination (Harley et al. 2003). These mutations are often missense in the high-mobility group (HMG) domain of the SRY protein (Harley et al. 2003). Functionally, SRY is a transcription factor which binds to the DNA in a typical sequence-specific manner, a function which if impaired leads to clinical XY gonadal dysgenesis (Harley et al. 1992; Nasrin et al. 1991). Of several mutations often seen in the SRY, the ones located in the HMG box specifically disrupt its DNA-binding capability (Harley et al. 1992; Mitchell and Harley 2002; Pontiggia et al. 1995; Schmitt-Ney et al. 1995; Tiepolo and Zuffardi 1976).

In addition to the *SRY*, genes located in the azoospermia factor (AZF) region on Y chromosome are also crucial for maintaining spermatogenesis (Vogt 1996). AZF comprises of three nonoverlapping subregions arranged from proximal to distal long arm euchromatin of the Y chromosome. These subregions are called AZFa, AZFb, and AZFc (Vogt 1996), and each has its own candidate genes. Ubiquitin-specific protease 9, Y chromosome (*USP9Y*) gene, covers almost half of the AZFa sequences (Brown et al. 1998; Mazeyrat et al. 1999; Vogt 1997). Loss of *USP9Y* from AZFa is generally associated with spermatogenic failures (Ferlin 1999; Foresta et al. 2000; Vogt 1997). Additionally, dead box on Y chromosome (*DBY*) and ubiquitous TPR motif on the Y (*UTY*) are two more candidate genes on the AZFa region (Lahn 1997; Mazeyrat et al. 1999). It has been suggested that *USP9Y* deficiency leads to altered spermatogenesis and this oligospermic phenotype is worsened with additional loss of *DBY* (Brown et al. 1998; Sun et al. 2000; Foresta et al. 2000). Though *DBY* is frequently deleted in infertile males, it is expressed ubiquitously.

Hence, the specific functional role of *DBY* in spermatogenesis and male germ cell development is still not completely explored.

Currently, translation initiation factor IA-isoform Y (*EIF1AY*), RNA-binding motif on Y (*RBMY*), and heat shock transcription factor on Y (*HSFY*) are the only three genes attributed to the AZFb region (Skaletsky et al. 2003; Tessari 2004). *EIF1AY* is expressed ubiquitously with one transcript specifically detected in testis. *EIF1AY* codes for a translation initiation factor, eIF-1A, and the biological function of eIF-1A still remains to be determined. *RBMY* belongs to a family of 20–50 genes and pseudogenes which are spread over both the arms of Y chromosome with one locus in AZFb (Prosser et al. 1996). *RBMY* is specifically expressed in male germline cells substantiating its role in spermatogenesis. Protein coded by *HSFY* contains an HSF type of DNA-binding domain, related to HSF2 gene on chromosome 6. This gene is active during spermatogenesis and embryogenesis only (Pirkkala 2001). Overall, the large Y chromosomal deletions detected in infertile males often have their breakpoints in the AZFb region (Ferlin 2003).

AZFc region is located in distal Yq11, and its deletion is the most frequent cause of azoospermia or severe oligozoospermia and hence male infertility (Krausz 1999). AZFc is composed of large blocks of repeat sequences leading to the ampliconic landscape (Kuroda-Kawaguchi et al. 2001). These amplicons are arranged into palindromes which are responsible for duplications, deletions, and gene conversions through intrachromosomal recombination. The large palindromes with high intrachromosomal sequence homology also ensures structural integrity. Currently, eight spermatogenesis genes have been attributed to the AZFc which are *BPY*, *CDY1*, *CSPG4LY*, *DAZ*, *GOLGA2LY*, *TTY3*, *TTY4*, and *TTY17* (Kuroda-Kawaguchi et al. 2001). Distinct proteins have only been detected for three of these genes. The deleted in azoospermia (*DAZ*) gene consists of nearly identical four copies. Each of these copies codes for DAZ repeats on C-terminus and an RNA-binding domain at the N-terminus (Mahadevaiah 1998; Reijo et al. 1995). Owing to its recurrent deletion in human males with spermatogenic failure, *DAZ* is considered to be one of the potential spermatogenesis genes located on AZFc (Lahn 1997; Wong et al. 1999; Yen 1998). Functional significance of other AZFc gene remains unknown except that most of them have multiple copies with testis-specific transcription. Based upon its location in the *DAZ* locus, one of the two *CDY* genes is also presumed to be essential for spermatogenesis (Yen 1998). Individual functional analyses of the *CDY* and *DAZ* genes will uncover the actual relative significance of these two genes in spermatogenesis.

In addition to specific genes, Y chromosomal microdeletions are also envisaged to cause spermatogenic failure and thus azoospermia and oligospermia. Sequence-tagged sites (STSs) are generally used to map these microdeletions and localize them to specific AZF regions during standard diagnosis regime (Vollrath et al. 1992). One key contradiction is detection of ~0.4% of all known microdeletions in normal males which are postulated to be part of natural polymorphism (Kent-First et al. 1999; Kobayashi et al. 1995; Pryor et al. 1997). There is a prevalence of microdeletions (up to 35%) in infertile males which may be due to pressure from natural selection or an actual diagnostic/prognostic

marker. Thus, the clinical relevance of microdeletions is still debatable. Nevertheless, microdeletions are frequently concentrated in AZFc region (up to 59.6%) including *DAZ* loci, AZFb region (up to 15.8%) including *RBMY* loci, and AZFa region (up to 4.9%). Microdeletions in the AZFc region are strongly linked with severe male infertility phenotypes like azoospermia, SCOS (Sertoli cell-only syndrome), and arrest of the spermatogenesis process. AZFc microdeletions are associated with azoospermia, severe oligozoospermia, and testicular pathologies like Sertoli cell-only syndrome (SCOS) to spermatogenic arrest and hypospermatogenesis. On the other hand, AZFa deletion mostly leads to SCOS and complete arrest of spermatogenesis which is often traced back to mutations and deletions of the *USP9Y* or *DBY* genes (Ferlin 1999, 2003; Foresta et al. 2000; Lin et al. 2005).

Intrachromosomal homologous recombination (ICHR) is prevalent on human Y chromosome due to palindromic repeats causing major interstitial deletions (Repping et al. 2002; Saxena et al. 2000; Sun et al. 2000). AZFc is more prone to ICHR followed by the AZFa region. Moreover, human endogenous retrovirus (HERV15) sequences in AZFa are known to recombine intrachromosomal and cause major deletions or duplications responsible for the azoospermia phenotype (Bosch 2003; Sun et al. 2000). The additional 100–200 bp repeat spread all over the AZFa and the HERV sequences further increase the instances of ICHR leading to minor and major deletion events. Similar to AZFa, the AZFc contains six major amplicons (115 kb gray to ~700 kb yellow) which are nearly identical (Skaletsky et al. 2003). These palindromes are major targets for ICHR. For instance, recombination between palindromes P5 and P1 leads to a massive deletion extending from the AZFb region to ~1.5 Mb into AZFc removing ~6 Mb total (Repping et al. 2002). Another recombination event between P5 and P1 removes ~7.7 Mb, which is one of the largest ICHR-mediated deletions known. ICHR mediated-deletions and structural reorganization in AZFc region are also known to be implicated in spermatogenic failure too (Vogt 2004; Repping et al. 2003). In conclusion, ICHR-mediated chromosomal alterations and microdeletions are largely common in patients with male infertility (Marshall Graves 2000; Premi et al. 2008).

7.5 Polymorphic Nature of the Y Chromosome

Segmental duplications and copy number variations (CNVs) are expensively described in human genome (Bailey 2002; Sebat 2004). Owing to its repetitive nature, the human Y chromosome endure extensive deletions, inversions, and neutral translocations (Bernstein et al. 1986; Schmid et al. 1984). The repetitive nature also contributes to Y polymorphism among different ethnic groups. However, due to lack of concrete data, the haplotypic polymorphism cannot yet be completely explained by palindrome-based ICHR.

Prenatal sex determination, archaeological analysis, forensic typing, and paternity testing are based on a routine assay for the amelogenin gene *AMELY* on Yp.

However, inverted repeat 3 (IR3)-based inversion has been reported to cause an inversion on Yp leading to two different orientations of the *AMELY* (Skaletsky et al. 2003). Similarly, ICHR between various copies of the testis-specific protein on Y (*TSPY*) gene is known to delete this gene from the Yp region in human males (Santos et al. 1998). Evolutionarily, *TSPY* deletions have been reported at least seven times independently. Moreover, some deletion events mediated by ICHR, often leading to deletion of *PRKY* and *TBL1Y*, are neutral and do not cause any clinical phenotypes (Jobling et al. 2007).

Palindromic repeats and their structural organization on the AZFc region are extremely prone to inversions and conversions which have been demonstrated by mapping sequence-tagged sites (STSs) and single-nucleotide variations (SNVs). Investigations on Y chromosome haplotypes coupled with STS mapping have demonstrated complete lack of *DAZ*1/*DAZ*2 and *BPY2.2/BPY2.3* duplets in the haplogroup N (Fernandes et al. 2004). This is construed to be normal genetic variation of the human genome and not linked with male infertility. Moreover, substantial Y chromosomal variations are prevalent among individuals across the Y haplogroups or even the same haplotypes. In addition to megabase deletions and duplications, the variations also include specific gene structures, sizes, and duplications/deletions. Four *DAZ* genes in a single locus vary in sizes with *DAZ*1, 65 kb; *DAZ*2, 70 kb; *DAZ*3, 50 kb; and *DAZ*4, 55 kb. The central genomic fragment separating two *DAZ* duplets at each *DAZ* locus is composed of tandemly repeated ~2.4 kb repeat units. Further, each *DAZ* gene harbors variable numbers and sequences of exon 7 and a 10.8 kb unit. Additionally, this 10.8 kb unit has three copies in *DAZ*1 and two copies in *DAZ*4 (Saxena et al. 2000). *DAZ* genes have an additional polymorphism in terms of copies of RNA-binding motif (RBM). Furthermore, *DAZ* genes carry variable numbers of RNA-binding motifs (RBM) (Premi et al. 2010).

Our own investigations have uncovered massive polymorphisms in Y chromosomal landscape which are either endogenous in normal males and males with sex chromosome-related anomalies or induced by exogenous factors like radiation exposure. We uncovered a unique structural reorganization in clinically normal males which did not agree with the structure reported earlier from full-length sequencing of human Y chromosome (Repping et al. 2002, 2003). Detailed investigation on copy number variations and STSs from ~1000 normal males revealed a singular structural organization (Premi et al. 2010). This investigation also identified a surprising translocation of inter-*DAZ* genomic fragments from Yq to Yp in Indian and European Y chromosome. The European Y chromosome was used as a control because the first whole length Y chromosomal sequencing was performed on DNA from a Caucasian male. This translocation still remains a functional and physiological mystery, more so when the promoters for all four *DAZ* genes are located in the translocated inter-*DAZ* region. Localization of this inter-*DAZ* segment on short arm instead of long arm of the Y chromosome implicates an unexplored promoter region either within or outside the *DAZ* locus on the long arm. This also suggests that in spite of harboring promoters for crucial spermatogenesis genes, the inter-*DAZ* region is just

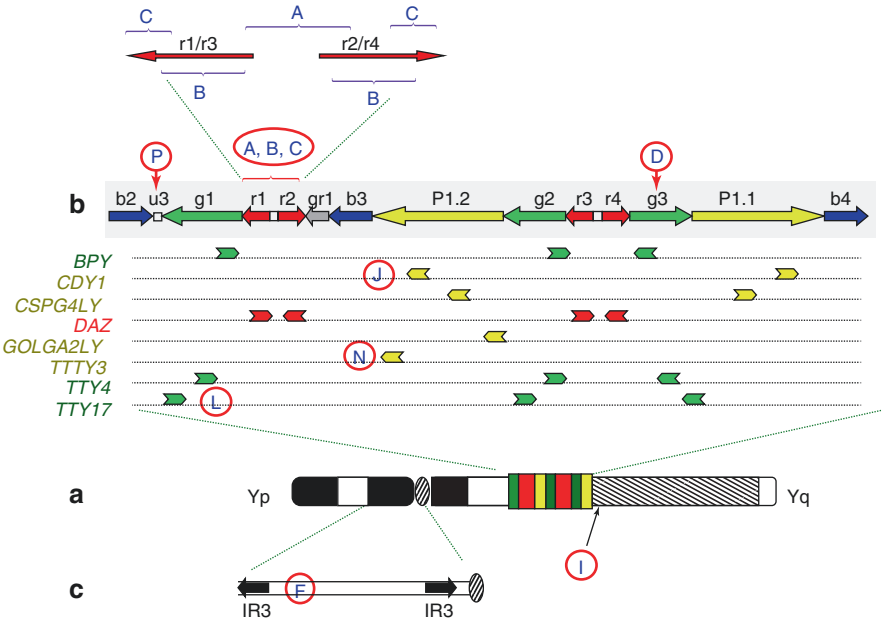

Fig. 7.2 Locations of various fluorescence in situ hybridization (FISH) probes that we used in our investigations on the polymorphic nature of human Y chromosome. *Red circles* indicate probes for that particular region which may encompass a single gene, palindrome, intergenic regions, or multiple genes. (**a**) Y chromosome cartoon with AZFc region as *vertical rectangular bars*. (**b**) Hypothetical expansion of the AZFc region to demonstrate locations of various FISH probes. (**c**) Inverted repeat 3 on long arm of the Y chromosome

another palindrome which translocated from Yq to Yp through ICHR. This translocation is explained in Figs. 7.2 and 7.3.

Further, our own studies suggest that the inter-*DAZ* region is just a small part of ICHR-mediated, intrachromosomal, and interchromosomal translocations. We localized the Y-specific AZFc amplicons *P1.1*/*1.2* on proximal and distal regions of chromosome 15 in addition to the Y chromosome. This polymorphism was a further substantiated localization of *P1.1*/*1.2* genes *TTY3* and *XKRY* onto the proximal 5p region, but not on the Y chromosome. We hypothesized that these Y chromosomal sequences might not be essential for sex determination or spermatogenesis, and hence genome tolerates their translocation onto the autosomes. However, the Caucasian Y map from the database still puts these regions onto the Y chromosome. However, such variations reiterate the fact that the Y chromosome acquired building blocks from various autosomes, or Y itself is a degraded version of an autosome. We also hypothesized that the *DAZ* genes underwent several rearrangement events leading to the current AZFc structural configuration where segments of the AZFc occupy both the Yp and Yq locations. Evolutionarily, the *DAZ* and *CDY* genes, both from the AZFc region, are present on short arm of the Y chromosome in pygmy chimpanzee and the Sumatran orangutan. Also, the *DAZ* genes have an autosomal

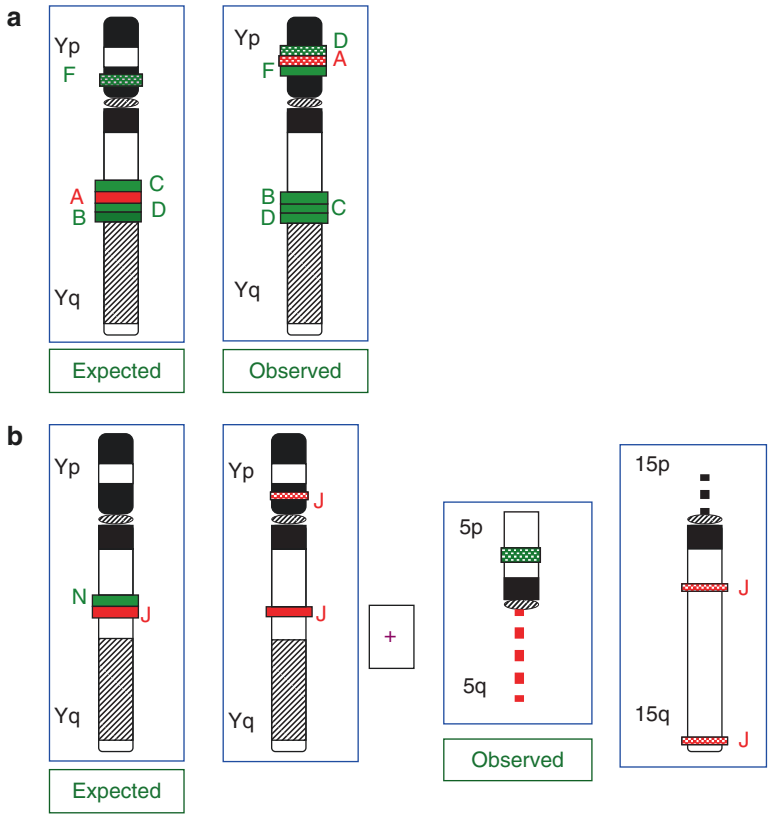

Fig. 7.3 Expected and observed structure of the Y chromosome palindromes and amplicon in 750 Indian males. The human Y chromosome map published in 2003 was taken as reference (http://www.nature.com/nature/focus/ychromosome/). (**a**) AZFc region, specifically the *DAZ* genes and inter-*DAZ* region. (**b**) *TTY3* and *CDY1* gene regions. Taken from Premi et al. (2010) *Chromosome Research* 18: 419–430

homologue called *DAZLA*. In conclusion, the AZFc sequences including important genes originated on the autosomes, translocated onto the Yp region and finally to the Yq region. Both *CDY* and *DAZ* are proposed to be present in highly repetitive sequence clusters on Y chromosome. AZFc region was never investigated empirically before. Our investigations of the MSY uncovered a unique organizational variation in terms of *DAZ* loci and megabase regions of the AZFc itself. This organization has never been reported, and it adds to the preexisting variations and polymorphisms of the Y chromosome. Analysis of large cohorts of Y chromosomes from different geographical regions all over the world is warranted to completely understand the complexity of its structural organization.

In spite of evolutionary degradation, Y chromosome has now adapted to endogenous insults and exogenous environmental factors; otherwise human species will cease to exist. For instance, exposure to high levels of natural radioactivity induces severe sequence-based and structural polymorphism on the Y chromosome.

Prolonged experimental irradiation and genome analyses is not ethically feasible for human beings. We exploited a natural setting in South Indian state of Kerala where the level of natural radioactivity is very high due to 10% thorium phosphate monazite present in the beach sand (Pask et al. 2000). Our own study revealed that Y chromosome use gene duplications, sequence polymorphisms, and structural alterations to buffer genotoxic effects of natural radioactivity (Premi et al. 2009).

As described above, absence of a homologue, repetitive landscape, and ICHR make human Y chromosome highly vulnerable to polymorphisms and sequence alterations (Crow 2000; Jehan et al. 2007). Since Y does not undergo homologous recombination, such polymorphisms and variations are passed onto the next generation through paternal inheritance. Since Y chromosome decides the sex, the lethal and clinical mutations or chromosomal anomalies are never inherited. Our analyses of the human Y chromosome from radiation-exposed males also uncovered that most of the microdeletions, sequence polymorphisms, and gene duplications did not follow a normal inheritance pattern. However, the radiation exposure did not affect the germline DNA (from sperms). This leads to the hypothesis that owing to its proneness, Y chromosome absorbs radiation genotoxicity in the form of somatic alterations, whereas germline remains unaffected so that the continuation of the species is warranted. An alternate expiation however is that meiotic germ cells have a high turnover and this removes any sick cells before they can transform into active spermatozoa. Nevertheless, a detailed analysis of various cellular pathways responsible for radiation genotoxicity is envisaged to uncover a full extent of damage induced by radiation exposure and its impact on Y chromosome physiology. Some of these pathways are tumor suppression, apoptosis, genome imprinting, epigenetic modification, and controls of signal transduction.

The radiation exposure also enhanced the polymorphism by inducing gene duplications. Sequence analyses revealed that of all the copies, sequence of at least one remains unaffected. This suggests the buffering effect of Y chromosome explained above so that radiation genotoxicity can be neutralized. Some examples of such polymorphic multi-copy genes are *SRY* and *CDY1*. In addition to gene duplication, radiation exposure also enhanced the expression of Y-linked genes in white blood cells. Based upon this, it is logical to conclude that radiation exposure modifies the transcriptional machinery so that it either does not differentiate between polymorphic copies or enhances the expression of the only normal gene copy. Both scenarios ensure sufficient supply of the transcripts from affected genes so that a normal function can still be maintained. Thus, exogenous factors like radiation exposure enhance the preexisting structural and sequence polymorphism on Y chromosome. However, the palindromes and segmental repeats protect the Y chromosomes by buffering and neutralizing the genotoxicity.

A complete map of all possible Y polymorphism is difficult to be generated. However, an overview of its structural integrity across various haplotypes is absolutely essential before drawing any clinical conclusions. Preliminary attempts have been made along these lines. In 2006, Jobling et al. analyzed Y chromosome from each of the 47 branches of human genealogy (Jobling et al. 2006). This analyses revealed an expected variation in the length of Yq heterochromatin and megabase rearrangements of the AZFc region.

7.6 Disintegrating Y Chromosome

Lack of a homologous chromosome is one of the major reasons for evolutionary degradation of the Y chromosome. It is speculated that genetic modifications lead to a steady decrease in recombinational frequencies and ultimately to cessation of homologous recombination (Brooks 2000). In addition, it is also believed that Y chromosome acquired palindromic repeats which induced inversions and similar megabase rearrangements which then inhibited any homologous recombination between X and Y (Premi et al. 2009). Loss of recombination caused Y degradation because there is no selection against deleterious mutations. Owing to these facts, it is widely believed that Y chromosome may disappear completely in about 5–10 million years (Lin et al. 2005). However, this hypothesis has been countered by several previous reports and our own investigations.

First objection to the proposed ongoing degradation of Y chromosome is the gene copy number. Modern methods like genome-wide arrays and genome hybridizations have revealed a large quantity of copy number polymorphisms (CNPs) which were either unknown or unexpected (Bailey 2002; Sebat 2004). CNPs are frequently found in the genic regions, especially the ones with segmental duplications. Many of which are implicated with resistance or susceptibility to a disease, responsiveness for a particular drug, or age-related ailments (Charlesworth and Hartl 1978). The *SRY* is long known to be a single copy gene. However, in our own investigation, we established multiple polymorphic copies of this gene in response to the radiation exposure and also in males with sex chromosome-related anomalies (Premi 2006; Premi et al. 2007, 2008, 2009, 2010). Some Turner's syndrome patients had as many as 16 copies of the *SRY* which suggested multiple rounds of duplication assuming that originally there was only a single copy. Absence of gene duplication in their parents meant de novo events of gene duplication.

These patients did not inherit 16 copies from the fathers, but instead acquired them by de novo duplications. To highlight the role of polymorphism in ongoing existence of the Y chromosome, a summary of gene duplications and sequence variations from ~250 patients with sex chromosome-related anomalies is given in Table 7.1. In addition to microdeletions, the gene duplications were also

Table 7.1 Summary of gene duplication in 236 patients shown in %

	Mosaic		One duplication		Two duplications		Three duplications	
Patient	SRY (%)	DAZ (%)	SRY (%)	DAZ (%)	SRY (%)	DAZ (%)	DAZ (%)	SRY (%)
Turner's syndrome (15)	95	80	1	5	1	10	3	5
Y polysomy (1)	0	0	100	0	0	0	0	100
Swyer syndrome (25)	95	90	2	3	2	5	1	2
Oligospermia (50)	10	5	30	85	40	5	20	5
Azoospermia (200)	10	5	20	85	50	5	20	5

Figures in parenthesis represent number of patients analyzed in a given category

inconsistent between males of two generations. Once again, the germline was normal in both generations (Premi et al. 2007, 2008) reestablishing the buffering effect of palindromic repeats. Moreover, higher frequency of CNP of Y-linked genes compared to that of autosomal *CDYL* and *CDYL2* supported the alteration-prone nature of the Y chromosome. Thus, by virtue of being haploid and accessible to segmental duplication, Y chromosome seems to be affected more than any of the autosomes or X chromosome. The copy number polymorphism of Y-linked genes and palindromic regions is surprisingly high. However, this explains a continual genetic integrity and sustenance of the Y chromosome by tolerating endogenous and exogenous genetic pressures (Bailey 2002; Premi et al. 2008; Sebat 2004). Multiplicity of the Y-linked sequences is probably due to nondisjunction which is supposed to maintain symmetrical copy numbers. Our investigations did not uncover this symmetry between numbers of Y chromosomes in patients and corresponding copies of the Y-linked genes. This indicates a non-correlative relation between nondisjunction of the Y chromosome and copy number polymorphism, especially in the case of the Y chromosome. Such copy number variations lead to enhanced genomic complexity. This is particularly true for Y chromosome since all the duplicate copies do not follow corresponding expressional and mutational profile, within and between two generations (Marshall Graves 2000; Premi 2006). The copy number changes were never associated with translocations suggesting a tandem mode of gene duplication (Marshall Graves 2000; Premi et al. 2008).

Classically, gene duplication leads to a normal parental copy and one or more copies with modified or new functional properties. Another proposed model for copy number change suggests that duplications are either followed by functional divergence or by functional complementation of the original copy (Premi 2006; Premi et al. 2007). In our own investigations, we concluded that exogenous stresses like radiation exposure initiate the duplication events which are followed by multiallelic tandem duplications to neutralize radiation genotoxicity. Ultimately, such duplication events are envisaged to prevent the degradation of the Y chromosome (Table 7.2).

Table 7.2 Fate of Y chromosome linked genes/loci in patients with SCRA and males exposed to NBR

Changes detected	NBR exposed	SCRA patients
Microdeletions	High frequency in AZFc and AZFa than AZFb	Almost same frequency in composite AZF region
Gene duplications	One to two rounds of duplication Multiple polymorphic copies of the *SRY*, *DAZ*, *CDY1*, and *XKRY* genes	One to three rounds of duplication Multiple polymorphic copies of the *SRY*, *DAZ*, *CDY1*, and *XKRY* genes
ICHR	No recombination	No recombination
Genic deletions	No deletion except few males lacking *DBY* in blood DNA	Frequently lack *XKRY*, *VCY*, *CDY1*, *CDY2*, *GOLGA2LY*, *TTY4*, and *BPY2* genes in random combinations

ICHR Intrachromosomal homologous recombination, *SCRA* Sex chromosome related-anomalies, *NBR* Natural background radiation

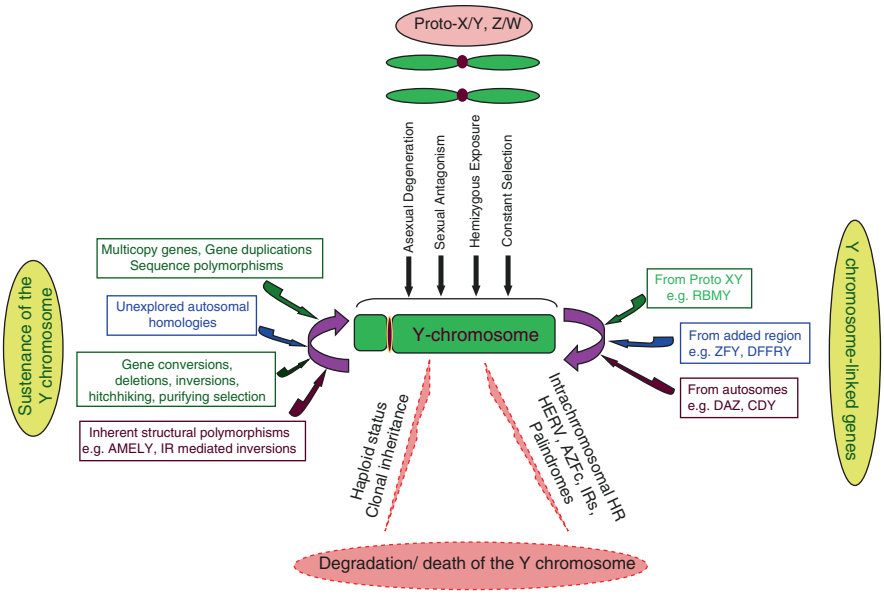

Fig. 7.4 Overall conclusion of evolution, degradation, and sustenance of human Y chromosome. It started with proto-XY/ZW chromosome which underwent various genetic/physical transformations to form modern day Y chromosome. These transformations include asexual degeneration, sexual antagonism, hemizygous exposure, and constant selection, all of which hugely reduced the physical size of Y chromosome. But along with these, the Y chromosome has been constantly supplemented by gene transfers from autosomes and its internal polymorphisms like copy number variations, palindromic and ampliconic duplications, gene conversions, etc. which helped it to sustain

Comparative sequence analyses between human and chimpanzee Y chromosome suggest that the Y has not undergone any degradation for the last ~6 million years (Hughes et al. 2005). This comparison also revealed that selective purification of sex determining genes is active on the Y chromosome. Intrachromosomal recombination further aids in maintaining the integrity of Y chromosome through gene conversions, duplications, deletions, inversions, purifying selection, and the hitchhiking of the sex determination factors. The role of such recombinational events in maintaining structural integrity of the Y chromosome (Sharp et al. 2005) is modeled in (Fig. 7.4). Thus, the proposed ongoing degradation and ultimate demise of the human Y chromosome is not supported by empirical data and may not be true at all.

Conclusions

Owing to its well-documented role in spermatogenesis, significance of the Y chromosome in sex determination has been strongly substantiated. However, exact number of Y-linked genes/loci and mechanisms controlling these phenomena are still obscure. Similarly, numbers of autosomal genes/loci implicated with testicular functions also remain a moot point. Thus, to be able to generate a consensus, current scenario warrants a comprehensive analysis of the global Y chromosomes both from

normal and abnormal genomes. This in turn is envisaged to help us undertake an accurate diagnosis of the individuals suffering from sex chromosome-related anomalies. This, however, cannot be achieved unless pathways of signal transduction and apoptosis leading to control and regulation of spermatogenesis are deciphered. Finally, simultaneous analysis of Y-linked genes, loci, and mRNA transcripts both from somatic tissue (blood) and germline (spermatozoa) would enable segregating those genes that are particularly active in the germline. Work on this line would form a rich and reliable basis of germline comparative genomics across the species.

A number of Y-based markers are currently being used extensively for paternity testing, particularly to ascertain the paternity of the male child. However, a number of still unexploited markers/marker systems of the Y chromosome are there that can be used both for paternity testing and DNA diagnosis. For instance, DYZ1 satellite in a single individual contains a total of 229 five-base-long repeat motifs "TTCCA" (Nakahori et al. 1986). This number is different in different individuals offering a rich source of innate sequence polymorphism to be exploited in the context of paternity. Likewise, a single array is about 3.4 kb, and since within the array there are regions that are highly variable, these regions can prove to be equally useful for unequivocal paternity testing based on PCR amplification, cloning, and sequencing. Systematic work along this line would have far reaching implications particularly, when a large number of males from across the different ethnic groups are screened. Even though comprehensive analyses have been performed on Y chromosome, there still lies a vacuum as far as its fine genetic prints and working reprints are concerned. We therefore propose screening of additional Y chromosomes for a given abnormality from across the globe. That would uncover innate, novel, and singular polymorphisms as well as the ones that can be correlated with abnormal phenotype. Expression studies of Y-linked genes from across the globe using samples representing normal and abnormal genomes may prove to be highly informative to pinpoint cause and effect of a disease on gene organization, regulation, and expression. Thus, there exists a scope to undertake analysis of the human Y chromosome for its large-scale biological applications and implications.

Competing Interests No competing interests to be disclosed.

References

Bailey JA. Recent segmental duplications in the human genome. Science. 2002;297(5583): 1003–7.

Bernstein R, Wadee A, Rosendorff J, Wessels A, Jenkins T. Inverted Y chromosome polymorphism in the Gujarati Muslim Indian population of South Africa. Hum Genet. 1986;74(3):223–9.

Bosch E. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. Hum Mol Genet. 2003;12(3):341–7.

Brooks R. Negative genetic correlation between male sexual attractiveness and survival. Nature. 2000;406(6791):67–70.

Brown GM, et al. Characterisation of the coding sequence and fine mapping of the human DFFRY gene and comparative expression analysis and mapping to the Sxrb interval of the mouse Y chromosome of the Dffry gene. Hum Mol Genet. 1998;7(1):97–107.
Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philos Trans R Soc Lond Ser B Biol Sci. 2000;355(1403):1563–72.
Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 1987;130(1):113–46.
Charlesworth B, Hartl DL. Population dynamics of the segregation distorter polymorphism of Drosophila melanogaster. Genetics. 1978;89(1):171–92.
Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet. 2000;1(1):40–7.
Delbridge ML, Lingenfelter PA, Disteche CM, Graves JAM. Nat Genet. 1999;22(3):223–4.
Delbridge ML, et al. A human candidate spermatogenesis gene, RBM1, is conserved and amplified on the marsupial Y chromosome. Nat Genet. 1997;15(2):131–6.
Dorus S. The CDY-related gene family: coordinated evolution in copy number, expression profile and protein sequence. Hum Mol Genet. 2003;12(14):1643–50.
Ferlin A. Human male infertility and Y chromosome deletions: role of the AZF-candidate genes DAZ, RBM and DFFRY. Hum Reprod. 1999;14(7):1710–6.
Ferlin A. The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. J Med Genet. 2003;40(1):18–24.
Fernandes S, et al. A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet. 2004;74(1):180–7.
Foresta C, Ferlin A, Moro E. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. Hum Mol Genet. 2000;9(8):1161–9.
Gavrilets S, Arnqvist G, Friberg U. The evolution of female mate choice by sexual conflict. Proc Biol Sci. 2001;268(1466):531–9.
Glas R. Cross-species chromosome painting between human and marsupial directly demonstrates the ancient region of the mammalian X. Mamm Genome. 1999;10(11):1115–6.
Graves JAM. The rise and fall of SRY. Trends Genet. 2002;18(5):259–64.
Harley VR, Clarkson MJ, Argentaro A. The molecular action and regulation of the testis-determining factors, SRY (Sex-determining region on the Y chromosome) and SOX9 [SRY-related high-mobility group (HMG) Box 9]. Endocr Rev. 2003;24(4):466–87.
Harley V, et al. DNA binding activity of recombinant SRY from normal males and XY females. Science. 1992;255(5043):453–6.
Hughes JF, et al. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature. 2005;437(7055):100–3.
Hurles ME, Jobling MA. A singular chromosome. Nat Genet. 2003;34(3):246–7.
Hurst LD, Randerson JP. An eXceptional chromosome. Trends Genet. 1999;15(10):383–5.
Jehan Z, et al. Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric CDC2L2. Genome Res. 2007;17(4):433–40.
Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet. 2003;4(8):598–612.
Jobling MA, et al. Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet. 2006;16(3):307–16.
Jobling MA, et al. Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet. 2007;16(3):307–16.
Kent-First M, et al. Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. Mol Reprod Dev. 1999;53(1):27–41.
Kobayashi K, et al. CORRIGENDUM. Hum Mol Genet. 1995;4(5):974.
Koopman P, Gubbay J, Collignon J, Lovell-Badge R. Zfy gene expression patterns are not compatible with a primary role in mouse sex determination. Nature. 1989;342(6252):940–2.

Krausz C. Y chromosome and male infertility. Front Biosci. 1999;4(1–3):e1.
Kuroda-Kawaguchi T, et al. The AZFc region of human Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet. 2001;29(3):279–86.
Lahn BT. Functional coherence of the human Y chromosome. Science. 1997;278(5338):675–80.
Lin Y-W, et al. Polymorphisms associated with the DAZ genes on the human Y chromosome. Genomics. 2005;86(4):431–8.
Mahadevaiah S. Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. Hum Mol Genet. 1998;7(4):715–27.
Marshall Graves JA. Human Y chromosome, sex determination, and spermatogenesis—a feminist view. Biol Reprod. 2000;63(3):667–76.
Mazeyrat S, Saut N, Mattei MG, Mitchell MJ. RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. Nat Genet. 1999;22(3):224–6.
Mitchell CL, Harley VR. Biochemical defects in eight SRY missense mutations causing XY gonadal dysgenesis. Mol Genet Metab. 2002;77(3):217–25.
Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 2001;17(9):481–5.
Nakahori Y, Mitani K, Yamada M, Nakagome Y. A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. Nucleic Acids Res. 1986;14(19):7569–80.
Nasrin N, et al. DNA-binding properties of the product of the testis-determining gene and a related protein. Nature. 1991;354(6351):317–20.
Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. Nat Genet. 2001;29(1):20–1.
Pask AJ, Harry JL, Renfree MB, Marshall Graves JA. Absence of SOX3 in the developing marsupial gonad is not consistent with a conserved role in mammalian sex determination. Genesis. 2000;27(4):145–52.
Pirkkala L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. FASEB J. 2001;15(7):1118–31.
Pontiggia A, Whitfield S, Goodfellow PN, Lovell-Badge R, Bianchi ME. Evolutionary conservation in the DNA-binding and -bending properties of HMG-boxes from SRY proteins of primates. Gene. 1995;154(2):277–80.
Premi S. Tandem duplication and copy number polymorphism of the SRY gene in patients with sex chromosome anomalies and males exposed to natural background radiation. Mol Hum Reprod. 2006;12(2):113–21.
Premi S, Srivastava J, Chandy SP, Ali S. AZFc somatic microdeletions and copy number polymorphism of the DAZ genes in human males exposed to natural background radiation. Hum Genet. 2007;121(3–4):337–46.
Premi S, Srivastava J, Chandy SP, Ali S. Unique signatures of natural background radiation on human Y chromosomes from Kerala, India. PLoS One. 2009;4(2):e4541.
Premi S, Srivastava J, Epplen JT, Ali S. AZFc region of the Y chromosome shows singular structural organization. Chromosom Res. 2010;18(4):419–30.
Premi S, Srivastava J, Panneer G, Ali S. Startling mosaicism of the Y-chromosome and tandem duplication of the SRY and DAZ genes in patients with Turner Syndrome. PLoS One. 2008;3(11):e3796.
Prosser J, et al. Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. Mamm Genome. 1996;7(11):835–42.
Pryor JL, et al. Microdeletions in the Y chromosome of infertile men. N Engl J Med. 1997;336(8):534–40.
Reijo R, et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet. 1995;10(4):383–93.
Repping S, et al. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet. 2002;71(4):906–22.

Repping S, et al. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet. 2003;35(3):247–51.
Santos FR, Pandya A, Tyler-Smith C. Reliability of DNA-based sex tests. Nat Genet. 1998;18(2):103.
Saxena R, et al. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet. 1996;14(3):292–9.
Saxena R, et al. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. Genomics. 2000;67(3):256–67.
Schmid M, et al. Satellited Y chromosomes: structure, origin, and clinical significance. Hum Genet. 1984;67(1):72–85.
Schmitt-Ney M, et al. Two novel SRY missense mutations reducing DNA binding identified in XY females and their mosaic fathers. Am J Hum Genet. 1995;56(4):862–9.
Sebat J. Large-scale copy number polymorphism in the human genome. Science. 2004;305(5683):525–8.
Sharp AJ, et al. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet. 2005;77(1):78–88.
Shen JHC, Ingraham HA. Regulation of the orphan nuclear receptor steroidogenic factor 1 by Sox proteins. Mol Endocrinol. 2002;16(3):529–40.
Skaletsky H, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 2003;423(6942):825–37.
Steinemann M, Steinmann S. Common mechanisms of Y chromosome evolution. Genetica. 2000;109(1/2):105–11.
Sun C, et al. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet. 2000;9(15):2291–6.
Tessari A. Characterization of HSFY, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis. Mol Hum Reprod. 2004;10(4):253–8.
Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. Hum Genet. 1976;34(2):119–24.
Ting C. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science. 1998;282(5393):1501–4.
Vogt P. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet. 1996;5(7):933–43.
Vogt PH. Report of the third international workshop on human Y Chromosome mapping 1997. Cytogenet Genome Res. 1997;79(1–2):1–20.
Vogt PH. Genomic heterogeneity and instability of the AZF locus on the human Y chromosome. Mol Cell Endocrinol. 2004;224(1–2):1–9.
Vollrath D, et al. The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science. 1992;258(5079):52–9.
Waters PD, Duffy B, Frost CJ, Delbridge ML, Graves JAM. The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80–130 million years ago. Cytogenet Genome Res. 2001;92(1–2):74–9.
Waters PD, et al. Autosomal location of genes from the conserved mammalian X in the platypus (*Ornithorhynchus anatinus*): implications for mammalian sex chromosome evolution. Chromosom Res. 2005;13(4):401–10.
Wolfenbarger LL, Wilkinson GS. Sex-linked expression of a sexually selected trait in the stalk-eyed fly, Cyrtodiopsis dalmanni. Evolution. 2001;55(1):103–10.
Wong J, Blanco P, Affara NA. An exon map of the AZFc male infertility region of the human Y chromosome. Mamm Genome. 1999;10(1):57–61.
Wu C-I, Wyckoff GJ, Wang W. Rapid evolution of male reproductive genes in the descent of man. Nature. 2000;403(6767):304–9.
Yen PH. A Long-range restriction map of deletion interval 6 of the human Y chromosome: a region frequently deleted in azoospermic males. Genomics. 1998;54(1):5–12.

Dr. Sanjay Premi Dr. Premi is a proficiently trained molecular biologist with eminent experience in human genetics, biochemistry, photobiology, and environmental carcinogens. During his PhD, he pioneered in establishing a buffering effect of chromosomal alterations in offsetting the genotoxic effects of natural radioactivity. In his postdoctoral training, he discovered carcinogenicity of the pigment melanin which was completely unanticipated. He and his colleagues termed it melanin chemiexcitation which generated pyrimidine dimers in complete absence of UV radiation exposure.

Dr. Jyoti Srivastava Dr. Srivastava is a distinguished cancer biologist with an impressive track record. She is expert in Molecular & Cellular Biology, Biochemistry, and Genomics,with a specific focus on translational medicine. She has published about 30 peer-reviewed publications, scientific reviews, and book-chapters, and received awards and recognition from respectable organizations for scientific achievements.

Human Microbiome: Implications on Health and Disease

8

Princy Hira, Utkarsh Sood, Vipin Gupta, Namita Nayyar, Nitish Kumar Mahato, Yogendra Singh, Rup Lal, and Mallikarjun Shakarad

8.1 Introduction

> The way to a man's heart is through his stomach.
>
> John Adams, 1814

The state of our gut not only governs the health of our body but also our mental health and emotional status. Although, this notion is more than a century old and sporadic interest was visible from the 1970s, serious research studies on gut microbiome and its implications on our health have just begun (Schmidt 2015). The human body consists of about 40 trillion cells (Bianconi et al. 2013) with about 22,000 human genes in each cell (Pertea and Salzberg 2010). However, with the association of microbes immediately after birth, the human body contains about 100 trillion cells and more than 2 million genes. The microbiota that gets associated with the human body makes up about 1–3% of the human body mass amounting to 2–6 pounds of microorganisms in a 200-pound adult (Turnbaugh et al. 2007, HMP 2007–2012). The additional cells as mentioned above are the microorganisms that, apart from the gut, also reside on the skin surface, in the deep skin layers, in the mouth, digestive tract and other human organ systems. The sum total of microorganisms that colonize the human body are collectively referred to as 'human microbiome or human microbiota'. The microbiome is central to human biology (Schnorr 2015). With so much of microbiota getting associated with the human body, it

Princy Hira and Utkarsh Sood contributed equally with all other contributors.

P. Hira • U. Sood • V. Gupta • N. Nayyar • N.K. Mahato • R. Lal (✉) • M. Shakarad (✉)
Department of Zoology, University of Delhi, Delhi 110007, India
e-mail: ruplal@gmail.com; beelab.ms@gmail.com

Y. Singh
CSIR Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India

L. Rawal, S. Ali (eds.), *Genome Analysis and Human Health*,
DOI 10.1007/978-981-10-4298-0_8

became imperative to understand the role of microbes that colonize the human body to fully understand and appreciate the human physiology and behaviour under healthy and diseased conditions. With the progress of research in this field, it is proposed that better understanding of human microbiome would pave the way for successful treatment of not only lifestyle diseases but also life-threatening diseases as well as non-genetic behavioural disorders. With the completion of phase 1 of the Human Microbiome Project (HMP 2008–2012) and the researches that have been carried out subsequently, it has become clear that the human microbiome is associated with obesity, cancer, mental health disorders, asthma and autism. While many other aspects of these associations are yet to be investigated, we are not clear whether the differential microbiome composition among the diseased individuals is a consequence of the disease itself or the differing microbiome causes the disease.

8.2 Inception of Microbiota with Human Body

It will be pertinent to mention here that the human foetus grows in an absolutely sterile environment of the uterus for about 266 days from the time of conception till parturition. The first encounter of human with the microbes is during the passage of the infant through the birth canal, specifically the vaginal tract and the vulva (Ravel et al. 2011). Perhaps skin surface microbes are the first colonizers followed by nasal (respiratory) and oral (digestive) tracts, brought about by the processes of breathing and external feeding. Establishment of an unwavering flora on the skin, oral cavity and intestinal tract occurs with handling and feeding of the foetus within the first 48 h. Mode of birth whether normal or caesarean also suggested to influence the microbial colonization of the human infant to a greater extent (Mackie et al. 1999; Dominguez-Bello et al. 2010).

8.3 Diversity of Human Microbiota

Recent investigations revealed that the human microflora is exceedingly intricate and includes more than 200 species of bacteria (Todar 2012). Various factors like genetics, age, sex, stress, nutrition and dietary habit of the individuals greatly influence the diversity and abundance of microflora. The estimated number of bacteria present on the human skin, inside the mouth and the gastrointestinal tract, is 10^{14}, 10^{10} and 10^{14}, respectively (Mikelsaar and Zilmer 2009). The number of bacteria in the human gut alone far exceeds the total number of human cells (Gerritsen et al. 2011). The digestive system alone accounts for 55% of the total human microbiota, followed by skin, respiratory system and urogenital system. Surprisingly, blood contains just about 1% of the total human microbiota, while the conjunctiva has negligible quantity of microbiota (Table 8.1). The microbiota of the human intestine

Table 8.1 Microbiota (prevalent genera) that colonize different human organ systems

S. no.	Body niche	No. of bacteria (as a % of the total microbiota in humans)	Prevalent genus
1	Skin	21	*Staphylococcus*, *Propionibacterium* and *Corynebacterium*
2	Gut	29	*Bacteroides*, *Clostridium*, *Fusobacterium*, *Enterococcus*, *Eubacterium*, *Ruminococcus*, *Peptococcus*, *Peptostreptococcus*, *Bifidobacterium*, *Escherichia* and *Lactobacillus*
3	Oral cavity	26	*Streptococci*, *Lactobacilli*, *Staphylococci*, *Corynebacterium* and *Bacteroides*
4	Vagina (urinogenital)	9	*Lactobacillus*, *Atopobium*, *Peptostreptococcus* and *Staphylococcus*
5	Conjunctiva	0	*Staphylococcus*, *Propionibacterium* and *Haemophilus*
6	Respiratory region	14	*Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus and Streptococcus*
7	Blood	1	*Staphylococcus*

Instability of human microbiome (Adopted from Peterson et al. 2009 (NIH Human Microbiome Project))

is suggested to not only help in digestion, produce vitamins and promote gastrointestinal motility but balance the immune system as well (Berg 1996), suggesting the larger implications on human health and diseases. The disturbance of microbiota–host relationship is associated with numerous chronic inflammatory diseases and metabolic syndrome (Chassaing et al. 2015).

8.4 Human Microbiome and Human Health

In order to better understand the impact of the human microbiome on human health and diseases, it is important to understand not only the microbial density/load but also to know the diversity of microbes colonizing different organ systems. Among the different organ systems that were assessed for the microbial diversity, the gut was found to have the highest diversity followed by the mouth and skin. Vaginal region had the least microbial diversity (Li et al. 2012). The highly diverse microflora of the digestive system are perhaps due to variable food habits of individuals, while the diversity of the skin microbiota might be related to the geographical differences (Kau et al. 2011). Shannon–Wiener index

(H′), Simpson's index (D), Brillouin index (HB), richness and evenness are some of the standard diversity measures adopted in elucidating community diversity (Zar 2010). Each of these diversity indices has their own limitations and advantages and is often used in combination for a better understanding. Even the use of these indices in combinations falls short of expectations when one has to understand the microbial communities across human body habitats, specifically the failure to capture low abundant taxa. Tail statistic, τ—a rank-based diversity measure that is similar to standard deviation statistic, σ—is suggested to best suit the 16S profiles that tend to exhibit a long-tailed distribution (Li et al. 2012).

The microbial community colonizing a healthy human body is dominated by four major phyla, viz. *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria*. At the genus level, the most predominant genera are *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Eubacterium*, *Fusobacterium*, *Peptococcus*, *Peptostreptococcus* and *Ruminococcus*. *Bacteroides* alone constitutes about 30% of the total gut microflora (Sears 2005). *Escherichia* and *Lactobacillus* are the other two genera present to a lesser extent (Khanna and Tosh 2014). Apart from Bacteria, Archaea and Fungi are the other group of microorganisms that are found in variable numbers in the human body. The common fungi include species of the genera *Candida*, *Saccharomyces*, *Aspergillus* and *Penicillium* (Hoffmann et al. 2013).

In addition to the characteristic and systematic differences in the microbial diversity in different human body habitats, differences among individuals were also reported (Li et al. 2012). It is now becoming clear that the microbial community differences among individuals hold a key to human health, diseases and treatment. Introduction (through dietary change) and/or extinction (due to antibiotic treatment) of particular microbial groups would alter the community and population structure of the microbiota that potentially bring about functional variation.

The developments of new sequencing technologies, computational algorithms and bioinformatic tools have made the exploration of the human microbiome a frontier enterprise. The main focus in the recent past has been to elucidate the 'core' microbiome occupying specific human body niches and to ascertain interindividual differences of healthy humans. However, it is critically important to discern the differences between healthy and diseased individuals. Although association of specific microbial communities under physiologically different conditions of healthy subjects was being sporadically reported from the late 1970s, there is a steady stream of publications reporting the microbial communities predominant in human subjects affected by different diseases. In this chapter we have restricted our discussion to the diseases manifested as a consequence of altered gut microbial community structure (Fig. 8.1) due to medication. The specific diseases are *Clostridium difficile* infection (CDI), autism spectrum disorder (ASD), diabetes, gastric cancer, obesity and inflammatory bowel disease (IBD).

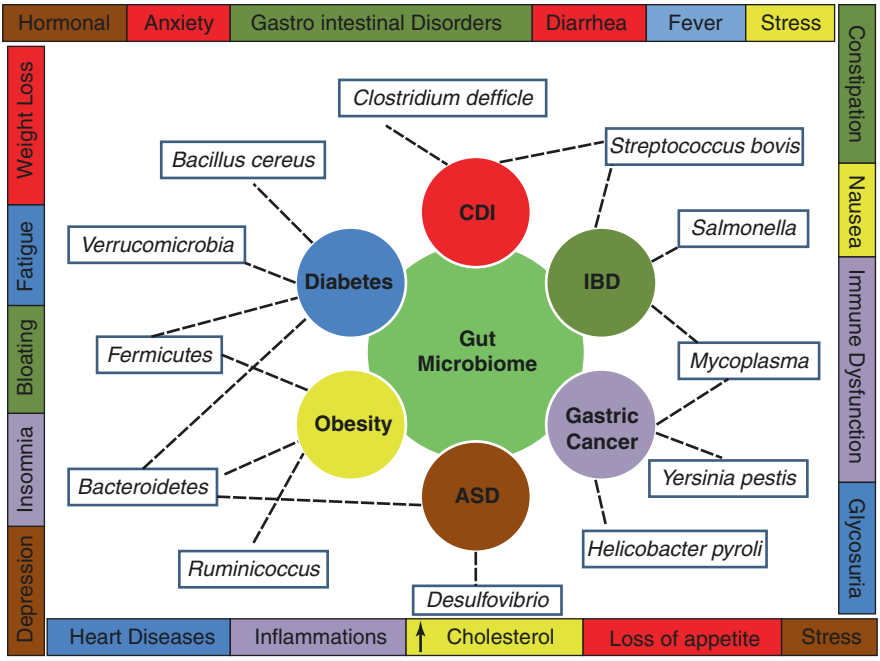

Fig. 8.1 Complex interplay between gut microbes, diseases and their symptoms. Colour of each microbial-associated disease corresponds to its coloured phenotype. The *dotted line* between the causative microbial lineages represents the cross talk through nonlinear signalling interdependence

8.5 Microbiome and *Clostridium difficile* Infection (CDI)

Clostridium difficile is a pathogenic native gut microorganism, found in 3 out of 100 adults and 7 out of 10 babies. In healthy individuals the population numbers of *C. difficile* are maintained at negligible level that is insufficient to cause disease. However, with administration of broad-spectrum antibiotics, the patients develop gastrointestinal illness, due to a toxin produced by *C. difficile* (Buss et al. 2015). The disease is referred to as *C. difficile* infection (CDI). Although our knowledge of CDI pathogenesis is still rudimentary (Britton and Young 2012), CDI is one of the most ubiquitous and expensive nosocomial infections. CDI occurs in 25% of all antibiotic-associated diarrhoea (Bartlett 2002). Another disease also called the antibiotic-associated diarrhoea is reported to coincide with the decline in the carbohydrate-fermenting butyrate-producing members of the phylum *Firmicutes* (Britton and Young 2012). Further, it has been shown that even short-term antibiotic treatment can bring about long-term changes in gut microbiota that is not necessarily reversible with the discontinuation of antibiotic treatment (Jakobsson

et al. 2010). The reduced microbial diversity would not only lead to invasion and proliferation of pathogenic flora due to lowered resistance (Chang et al. 2008; Britton and Young 2012) but is responsible for the progression of the disease (Freter 1955).

Over the past decade, increased morbidity and mortality, as well as relapse of *C. difficile* infection, have become more common (Khanna et al. 2012) due to the emergence of strain 027 of *C. difficile* (Karas et al. 2010; Marsh et al. 2012). Antibiotic resistance, sporulation ability and toxin production are suggested to be the potential contributors to virulence of historical ribotypes and *C. difficile 027* (Warny et al. 2005; Drudy et al. 2007; Merrigan et al. 2010; Lanis et al. 2010, 2012, 2013). TcdA and TcdB are two large clostridial toxins produced by *C. difficile* responsible for major virulence causing extensive tissue damage in human disease (Taylor et al. 1981; Libby et al. 1982; Lyerly et al. 1986). Among the two toxins, TcdB is the critical virulence factor (Lyras et al. 2009), antigenically variable and more lethal and causes more extensive brain haemorrhage (Lanis et al. 2013).

Due to ever-increasing severity of CDI, many studies have been initiated to unravel the details of the disease progression that perhaps would aid in designing effective treatment. In humans, bile acids are secreted in the small intestine in response to consumption of food so as to facilitate absorption of fats and fat-soluble vitamins and nutrients (Britton and Young 2012). Cholate and chenodeoxycholate are the primary bile acids that are conjugated to either of the two amino acids—glycine and taurine (Ridlon et al. 2006). Deoxycholate, a secondary bile acid produced by the action of 7-dehydroxylase on cholate, was reported to be a potent *C. difficile* spore germinant but highly toxic to its vegetative cells. Further, bile acid (taurocholate) and amino acid (glycine) were shown to enhance *C. difficile* spore germination by 1000-fold (Sorg and Sonenshein 2008). Antibiotic treatment perhaps reduced members of microbiota that were involved in the conversion of cholate to deoxycholate, thus resulting in increased levels of cholates and their derivatives. This in turn facilitates the germination of spores and growth and propagation of vegetative cells of *C. difficile* (Britton and Young 2012). However, chenodeoxycholate is shown to inhibit spore germination (Sorg and Sonenshein 2008), and hence non-metabolizable derivates of chenodeoxycholate could serve as therapeutics (Sorg and Sonenshein 2010). Competitive exclusion of toxigenic *C. difficile* by non-toxigenic *C. difficile* (Sambol et al. 2002), direct antagonism by intestinal microbiota such as *Bacillus thuringiensis* that secretes thuricin CD bacteriocin with narrow-spectrum activity against *C. difficile* spores (Rea et al. 2010) and faecal transplantation (Khoruts et al. 2010) are suggested as potential curative measures.

8.6 Autism Spectrum Disorders (ASD) and Gut Microbiome

The gut microbes are now reported to make neuroactive compounds, including neurotransmitters and metabolites that act on brain via the vagus nerve that connects the brain and the digestive tract (Schmidt 2015). Disruptions of the healthy

microbiome are suggested to result in anxiety, depression and even autism. Autism spectrum disorders (ASD) are complex neurobiological disorders characterized by stereotyped behavioural patterns leading to visible impairment in social interactions and communications (Johnson and Myers 2007). Both genetic and environmental factors play an important role in ASD aetiology. Genetically, ASD is linked with autosomal recessive inheritance, X-linked inheritance and sporadic chromosomal anomalies. Among the environmental factors, gut microbes have the potential to interact with central nervous system (Collins and Bercik 2009). Autistic children's gut had reduced bacterial richness compared to neurotypical children. Altered gut microbiota was not due to demographics or special diets but due to antibiotic treatment that is suggested to aggravate ASD-related behavioural symptoms (Kang et al. 2013). High levels of gram-negative bacteria *Bacteroides vulgatus* and *Desulfovibrio* have been reported in autistic children (Finegold et al. 2010). Lipopolysaccharides (LPS) present in the cell walls of many pathogenic gram-negative bacteria are suggested to damage many tissues including the brain (Minami et al. 2007) leading to increased permeability of the blood–brain barrier, thus facilitating the accumulation of high levels of mercury in the cerebrum that may aggravate ASD symptoms (Adams et al. 2008). Glutathione—an important antioxidant responsible for heavy metal detoxification in the brain—has been shown to be reduced in rats exposed to LPS (Zhu et al. 2007). Depletion of glutathione could also be caused by p-cresol—formation of which is catalysed by a glycyl radical enzyme (p-hydroxyphenylacetate decarboxylase) from *C. difficile*, a gram-positive bacteria (Selmer and Andrei 2001). As discussed above *C. difficile* is known to play a crucial role in development of gastrointestinal illness (GI). Thus the presence of autistic symptoms and their correlated GI severity seems to be linked to reduced richness and diversity of gut microflora that in turn might alter the physiological functionality and microbial GI robustness due to decrease in microbial redundancy in ASD children (Kang et al. 2013). Although a statistically significant correlation between autistic symptoms and abundances of unclassified *Veillonellaceae, Prevotella* and *Coprococcus* genera is established, severity of GI symptoms is not a significant predictor of these microbial changes among autistic children (Kang et al. 2013). ASD children are reported to have a strong preference for starches, snack and processed foods while rejecting most fruits, vegetables and proteins (Field et al. 2003; Sharp et al. 2013). Although the aetiological factors contributing to feeding problems in ASD patients remain elusive (Mulle et al. 2013), neurobehaviourally influenced aetiology of higher rates of constipation and encopresis is reported in ASD (Ibrahim et al. 2009). The major function of gut microbiome of healthy individuals is to help in breaking down complex plant polysaccharides and other dietary matter. The altered gut microbiome of the ASD patients reported is unable to assist in the breakdown of the plant polysaccharides thus causing GI distress (Mulle et al. 2013). Hence, interventions aimed at restoring the microbial balance in the gut of ASD individuals might improve behaviours (Mulle et al. 2013).

8.7 Gut Microbiome, Obesity and Diabetes

The relation between obesity and gut microbiota was known as early as three decades ago, and the gut microbiota is shown to shift in response to host adiposity and nutrient intake (Musso et al. 2011). Several studies have suggested the involvement of gut microbiota in host metabolism, energy utilization and storage (Musso et al. 2011) leading to the development of fat mass and fat storage (Backhed et al. 2004; Everard and Cani 2013). *Bacteroides intestinalis*, *Bacteroides fragilis* and *Escherichia coli* are suggested to be involved in generation of secondary bile acids in the colon (Fukiya et al. 2009), and bile acids are known to exert metabolic regulatory functions in addition to favouring dietary lipid absorption (Keitel et al. 2008; Lefebvre et al. 2009). The development of obesity was found to be associated with the enrichment of *Firmicutes*—specifically *Mollicutes*—at the expense of *Bacteroidetes* in mice fed with high-fat/high-sugar diet compared to those fed with low-fat/high-polysaccharide diet (Turnbaugh et al. 2008). The microbiome of the obese mice showed enrichment in genes coding for enzymes that enable the extraction of energy from otherwise indigestible alimentary polysaccharides suggesting increased energy extraction capacity of the gut flora of obese individuals (Turnbaugh et al. 2006; Musso et al. 2011). Further, gut microbiota is shown to play a major role in the onset of insulin resistance and type 2 diabetes (Bäckhed et al. 2004, 2007; Cani et al. 2007a; Shen et al. 2013) triggering low-grade inflammation—a common feature characterizing obesity and several other metabolic disorders (Everard and Cani 2013). Microbiota-derived lipopolysaccharides (LPS) are reported to be the key molecule involved in early development of inflammation and metabolic diseases (Cani et al. 2007b). Animal model studies have established that obesity is transmissible along with gut microbiota (Musso et al. 2011) as transplantation of microbiota from obese mice to germ-free wild-type recipient mice resulted in increased adiposity compared to those that received microbiota from conventionally raised lean wild-type littermates (Turnbaugh et al. 2006).

Diet also plays an important role in changing the microbial diversity of gut microbiome. High-fat diet when given to both obese and lean genotypes was found to be associated with a decrease in *Bacteroidetes* and an increase in both *Firmicutes* and *Proteobacteria* (Hildebrandt et al. 2009; Turnbaugh et al. 2009). On the other hand, germ-free mice were found to be resistant to diet-induced obesity caused by consumption of a high-fat or high-sugar 'Western' diet (Backhed et al. 2007). A study by Ley et al. (2005) clearly demonstrated that both—genetic obese and diet-induced obese—had increased abundance of *Firmicutes* in their gut microbiome.

While type 2 diabetes is a metabolic disorder caused due to obesity-linked insulin resistance, type 1 diabetes (T1D) is a T-cell-mediated autoimmune disease due to slow and progressive destruction of insulin-producing β cells (Zipris 2008). Both genetic and environmental factors are known to contribute to autoimmunity disorders. Altered gut microbiota, impaired intestinal mucosal barrier and mucosal immunity are reported to contribute to T1D pathogenesis (Musso et al. 2011). Although specific details of how the gut microbiota regulates the T1D are unknown, T1D-resistant MyD88 KO mice were shown to harbour a lower *Firmicutes*/*Bacteroidetes*

ratio with an increased proportion of *Lactobacilli*, *Rikenellae* and *Porphyromonadaeae* (Wen et al. 2008). The dynamic link between gut microbiota, adiposity and diabetes indicates that manipulation of gut microbial communities by dietary interventions (e.g. probiotics or prebiotics) and translocation could be an approach to treat obesity and improve metabolic health (Flint et al. 2014).

8.8 Gut Microbiome and Inflammatory Bowel Disease (IBD)

Inflammatory bowel disease (IBD) involves chronic and recurring immune responses with relapsing and remitting inflammations in gastrointestinal tract. Aetiology is multifarious including genetic, microbial and environmental factors contributing to disease development (Cho and Blaser 2012). IBD primarily includes two subtypes, namely, ulcerative colitis (UC) and Crohn's disease (CD). UC remains confined to the colon and rectum, while CD can affect different areas of GI tract including the mouth. These are characterized as autoimmune diseases with the identification of pathways involving NOD2, ROS, CARD9 and Th17 cells in genetically susceptible hosts (Cho and Blaser 2012). Genetic predisposition is in itself not sufficient for the onset and progression of inflammation. Microbial dysbiosis plays a key role in the onset and progression of IBD, indicating the complex interplay between the gut microbiome and genetic susceptibility to IBD (Knights et al. 2013).

Microbial dysbiosis refers to the shift in relative abundances of dominant taxa and decrease in overall diversity of gut community (Sokol and Seksik 2010). It remains unclear whether this dysbiosis is the cause of or the response to the disease; nevertheless stable and healthy gut commensal bacteria are necessary to suppress the pathogenic infection (Kamada et al. 2012). Broadly, IBD is associated with reduced gut diversity, an increase in proportion of *Gammaproteobacteria* and reduced number of *Firmicutes* (Sokol and Seksik 2010). A significant decrease in abundance of two genera *Roseburia* and *Phascolarctobacterium* is associated with both UC and CD subjects (Morgan et al. 2012). In gut, species of the genus *Roseburia* are associated with production of butyrate and utilization of acetate (Duncan et al. 2002), whereas species of the genus *Phascolarctobacterium* are associated with production of propionate in coculture with *Paraprevotella* (Watanabe et al. 2012). Therefore, apart from changes in composition, functional imbalance has also been witnessed in IBD subjects including upregulation of sulphur metabolism pathways and downregulation of butanoate and propanoate metabolism. Few microbial clades are differentially abundant in CD and UC patients; proportion of *Faecalibacterium* of *Ruminococcaceae* family (acetate producers) is reduced, and members of the family *Enterobacteriaceae* show significant increase in abundance in CD (Kang et al. 2010), whereas a significant reduction in members of *Leuconostocaceae* is seen in UC (Morgan et al. 2012).

Epidemiological studies on concordance rates for IBD in German monozygotic twins (16% for UC and about 35% for CD) suggest stronger genetic influence in CD as compared to UC and also indicate the role of environmental factors in the development of chronic inflammation (Spehlmann et al. 2008).

Dietary intake is also correlated with incidence of IBD. Diet with high amounts of total fats, PUFAs, omega-6 fatty acids and meat was associated with an increased risk of CD and UC, whereas high fibre and fruit intake were related to decreased risk for CD. High vegetable intake was linked with decreased risk for UC (Hou et al. 2011). Recently, blow-out of 'Western diet', rich in protein but low in fruits and vegetable, is also being considered as a reason for increasing IBD incidence.

Hence, there exists an interaction network between genetics, host gut microbiome and diet providing feedback to host immune responses. For instance weakened immune response to commensal bacteria in gut can result from mutations in *NOD*2 and *GPR*35 and, as a result, cause imbalance in taxonomic structure of gut microbiota which can subsequently lead to metabolic dysbiosis. Altered metabolic capabilities of gut microbiome may further lead to diminished antibacterial activity through different pathways and consequent taxonomic and metabolic imbalance (Knights et al. 2013). Recently, even the alterations in gut virome have been observed in IBD patients (Ray 2015).

Based on studies done so far, treatments used for IBD are accompanied with potential risks and side effects. However, use of probiotics and prebiotics with clinical course is being tested for its cure of which using *Faecalibacterium* as a probiotic is a promising strategy in counterbalancing the gut commensal bacteria composition in CD patients (Sokol et al. 2008). Symbiosis factors from microbes can also be employed in therapeutics for inflammation, for example, PSA (polysaccharide A) produced by *B. fragilis* is reported to suppress the production of interleukin-17 (pro-inflammatory) from intestinal immune cells (Mazmanian et al. 2008). Apart from these, researchers are trying faecal bacteriotherapy (FBT) in which faeces from healthy donor are transplanted into the gut as a treatment of UC, though it has not yet approved regulatory authorities.

8.9 Microbiome and Gastric Cancers

It is evident from the preceding discussion that the gut microbiota has significant influence on inflammation of the gut particularly the distal large intestine (Louis et al. 2014). The chronic inflammation of the gastrointestinal tract progresses to inflammatory bowel disease (IBD), and IBD patients are reported to show an increased incidence of colorectal cancer (CRC) also known as colitis-associated cancer (CAC) (Jess et al. 2005; Danese et al. 2011). More than 95% of CRC cases show an association with dietary lifestyle and more recently gut microbiota, while less than 5% are hereditary (Rustgi 2007; Watson and Collins 2010; Irrazábal et al. 2014). CRC is ranked third among the most common causes of cancer-related deaths in the world (AICR 2007; Jemal et al. 2011; Irrazábal et al. 2014). Several pathogenic bacteria have been implicated in promoting CRC via pro-inflammatory interactions with host cells (Sears and Garrett 2014; Zackular 2014; Zackular et al. 2014; Louis et al. 2014). Relative abundance of *Ruminococcaceae*, *Clostridium*, *Pseudomonas* and *Porphyromonadaceae* was higher, while the relative abundances

of *Bacteroides*, *Lachnospiraceae*, *Clostridiales* and *Clostridium* were found to be less in patients with adenomas (Zackular et al. 2014). Further, patients with carcinomas had higher relative abundances of *Fusobacterium*, *Porphyromonas*, *Lachnospiraceae* and *Enterobacteriaceae* and lower abundances of *Bacteroides*, *Lachnospiraceae* and *Clostridiales* (Zackular et al. 2014). Furthermore, *Helicobacter pylorus* has been identified as the primary cause of gastric cancer (Tu et al. 2008). However, it is now becoming increasingly clear that collective activities of the metabolic products of the microbiota greatly influence the predisposition to and protection against CRC (Gill and Rowland 2002; Schwabe and Jobin 2013). Nitrosation of amines produced by fermentation of proteins in the large intestine by *Bacteroides* and *Firmicutes* leads to formation of N-nitroso compounds that have the potential to promote cancer (Rowland 2000; Louis et al. 2014) as indicated by the positive correlation between dietary intake of NOCs and CRC in European populations (Loh et al. 2011). Nitroreductases and nitrate reductases encoded by *Proteobacteria* are suggested to be facilitating nitrosation (Louis et al. 2014). Ammonia—a product of protein fermentation—is reported to be potentially carcinogenic at low concentrations (Windey et al. 2012). Although polyamines are essential for maintenance of structural integrity of membranes and nucleic acids, higher levels of polyamines are associated with several diseases including cancer (Louis et al. 2014) and certain gut bacteria including enterotoxigenic *B. fragilis* that upregulate polyamine production (Pegg 2013). Further, pathogens such as *Shigella flexneri*, *Streptococcus pneumoniae*, *Salmonella enterica* and *H. pylori* are known to exploit polyamines to increase their virulence (Di Martino et al. 2013). Colonocyte barrier breakdown by toxic sulphide produced as hydrogen sulphide in the gut by sulphate-reducing bacteria related to *Disulfovibrio* spp. could be another causative agent of CRC as indicated by higher stool sulphide levels in CRC patients, although increased levels of *Disulfovibrio* spp. have not been reported (Carbonero et al. 2012). However, several bacterial pathogens such as *B. fragilis*, *E. coli* NC101 strain, *Fusobacterium* spp. and *Campylobacter* spp. seem to be directly and specifically involved in promoting CRC (Sears 2009; Arthur et al. 2012; Kostic et al. 2013). Further, there is a complex interplay between diet, bile acid and gut microbiota (Louis et al. 2014). Higher-fat intake is positively correlated with secondary bile acids (Ou et al. 2013), and secondary bile acid deoxycholic acid is reported to promote liver cancer (Yoshimoto et al. 2013). Furthermore, higher levels of bile acids are reported from faecal samples of CRC patients (Ou et al. 2012).

Both animal and human studies suggest that dietary supplementation with non-digestible carbohydrates can reduce protein fermentation in the large intestine, leading to decrease in the genotoxicity of faecal water (Windey et al. 2012), thus reducing the incidence of IBD as well as CRC.

Conclusions

We owe our very persistence in nature to the plethora of the microbiota that has colonized our various organ systems. Particularly, the gut microbiota provides important benefits in terms of primary breakdown of the food ingested, immune development as well as mental wellbeing. However, the full import of the role of

human microbiome on the health as well as disease has just begun to emerge with the advent of culture independent research technologies. Although specific microbes have been implicated in causing and/or promoting specific diseases, it is now becoming clear that it is the overall community structure of microbiota that is the 'Lakshman Rekha' that separates health and disease, and diet seems to play a very crucial role in altering the community structure of the gut microbiota. The way to a man's heart is certainly through his stomach but via the microbiota.

Acknowledgements P.H., U.S., V.G., N.N. and N.K.M. gratefully acknowledge Council of Scientific and Industrial Research (CSIR) and University Grant Commission (UGC) for providing research scholarships. We also gratefully acknowledge Department of Biotechnology (DBT), Government of India (under project BT/PR3301/BCE/08/875/2011) for providing infrastructural support.

Competing Interests No competing interests to be disclosed.

References

Adams JB, Romdalvik J, Levine KE, Hu LW. Mercury in first-cut baby hair of children with autism versus typically-developing children. Toxicol Environ Chem. 2008;90:739–53.

Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012;338:120–3.

Backhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanism underlying the resistance to diet-induced obesity in germ free mice. Proc Natl Acad Sci USA. 2007;104:979–84.

Bäckhed F, et al. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci USA. 2004;101:15718–23.

Bartlett JG. Antibiotic-associated diarrhea. N Engl J Med. 2002;346:334–9.

Berg RD. The indigenous gastrointestinal microflora. Trends Microbiol. 1996;4:430–5.

Bianconi E, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40:463–71.

Britton RA, Young VB. Interaction between the intestinal microbiota and host in Clostridium difficile colonization resistance. Trends Microbiol. 2012;20:313–9.

Buss SN, et al. Gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. J Clin Microbiol. 2015;53:915–25.

Cani PD, et al. Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes. 2007a;56:1761–72.

Cani PD, et al. Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. Diabetologia. 2007b;50:2374–83.

Carbonero F, Benefiel AC, Gaskins HR. Contributions of the microbial hydrogen economy to colonic homeostasis. Nat Rev Gastroenterol Hepatol. 2012;9:504–18.

Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, Young VB. Decreased diversity of the fecal microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis. 2008;197:435–8.

Chassaing B, et al. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature. 2015;519:92–6.

Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13:260–70.

Collins SM, Bercik P. The relationship between intestinal microbiota and the central nervous system in normal gastrointestinal function and disease. Gastroenterology. 2009;136:2003–14.

Danese S, Malesci A, Vetrano S. Colitis-associated cancer: the dark side of inflammatory bowel disease. Gut. 2011;60:1609–10.
Di Martino ML, et al. Polyamines: emerging players in bacteria–host interactions. Int J Med Microbiol. 2013;303:484–91.
Dominguez-Bello MG, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci USA. 2010;107:11971–5.
Drudy D, Kyne L, O'Mahony R, Fanning S. gyrA mutations in fluoroquinolone-resistant Clostridium difficile PCR-027. Emerg Infect Dis. 2007;13:504–5.
Duncan SH, Hold GL, Barcenilla A, Stewart CS, Flint HJ. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol. 2002;52:1615–20.
Everard A, Cani PD. Diabetes, obesity and gut microbiota. Best Pract Res Clin Gastroenterol. 2013;27:73–83.
Field D, Garland M, Williams K. Correlates of specific childhood feeding problems. J Paediatr Child Health. 2003;39:299–304.
Finegold SM, et al. Pyrosequencing study of fecal microflora of autistic and control children. Anaerobe. 2010;16:444–53.
Flint HJ, Duncan SH, Louis P. Gut microbiome and obesity in treatment of the obese patient. 2014:73–82.
Freter R. The fatal enteric cholera infection in the guinea pig achieved by inhibition of normal enteric flora. J Infect Dis. 1955;97:57–65.
Fukiya S, et al. Conversion of cholic acid and chenodeoxycholic acid into their 7-oxo derivatives by Bacteroides intestinalis AM-1 isolated from human feces. FEMS Microbiol Lett. 2009;293:263–70.
Gerritsen J, Smidt H, Rijkers GT, de Vos WM. Intestinal microbiota in human health and disease: the impact of probiotics. Genes Nutr. 2011;6:209–40.
Gill CIR, Rowland IR. Diet and cancer: assessing the risk. Br J Nutr. 2002;88:s73–87.
Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen YY, Knight R, Ahima RS, Bushman F, Wu GD. High-fat diet determines the composition of the murine gut microbiome independently of obesity. Gastroenterology. 2009;137:1716–24.
Hoffmann C, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One. 2013;8:e66019.
Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. Am J Gastroenterol. 2011;106:563–73.
Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012;486:215–21.
Ibrahim SH, Voigt RG, Katusic SK, Weaver AL, Barbaresi WJ. Incidence of gastrointestinal symptoms in children with autism: a population-based study. PLoS One. 2009;124:680–6.
Irrazábal T, Belcheva A, Girardin SE, Martin A, Philpott DJ. The multifaceted role of the intestinal microbiota in colon cancer. Mol Cell. 2014;54:309–20.
Jakobsson HE, et al. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. PLoS One. 2010;5:e9836.
Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69–90.
Jess T, Gamborg M, Matzen P, Munkholm P, Sørensen TI. Increased risk of intestinal cancer in Crohn's disease: a meta-analysis of population-based cohort studies. Am J Gastroenterol. 2005;100:2724–9.
Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007;120:1183–215.
Kamada N, et al. Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. Science. 2012;336:1325–9.
Kang S, et al. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. Inflamm Bowel Dis. 2010;16:2034–42.
Kang DW, et al. Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. PLoS One. 2013;8:e68322.

Karas JA, Enoch DA, Aliyu SH. A review of mortality due to Clostridium difficile infection. J Infect. 2010;61:1–8.
Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome, and immune system: envisioning the future. Nature. 2011;474:327–36.
Keitel V, Kubitz R, Häussinger D. Endocrine and paracrine role of bile acids. World J Gastroenterol. 2008;14:5620–9.
Khanna S, Tosh PK. A clinician's primer on the role of the microbiome in human health and disease. Mayo Clin Proc. 2014;89:107–14.
Khanna S, et al. The epidemiology of community-acquired Clostridium difficile infection: a population-based study. Am J Gastroenterol. 2012;107:89–95.
Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J Clin Gastroenterol. 2010;44:354–60.
Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. Gut. 2013;62:1505–10.
Kostic AD, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe. 2013;14:207–15.
Lanis JM, Barua S, Ballard JD. Variations in TcdB activity and the hypervirulence of emerging strains of Clostridium difficile. PLoS Pathog. 2010;6:e1001061.
Lanis JM, Heinlen LD, James JA, Ballard JD. Clostridium difficile 027/BI/NAP1 encodes a hypertoxic and antigenically variable form of TcdB. PLoS Pathog. 2013;9:e1003523.
Lanis JM, Hightower LD, Shen A, Ballard JD. TcdB from hypervirulent Clostridium difficile exhibits increased efficiency of autoprocessing. Mol Microbiol. 2012;84:66–76.
Lefebvre P, Cariou B, Lien F, Kuipers F, Staels B. Role of bile acids and bile acid receptors in metabolic regulation. Physiol Rev. 2009;89:147–91.
Ley RE, Backed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci USA. 2005;102:11070–5.
Li K, Bihan M, Yooseph S, Methé BA. Analyses of the microbial diversity across the human microbiome. PLoS One. 2012;7:e32118.
Libby JM, Jortner BS, Wilkins TD. Effects of the two toxins of Clostridium difficile in antibiotic-associated cecitis in hamsters. Infect Immun. 1982;36:822–9.
Loh YH, Jakszyn P, Luben RN, Mulligan AA, Mitrou PN, Khaw KT. N-nitroso compounds and cancer incidence: the European Prospective Investigation into Cancer and Nutrition (EPIC)—Norfolk Study. Am J Clin Nutr. 2011;93:1053–61.
Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol. 2014;12:661–72.
Lyerly DM, Phelps CJ, Toth J, Wilkins TD. Characterization of toxins A and B of Clostridium difficile with monoclonal antibodies. Infect Immun. 1986;54:70–6.
Lyras D, et al. Toxin B is essential for virulence of Clostridium difficile. Nature. 2009;458:1176–9.
Mackie RI, Sghir A, Gaskins HR. Developmental microbial ecology of the neonatal gastrointestinal tract. Am J Clin Nutr. 1999;69:1035s–45s.
Marsh JW, Shutt KA, Brooks MM, Pasculle AW, Muto CA, Harrison LH. Perirectal swab surveillance for Clostridium difficile by use of selective broth preamplification and real-time PCR detection of tcdB. J Clin Microbiol. 2012;50:4078–82.
Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. Nature. 2008;453:620–5.
Merrigan M, et al. Human hypervirulent Clostridium difficile strains exhibit increased sporulation as well as robust toxin production. J Bacteriol. 2010;192:4904–11.
Mikelsaar M, Zilmer M. Lactobacillus fermentum ME-3—an antimicrobial and antioxidative probiotic. Microb Ecol Health Dis. 2009;21:1–27.
Minami T, Oda K, Gima N, Yamazaki H. Effects of lipopolysaccharide and chelator on mercury content in the cerebrum of thimerosal-administered mice. Environ Toxicol Pharmacol. 2007;24:316–20.

Morgan XC, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79.
Mulle JG, Sharp WG, Cubells JF. The gut microbiome: a new frontier in autism research. Curr Psychiatry Rep. 2013;15:337–49.
Musso G, Gambino R, Cassader M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu Rev Med. 2011;62:361–80.
Ou J, DeLany JP, Zhang M, Sharma S, O'Keefe SJ. Association between low colonic short-chain fatty acids and high bile acids in high colon cancer risk populations. Nutr Cancer. 2012;64:34–40.
Ou J, et al. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. Am J Clin Nutr. 2013;98:111–20.
Pegg AE. Toxicity of polyamines and their metabolic products. Chem Res Toxicol. 2013;26:1782–800.
Pertea M, Salzberg SL. Between a chicken and a grape: estimating the number of human genes. Genome Biol. 2010;11:206–12.
Peterson J, et al. The NIH human microbiome project. Genome Res. 2009;19:2317–23.
Ravel J, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci USA. 2011;108:4680–7.
Ray K. IBD: gut microbiota in IBD goes viral. Nat Rev Gastroenterol Hepatol. 2015;12(3):122.
Rea MC, et al. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. Proc Natl Acad Sci USA. 2010;107:9352–7.
Ridlon JM, Kang DJ, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res. 2006;47:241–59.
Rowland RHIR. Metabolic activities of the gut microflora in relation to cancer. Microb Ecol Health Dis. 2000;12:179–85.
Rustgi AK. The genetics of hereditary colon cancer. Genes Dev. 2007;21:2525–38.
Sambol SP, Merrigan MM, Tang JK, Johnson S, Gerding DN. Colonization for the prevention of Clostridium difficile disease in hamsters. J Infect Dis. 2002;186:1781–9.
Schmidt C. Mental health: thinking from the gut. Nature. 2015;518:S12–5.
Schnorr SL. The diverse microbiome of the hunter-gatherer. Nature. 2015;518:S14–5.
Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13:800–12.
Sears CL. A dynamic partnership: celebrating our gut flora. Anaerobe. 2005;11:247–51.
Sears CL. Enterotoxigenic Bacteroides fragilis: a rogue among symbiotes. Clin Microbiol Rev. 2009;22:349–69.
Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe. 2014;15:317–28.
Selmer T, Andrei PI. p-Hydroxyphenylacetate decarboxylase from Clostridium difficile. A novel glycyl radical enzyme catalysing the formation of p-cresol. Eur J Biochem. 2001;268:1363–72.
Sharp WG, Berry RC, McCracken C, Nuhu NN, Marvel E, Saulnier CA, Klin A, Jones W, Jaquess DL. Feeding problems and nutrient intake in children with autism spectrum disorders: a meta-analysis and comprehensive review of the literature. J Autism Dev Disord. 2013;43:2159–73.
Shen J, Obin MS, Zhao L. The gut microbiota, obesity and insulin resistance. Mol Asp Med. 2013;34:39–58.
Sokol H, Seksik P. The intestinal microbiota in inflammatory bowel diseases: time to connect with the host. Curr Opin Gastroenterol. 2010;26:327–31.
Sokol H, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci USA. 2008;105:16731–6.
Sorg JA, Sonenshein AL. Bile salts and glycine as cogerminants for Clostridium difficile spores. J Bacteriol. 2008;190:2505–12.
Sorg JA, Sonenshein AL. Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol. 2010;192:4983–90.
Spehlmann ME, et al. Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. Inflamm Bowel Dis. 2008;14:968–76.

Taylor NS, Thorne GM, Bartlett JG. Comparison of two toxins produced by Clostridium difficile. Infect Immun. 1981;34:1036–43.
Todar K. The normal bacterial flora of humans. Todar's Online textbook of Bacteriology, 2012. http://www.textbookofbacteriology.net/normalflora_3.html
Tu S, et al. Overexpression of interleukin-1β induces gastric inflammation and cancer and mobilizes myeloid-derived suppressor cells in mice. Cancer Cell. 2008;14:408–19.
Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe. 2008;3:213–23.
Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449:804–10.
Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027–31.
Turnbaugh PJ, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457:480–4.
Warny M, et al. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. Lancet. 2005;366:1079–84.
Watanabe Y, Nagai F, Morotomi M. Characterization of *Phascolarctobacterium succinatutens* sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces. Appl Environ Microbiol. 2012;78:511–8.
Watson AJ, Collins PD. Colon cancer: a civilization disorder. Dig Dis. 2010;29:222–8.
Wen L, et al. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature. 2008;455:1109–13.
Windey K, De Preter V, Verbeke K. Relevance of protein fermentation to gut health. Mol Nutr Food Res. 2012;56:184–96.
World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR, 2007.
Yoshimoto S, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature. 2013;499:97–101.
Zackular J. Characterizing the role of the gut microbiome in colorectal cancer. Doctoral dissertation, University of Michigan; 2014.
Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res. 2014;7:1112–21.
Zar JH. Measures of variability and dispersion, in Biostatistical analysis 5th Ed. (Editor: Lynch D) Pearson Publishing Company, USA 2010, 33–46.
Zhu Y, Carvey PM, Ling Z. Altered glutathione homeostasis in animals prenatally exposed to lipopolysaccharide. Neurochem Int. 2007;50:671–80.
Zipris D. Innate immunity and its role in type 1 diabetes. Curr Opin Endocrinol Diabetes Obes. 2008;15:326–31.

Dr. Rup Lal Dr. Rup Lal is an eminent professor of the University of Delhi, Delhi. His primary research interests include microbial diversity at pesticide-polluted sites, genetic and biochemistry of hexachlorocyclohexane (HCH) degradation, and development of HCH bioremediation technology. Prof. Rup Lal is also interested in the evolution of lin genes and genetic manipulation of rifamycin producer *Amycolatopsis mediterranei*. His group has already sequenced genomes of *Amycolatopsis mediterranei* S699, *Thermus* sp. RL, *Acinetobacter* sp. HA, and *Sphingobium indicum* B90A.

Dr. Mallikarjun Shakarad Dr. Mallikarjun Shakarad is a professor at the Department of Zoology, University of Delhi. His area of expertise is in the evolution of life histories in insect systems and the influence of microbiota in altering the health span and life histories in higher animal systems.