

INVESTIGATING  
THE  
HUMAN  
GENOME

INSIGHTS INTO HUMAN VARIATION  
AND DISEASE SUSCEPTIBILITY

MOYRA SMITH

INVESTIGATING THE  
HUMAN GENOME

*Insights into Human Variation  
and Disease Susceptibility*

MOYRA SMITH

Vice President, Publisher: Tim Moore  
Associate Publisher and Director of Marketing: Amy Neidlinger  
Acquisitions Editor: Kirk Jensen  
Editorial Assistant: Pamela Boland  
Operations Manager: Gina Kanouse  
Senior Marketing Manager: Julie Phifer  
Publicity Manager: Laura Czaja  
Assistant Marketing Manager: Megan Colvin  
Cover Designer: Chuti Prasertsith  
Managing Editor: Kristy Hart  
Senior Project Editor: Lori Lyons  
Copy Editor: Hansing Editorial Services  
Proofreader: Kathy Ruiz  
Senior Indexer: Cheryl Lenser  
Compositor: Nonie Ratcliff  
Manufacturing Buyer: Dan Uhrig

© 2011 by Pearson Education, Inc.  
Publishing as FT Press  
Upper Saddle River, New Jersey 07458

FT Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact U.S. Corporate and Government Sales, 1-800-382-3419, [corpsales@pearsontechgroup.com](mailto:corpsales@pearsontechgroup.com). For sales outside the U.S., please contact International Sales at [international@pearson.com](mailto:international@pearson.com).

Company and product names mentioned herein are the trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

First Printing June 2011

ISBN-10: 0-13216814-6

ISBN-13: 978-0-13216814-4

Pearson Education LTD.  
Pearson Education Australia PTY, Limited.  
Pearson Education Singapore, Pte. Ltd.  
Pearson Education Asia, Ltd.  
Pearson Education Canada, Ltd.  
Pearson Educación de Mexico, S.A. de C.V.  
Pearson Education—Japan  
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Cataloging-in-Publication Data  
Smith, Moyra.

Investigating the human genome : insights into human variation and disease  
susceptibility / Moyra Smith.

p. ; cm.

Includes bibliographical references.

ISBN-13: 978-0-13-216814-4 (hardback : alk. paper)

ISBN-10: 0-13-216814-6 (hardback : alk. paper)

1. Human genome. 2. Disease susceptibility. I. Title.

[DNLM: 1. Genome, Human. 2. Genetic Predisposition to Disease. QU 470]

QH447.S625 2011

611'.01816—dc22

2011006494

*This work is dedicated to the memory of three  
remarkable people who inspired and  
encouraged me long ago:*

*My mother Florence Van Eyk Smith,  
my grandfather Manard James Van Eyk,  
and our beloved family physician,  
Dr. Colin Roy Cockcroft.*

*This page intentionally left blank*

# Contents

	Preface .....	vii
Chapter 1	Genome Architecture and Sequence Variation in Health and Disease .....	1
Chapter 2	Genes and Transcripts: Insight into Regulation at Different Levels .....	23
Chapter 3	Epigenetics: Modifications of DNA, Chromatin, and Gene Expression .....	37
Chapter 4	Gene Environment Interactions .....	53
Chapter 5	Pathways, Phenotypes, and Phenocopies .....	67
Chapter 6	Dynamic Function, Synaptic Activity, and Plasticity .....	81
Chapter 7	Late Onset Neurodegenerative Diseases .....	101
Chapter 8	Genes and Genomes in Cancer: Targeted Therapies .....	131
Chapter 9	Functional Genomics: Personalized Medicine and Therapeutics .....	153
	Epilogue .....	167
	References .....	171
	About the Author .....	203
	Index .....	205

# Acknowledgments

I express my gratitude to researchers and clinicians who have worked hard and long to further our knowledge of genetics and genetic diseases. Preparation of this book would not have been possible without the amazing resources of the University of California library services and the online resources of the National Library of Medicine—I want to acknowledge all who contribute to making these resources available.

I want to acknowledge Kirk Jensen, who encouraged me to write this book. I thank Lori Lyons of Pearson for moving this manuscript along in the publishing process and copy editor Krista Hansing for her thorough input.

# Preface

In 2009, Georgina Ferry wrote, “The optimism of science is two-fold: that its methods might reveal, one tiny pixel at a time, more of the wonder of the natural world and that this knowledge might be able to solve practical human problems.”

Progress in the fields of genetics and genomics since a draft sequence of the human genome was published in 2001 is indeed a cause for optimism. However, this discovery has left some people disappointed because development of new therapies to treat disease has been slower than anticipated. Availability of this sequence information has fueled groundbreaking studies in genetics, genomics, and epigenetics that provide insight into human variation and the pathogenesis of both common and rare diseases. The goal of this book is to briefly review several of those groundbreaking studies and new insights.

My own experiences during a 40-year career as a clinical geneticist and researcher in genetics and genomics influenced the choice of topics discussed in this book. I explore new insights into human origins, migrations, and human population diversity gained through genomic analyses. I consider insights into the etiology of common diseases such as diabetes and coronary heart disease. I also consider studies on synapses and synaptic plasticity, representing pathways to understanding mind and cognition.

I discuss complexities of late-onset neurological diseases and efforts to utilize genetic and genomic methodologies to unravel the pathogenesis of these disorders. I also consider new insights into aspects of protein misfolding and clearance or deposition as aggregates that sequester other proteins. In considering gene environment interactions, I focus on aspects of DNA damage and repair and DNA

instability. An appropriate movement is underway toward translational research and greater emphasis on treatment. I review examples of treating primary defects and downstream effects of genetic disorders. I review new information on regulating gene expression at the levels of transcription, translation, and post-translational modifications. Growing evidence indicates that modifications of DNA, histones, and of nonhistone proteins greatly impact gene expression and the function of gene products, and I review aspects of research in these areas, sometimes referred to as epigenetics. In a chapter related to cancer, I review new discoveries in genetics and genomics that have direct relevance to therapy.

Growing evidence points to the importance of protein interactions and webs of molecular interactions that determine regulation and growth and the operation of systems, and I consider these topics. In a closing chapter I consider the relevance of genomics and systems biology to personalized medicine.

# 1

---

## Genome architecture and sequence variation in health and disease

Availability of information on DNA sequence in human genomes and advances in technologies to amplify and sequence DNA have led to significant progress in delineating sequence differences that lead to disease. These techniques have also led to the discovery of sequence variants that occur in healthy individuals.

Studies of variation in the human genome are greatly facilitated through the availability of microarrays designed to detect single nucleotide polymorphisms (SNPs) that occur with frequencies greater than 1% to 5% in the population. Gene loci that are close to each other are often coinherited. SNP analyses can determine a series of alleles of loci in a specific region (a haplotype). Microarray technologies enable analysis of as many as one million SNPs on each array. These microarrays can also determine structural variation and copy number changes, defined as deletion or duplications greater than 1 kilobase (kb). Specific probes for regions known to frequently harbor copy number changes are also present on SNP microarrays such as the Affymetrix 6.0 array. Advances in technologies in DNA sequencing include massively parallel sequencing, often referred to as next-generation sequencing.

This chapter explores aspects of structural genomic variation and sequence variation in different populations and the role of sequence differences in the etiology of common disorders such as diabetes

mellitus, obesity, and coronary heart disease. It also covers next-generation sequencing and examples of its application to the discovery of gene defects that lead to disease.

Through the use of polymerase chain reaction techniques, samples with low concentrations of DNA can be used to derive material for DNA sequencing. This chapter discusses applications of these techniques to discover how the sequence in modern humans differs from that of Neanderthals and early modern humans. Also presented are reports of studies of DNA extracted from two teeth from a man who died in 1783. DNA analysis enabled researchers to diagnose the disease that afflicted him and analyze the specific mutation and surrounding polymorphisms that connected him to present-day patients with the same disease.

## **Structural variation**

In the human genome, segmental duplications with highly identical sequence are usually interspersed and separated by more than 1 megabase. She, et al. (2006), identified more than 400 duplication blocks within the human genome. Segmental duplications are frequently clustered in pericentric and subtelomeric regions (Marques-Bonet, et al., 2009). Evidence indicates that pericentric and subtelomeric duplications evolved independently from intrachromosomal duplications. Core duplicons of 5–30kb occur in intrachromosomal duplications. One example of a core duplicon is LCR16a, which is rich in Alu repeats.

Unequal crossover between directly oriented duplicated segments may lead to dosage changes or altered structure and function of a gene. Marques-Bonet, et al., noted that most copy number polymorphisms result from this mechanism.

Regions between segmental duplications may be deleted, duplicated, or inverted as a result of unequal crossover. The existence of

highly similar duplicated segments on two different chromosomes may lead to translocation events. Polymorphisms also exist within the segmental repeats, and in different individuals, these regions may be larger or smaller. Segmental duplications are particularly abundant in certain chromosome regions, such as 15q11-q13, and these regions are frequent sites of deletions and duplications.

A key question is whether a specific structural variant, such as a deletion or a duplication (copy number variant) that includes unique sequence DNA, is a direct cause of phenotypic abnormality. Genomic syndromes often occur as a result of deletion or duplication of genomic regions that are flanked by segmental duplication blocks. In these syndromes, specific phenotypes result from the deletion of specific regions; for example, Williams syndrome results from the deletion of chromosome 7q11.2. Characteristic phenotypic features of this syndrome include cognitive and behavioral impairments, distinct facial features, and cardiac malformations.

Girirajan and Eichler (2010), reviewed findings in a subset of genomic structural changes in which a particular genomic change results in a series of phenotypes in which specific clinical features differ in different individuals. Differences occur in the degree to which individuals with the same defect are affected—that is, there are varying degrees of penetrance. The clinical consequences of a particular dosage change in a specific region may be influenced by dosage changes or mutations elsewhere in the genome.

Examples of specific regions where deletions are associated with a variety of phenotypes include 16p11.2. In some cases with deletion in this region, severe obesity occurs; other cases with the same deletion are diagnosed with autism, while in others, congenital malformations and developmental delay occur. Diverse phenotypes have been described in cases with deletion of 17q12; some cases present with hereditary neuropathy, with a tendency to pressure palsy (HNPP);

and in other cases, schizophrenia occurs. Other diagnoses encountered in patients with 17q12 deletion include renal cystic disease or maturity-onset diabetes of the young. Deletion in 1q21.1 may be associated with a learning disability in some cases and with congenital heart disease or schizophrenia in others.

The copy number variants associated with diverse phenotypes are sometimes found with low frequency in control populations. One question that arises is whether the different phenotypic consequences result from slight differences in the position of deletion breakpoints and whether sequence differences occur in the same region on the homologous chromosome.

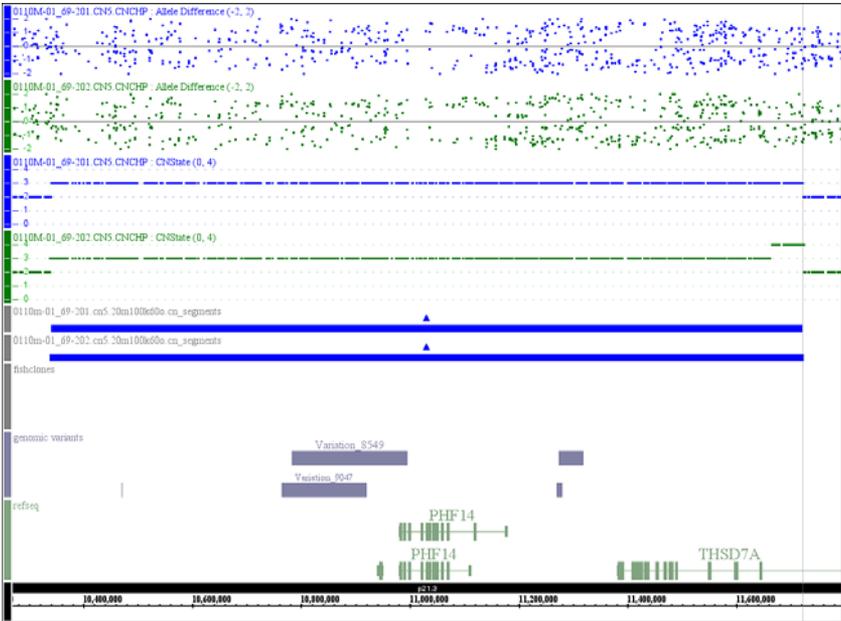
Another genetic factor that may play a role in some cases is that the deletion of a specific locus on one chromosome unmasks a recessive mutant allele at that locus on the homologous chromosome. Other important possible explanations for the phenotypic variation are that additional genetic modifiers elsewhere in the genome modify the phenotype.

Girirajan and Eichler (2010), proposed that a two-hit genomic model most likely explains the variable phenotypes in individuals with copy number variants in 16p12.1 or 22q11.2.

Copy number variations and deletions, in particular, are most often considered to be of clinical relevance if they arise *de novo*—that is, if they are present in a child but absent in the parents. However, growing evidence indicates that parents who carry specific copy number variants may have subclinical manifestations attributable to the genomic change. CNVs and microarray are illustrated in Figure 1.1.

## Human genetic sequence variation

During the past decade there have been significant advances in technologies for DNA sequencing that have facilitated studies of variation in ancient and modern humans.



**Figure 1.1** Results of analyzing SNP alleles and copy number variants using an Affymetrix 6.0 array and genotyping console on twin females with autism. Rows 1 and 2 at the top of the figure show the distribution of A and B alleles of specific SNPs. Note the identical patterns of alleles in the twins. Rows 3, 4, 5, and 6 show a chromosomal region with a copy number variant. Each twin has three copies of the CNV that encompasses three genes, shown at the bottom of the figure. A known population variant region is indicated, but the variant region is shorter and does not encompass a gene.

### *Studies in ancient fossil remains*

In 2010, Green, et al., published data on four billion nucleotides of DNA sequence from three different Neanderthal individuals. They noted that DNA extracted from these late Pleistocene remains had degraded to segments less than 200 nucleotides in length and that it had been chemically modified. In addition, they found substantial contamination from microorganisms. To enrich the Neanderthal DNA, samples were digested with restriction endonucleases that selectively cleave microbial DNA.

Green, et al., examined the DNA sequence in loci with specific alleles that are known to differ in different modern human populations. They determined that Neanderthals shared 1% to 4% of genotypes at the sequences with Europeans and Asians. At these loci, Neanderthals did not share alleles with Sub-Saharan African populations.

Sequence analyses also led to the identification of genes that apparently underwent positive selection in modern humans. Specific sequences in these genes impact protein function. Green, et al., identified specific functional sequence changes that occurred in modern humans but were absent from Neanderthals, and the Neanderthal sequence matched the sequence present in chimpanzees.

### ***Studies on an ancient Saqqaq individual***

In the past decade, we have seen the confluence of paleontological analyses of bone fossils and cultural artifacts with DNA analyses. Rasmussen, et al. (2010), examined DNA recovered from the hair roots of an individual from Greenland, estimated to have lived 4,000 years ago, who was of member of the Saqqaq culture. Analysis of DNA polymorphisms from the hair roots indicated that the closest match was with individuals from eastern Siberia.

One advantage of analyses from hair roots is that they are less contaminated with fungi and bacteria than samples isolated from bone fossils. The high quality of the DNA isolated from hair roots of the Saqqaq individual enabled the analysis of 350,000 SNPs. Earlier studies by Rasmussen's group generated information on the complete mitochondrial DNA gene from permafrost-preserved Saqqaq individuals.

Given the length of the tracts of homozygosity they found, Rasmussen, et al., concluded that the inbreeding coefficient was high. Rasmussen studied DNA sequence at functional polymorphic sites. The combination of SNPs at the *HERC2* and *OCA2* (oculocutaneous

albinism gene<sup>2</sup>) indicated that the individual most likely had brown eyes and dark hair. Analyses also revealed that the Saqqaq individual was most closely related to three Northern Old World Arctic populations and was more distantly related to New World Amerinds. Researchers were not able to detect evidence of West Eurasian population admixture. Nuclear DNA SNP analyses and studies on the mitochondrial and Y chromosome haplotypes of the Saqqaq individuals matched most closely with those of North East Asian populations.

The Saqqaq culture is a component of the Arctic small tool transition and is estimated to have existed between 4,750 and 2,500 years ago.

### ***Sequence variation in different populations and regions***

In an analysis of 650,000 common SNPs, Li, et al. (2008), collected samples from populations in 51 geographic regions. Populations studied were drawn from Sub-Saharan Africa, North Africa, the Middle East, Europe, East, South, and Central Asia, Oceania, and the Americas. They carried out haplotype analysis to identify linked alleles at specific loci. They detected finer haplotype substructure in different regions. They noted, for example, that Palestinians, Druze, and Bedouins have haplotype contributions from the Middle East, Europe, and South and Central Asia.

Li, et al., concluded that nonrandom differences between populations have accumulated at a number of different loci. However, they also concluded that within population differences accounted for most of the genetic diversity. Results of their analyses revealed that heterozygosity was greatest in Africa and was reduced as geographic distance from Addis Ababa increased.

Tishkoff, et al. (2009), studied genotypes in 121 African populations, in 60 non-African populations, and in the African-American population. They studied microsatellite repeat polymorphisms and insertion deletion polymorphisms. They obtained evidence for

regional differences in the allele frequency of markers; however, their analyses also revealed evidence for substantial population admixture.

### ***Homozygosity mapping***

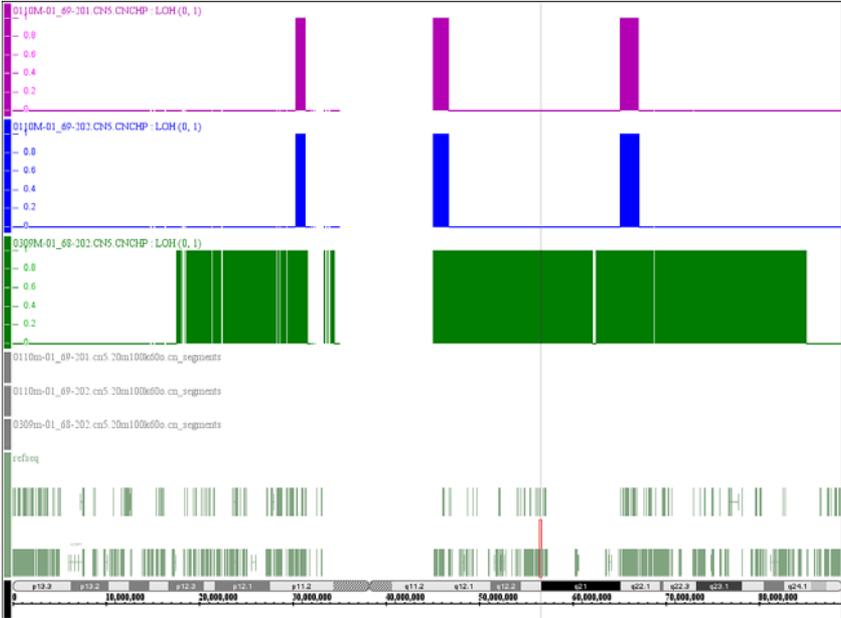
Because recombination occurs between homologous chromosomes during meiosis, the presence of identical alleles over long stretches of genomic DNA on homologous chromosomes (homozygosity) was thought to occur only in consanguineous pedigrees or inbred populations. Gibson, et al. (2006), studied 262 individuals in four different populations and identified 20 different genomic regions where homozygosity extended over 1 megabase or more. They noted that the lowest number of homozygous tracts occurs in the Yoruba population and that this reflects the more ancient roots of the population; over longer time periods, segments of chromosomes have broken up. Gibson, et al., noted that, in modern populations, tracts of homozygosity often occurred in similar genomic regions, indicating regions with a lower frequency of recombination. Examples of blocks of homozygosity are illustrated in Figure 1.2.

### ***Studies on hereditary disorders and population history***

Currently, 36 disorders are considered to comprise the Finnish disease heritage. Norio (2003) reviewed the history and studies of these disorders and related them to the historical origins, migrations, and settlements of the Finnish population. Thirty-two of these disorders have autosomal recessive inheritance patterns, two are autosomal dominant, and one is an X-linked disorder.

Norio noted that the Finnish population has been relatively stable for many years, without evidence of continuous migration into Finland. Internal migration of families from the southeastern parts of the country around Sevo into middle and northern regions of the country occurred around 1600. The migrant families settled in small clusters. Each cluster was often located at some distance from other

clusters, and with little admixture between clusters, mating occurred within clusters. In later generations, couples who married often shared founders six or seven generations ago.



**Figure 1.2** Blocks of homozygosity that are identical in twins (rows 1 and 2). Strikingly large blocks of homozygosity are present in the individual illustrated in row 3, likely due to consanguinity of parents. Rows 4 and 5 indicate positions of genes on chromosome 16.

In 1999, Peltonen, et al., reported that gene loci for 32 of the Finnish hereditary diseases were mapped to specific chromosomes and causative genes for 17 of the disorders were isolated. As expected, marked linkage disequilibrium occurred for markers in the vicinity of the disease alleles, and homogeneity of the disease-causing alleles was observed.

Peltonen and colleagues studied molecular mechanisms in these genetic disorders. Molecular analyses of products encoded by genes in regions where the diseases mapped resulted in the discovery of new proteins and enzymes. Peltonen emphasized that analysis of the

disease genes facilitated disease diagnosis, and, importantly, health-care was available following diagnosis.

Peltonen and coworkers also reported that analyzing linkage disequilibrium is useful in identifying gene loci that contribute to the risk of complex common diseases. Kilpinen, et al. (2009), studied regions of linkage disequilibrium in a unique pedigree with multiple cases of autism. Individuals in this pedigree shared ancestors in the 17th century. Analyses revealed areas of linkage disequilibrium in three chromosomal regions at 15q11-q13, 19p13, and 1q23.

### ***Genetic variations, single nucleotide polymorphisms (SNPs) and genome wide association studies (GWAS)***

The design of genome wide association studies (GWAS) is predicated on the hypothesis that common DNA sequence variants contribute to the etiology of common disease. Results indicate that even when statistically significant associations between disease and a specific SNP are determined, the overall contribution of specific SNPs to disease risk is often low.

### ***Genome wide association studies and insight into etiology of type 2 diabetes and obesity***

In type 2 diabetes, the pancreatic beta cell–secreting capacity becomes inadequate to overcome the progressive peripheral resistance to insulin uptake. Factors that play roles in the development of peripheral insulin resistance include age, inactivity, and weight gain. McCarthy (2010) reviewed the discovery of genes that impact susceptibility to diabetes and obesity. He considered three waves of discovery. The first included family-based linkage studies. These studies led to the identification of genes involved in a number of Mendelian forms of early-onset diabetes, including neonatal diabetes and maturity-onset diabetes of the young (MODY). Genes that were found to

play a role in MODY included *NEUROD1* (neurogenic differentiation 1); *GCK* (glucokinase); hepatic nuclear factor genes *HNF1A*, *HNF1B*, and *HNF4A*; and *IPF* (insulin promoter factor). Family studies also led to the discovery of a mitochondrial DNA mutation that predisposes carriers to diabetes and deafness.

McCarthy noted that family studies of childhood obesity led to the discovery of rare forms of this condition due to mutations in any one of three genes: leptin, leptin receptor, and pro-opiomelanocortin.

The second phase of investigation into diabetes and obesity involved searching for variants in candidate genes. These studies led to the identification of common variants of modest effect in *PPARG* (peroxisome proliferation activated receptor gamma) and *KCNJ11* (potassium inwardly-rectifying channel, subfamily J, member 11). Resequencing of the melanocortin 4-receptor gene led to the identification of associated variants in 2% to 3% of cases of obesity.

A third wave of studies involved large-scale analysis of common DNA sequence variants (SNPs). McCarthy considered this to be the most successful wave of studies. Important diabetes-associated loci identified in these studies include the transcription factor that modulates pancreatic function *TCF7L2*; cyclin-dependent kinases *CDKAL1*, *CDKN2A*, and *CDKN2B*, which regulate cyclin; and *HHEX*, a gene involved in beta cell development. Each copy of a susceptibility allele at one of these loci leads to a 15% to 20% increase in the risk for diabetes.

McCarthy reported at least 40 known loci with alleles associated with increased risk of diabetes. Of interest is the fact that five of the loci with common variant alleles associated with increased risk of diabetes also harbor rare variants involved in familial or syndromic diabetes. These four loci are wolframin (*NFS1*); hepatocyte nuclear factors *HNF1A* and *HNF1B*; the melatonin receptor *MTNR1B*; and *IRS1* insulin receptor substrate 1, which impacts insulin action.

## Pathways involved in diabetes

McCarthy reported that the loci with the strongest evidence of association with type 2 diabetes impact insulin secretion. Examples include cyclin-dependent kinases CDKAL1, CDKNL2A, and CDKN2B. At these loci, risk variants predispose to reduced pancreatic beta cell mass. The diabetes risk alleles in TCF7L2, MTNR1B, and KCNJ11 predispose to beta cell dysfunction.

Risk alleles in the FTO locus (gene related to fat mass and obesity) contribute to obesity and to peripheral resistance to insulin. The PPARG and IRS1 (insulin receptor substrate 1) loci impact insulin resistance and obesity.

In the monogenic forms of diabetes and in autoimmune diabetes in adults, information about the underlying causative gene can influence therapeutic decisions, such as whether insulin is required, whether dietary management may be sufficient, or whether sulfonylureas are required. In type 1 diabetes commonly associated with HLA variants (latent autoimmune diabetes) or with defects in the insulin genes INS or PTPN22 (protein tyrosine phosphatase non receptor type 22), insulin is likely necessary. Maturity-onset diabetes of the young (MODY) due to GCK (glucokinase) deficiency may respond adequately to dietary management. MODY due to HNF1A deficiency may require treatment with sulfonylureas.

McCarthy reviewed the results of genome wide association studies designed to identify common variants associated with increased body mass index (BMI) and noted that at least 30 such loci have been associated. The strongest signal is associated with the FTO locus (gene related to fat mass obesity). He noted that signals of risk alleles were also detected in genes with neuronal function, such as BDNF (brain derived neurotrophic factor), SH2B1 (signaling protein), and NEGR1 (neuronal growth regulator). He indicated that obesity may be partly a disease of disordered hypothalamic function. In studies that involved analyzing fat mass distribution, risk alleles in 15 loci

were identified. Evidence indicates that risk alleles at these loci impact adipocyte development and function.

McCarthy noted that clinical translation of these findings is impacted at least partly by the modest effect of the risk alleles. Homozygotes for the FTO risk allele are an average of 2 to 3 kilograms heavier than in individuals without the risk allele. However, he noted that identifying risk-altering genes contributes to our understanding of the biology of disease. Another important consideration is that most of the risk alleles lie outside the coding regions of genes, and it's not clear how they impact the regulation of gene expression.

McCarthy predicted that large-scale genome-wide resequencing efforts now underway would clarify relationships between sequence variants and clinical phenotypes.

## **Tracking genes involved in coronary heart disease after GWAS**

In 2007, Samani, et al., carried out genome wide association studies in coronary heart disease subjects. They identified several genetic loci that affect the risk of coronary artery disease (CAD), including loci at 9p21.3 and 1p13.5. In 2008, Kathiresan, et al., identified two loci associated with abnormal levels of low-density lipoprotein cholesterol (LDL cholesterol), one locus mapped to chromosome 1p13 and the other mapped to 19p13. They noted that the 1p13 locus maps near the gene SORT1 (sortilin 1).

Kjolby, et al. (2010), demonstrated that sortilin protein encoded by SORT1 is an intracellular sorting receptor for apolipoprotein ApoB100. They noted that SORT1 regulates plasma low-density lipoprotein levels through hepatic export of ApoB100 containing lipoproteins. In studies on mice, they determined that sortilin 1 over-expression stimulates the hepatic release of lipoproteins and increases plasma LDL levels.

Musunuru, et al. (2010), carried out studies in cohorts of human subjects and in human-derived hepatocytes. They determined that a noncoding polymorphism SNP rs12740374 in 1p13 impacts a transcription factor binding site that alters hepatic expression of SORT1. The risk allele G in rs12740374 disrupts the C/EBP transcription factor binding site and is significantly associated with LDL cholesterol levels,  $p=1 \times 10^{-170}$ . In studies on mouse livers, Musunuru, et al., determined that sortilin 1 impacts plasma levels of low-density lipoprotein (LDL) and very low-density lipoprotein (VLDL) cholesterol. They demonstrated that knockdown of sortilin 1 expression in mice led to a 46% increase in total cholesterol compared with controls.

The studies of Musunuru, et al., demonstrated the clinical relevance of non-protein-coding DNA variants identified through GWAS. They concluded that the sortilin pathway is a promising new target for therapeutic intervention in hyperlipidemia and myocardial infarction. These investigators noted that, in some individuals, aggressive treatment with statins fails to lower the levels of LDL cholesterol. Statins inhibit cholesterol synthesis through inhibiting hydroxy-3methyl-glutaryl coenzyme A reductase and reduce levels of both LDL cholesterol and total cholesterol.

In GWAS analysis of blood lipids on 100,000 individuals, Linsel-Nitschke, et al. (2010), reported evidence for the involvement of 18 genes that were previously shown to play roles in Mendelian lipid disease. The significance values for association were much higher with these variants than those obtained for other variants. Highly associated loci included LPL (lipoprotein lipase)  $2 \times 10^{-115}$ , APOA1 (Apolipoprotein A1)  $7 \times 10^{-240}$ , CETP (cholesterol ester transfer protein)  $7 \times 10^{-350}$ , LDLR (LDL receptor)  $4 \times 10^{-117}$ , APOE (apolipoproteins A)  $9 \times 10^{-147}$ , APOB (Apolipoprotein B)  $4 \times 10^{-114}$ , SORTL1 (sortilin1)  $1 \times 10^{-170}$ , and GCKR (glucokinase regulator)  $6 \times 10^{-133}$ .

Therefore, evidence indicates that genes that play roles in the etiology of rare Mendelian forms of diseases such as diabetes and hyperlipidemia also play roles in the common polygenic forms of these diseases.

### ***Disease-specific mutation in a Hunterian museum skeleton and his living relatives***

In 2011, Chahal, et al., reported that they had identified a specific mutation in the arylhydrocarbon receptor interacting protein (AIP) in four families from Northern Ireland in whom familial isolated pituitary adenoma occurred. The specific mutation in these families was a nucleotide substitution c.(910 CtoT); p.(R304X). A termination codon replaces amino acid 304, leading to a loss of 26 amino acids from the AIP protein.

Chahal, et al., obtained permission from the directors of the Hunterian museum in London to extract DNA from two teeth of the skeleton of an Irish giant who died in 1783. Harvey Cushing examined this skeleton in 1909. He concluded on the basis of the degree of enlargement of the pituitary fossa that the man had a pituitary adenoma. Chahal, et al., discovered that the same AIP mutation was present in the Hunterian giant with pituitary adenoma and in the four families from Northern Ireland they studied. Analysis of DNA polymorphisms (microsatellite repeat polymorphisms) revealed that the giant skeleton DNA and adenoma patients in the four Irish families shared a haplotype that extended for 2,068 megabases on chromosome 11q13.2 and included the AIP gene. Taking into account polymorphisms, mutation rates, and generation length, Chahal, et al., concluded that the skeleton and the four families shared a common ancestor 57 to 66 generations ago.

### ***Discovery of familial-inherited adenomas in different populations and the role of AIP***

In 2006, Vierimaa, et al., reported two clusters of families from Northern Finland who had familial pituitary adenoma that led to increased secretion of growth hormone and prolactin.

Analysis of SNP polymorphisms in these families revealed a link between adenoma development and chromosome 11q12-q13.

Sequencing of genes in this region revealed a defect in the aryl hydrocarbon interacting protein AIP. Subsequent analyses in the Finnish population led to the identification of a Q14X mutation in 6 out of 45 patients with acromegaly.

Karhu and Aaltonen (2007) noted that the function of AIP was not known.

The amino-terminal region of AIP contains FKBP domains. These domains usually are involved in protein folding and trafficking. In the carboxyterminal region of AIP are three tetratricopeptide repeats. These repeats usually form scaffolds for the formation of multiprotein complexes. The carboxy-terminal region of AIP interacts with arylhydrocarbon receptor and with the HSP 90 heat shock protein. Low expression of AIP in pituitary adenomas is a marker for invasive growth hormone producing tumors.

In 2009, Jennings, et al., reported studies on Polynesian kindred with three members who presented with pituitary macro-adenoma in childhood or adolescence. These patients had AIP mutation R271W. They presented with headaches, visual disturbances, and excessive height. Features of acromegaly were absent. Acromegaly features include frontal bossing and overgrowth of hands and feet.

In 2010, Daly, et al., reported results of an international study on 96 patients with germline AIP mutations and pituitary adenomas. They noted that the patients were usually young and that the first symptoms occurred in children or adolescents. Males constituted 63.6% of the patients. The majority of the tumors were macro-adenomas. Excessive secretion of growth hormone occurred in 78% of tumors. In 13 of the 96 cases, prolactin secretion was excessive; 7 tumors were nonsecreting.

In 2010, Chahal, et al., reported that they had identified 49 different AIP mutations in patients with familial-inherited pituitary adenomas. These included deletions, insertion, segmental duplications, nonsense, missense, and splice site mutations. In addition, whole

exon deletion or deletion of the entire AIP gene occurred in some patients. They noted that in the cohort of families they studied, approximately 30% of the individuals who carried a germline AIP mutation presented with pituitary tumors.

Chahal, et al. (2010), concluded that the physiological role of the arylhydrocarbon receptor (ARH) likely includes cell proliferation and differentiation. ARH occurs in the cytoplasm as a multiprotein complex with AIP, HSP90, and co-chaperone p23. This complex binds to xenobiotics. It is transferred to the nucleus, where it binds with hypoxia inducible factor HIF1b, also known as ARNT. They noted that several proteins involved in the regulation of hypoxia-induced proteins play a role in tumor susceptibility. These include succinate dehydrogenase fumarate hydratase and Von Hippel Lindau proteins. (These proteins are discussed further in Chapter 5, “Pathways, Phenotypes, and Phenocopies.”)

Evidence also indicates that AHR binds to ubiquitin ligase and plays a role in the degradation of estrogen and androgen receptors.

### ***Next-generation sequencing***

Key elements in next-generation sequencing are the miniaturization of sequencing reactions, the sequencing of short fragments of DNA bound to solid matrices, and real-time photo-capture of the sequencing reactions. Different companies have developed a number of different platforms for next-generation sequencing; precise methods for capturing fragments and sequencing vary depending on the sequence platform used. In some cases, fragments are ligated with specific oligonucleotides at each end, and these are hybridized to matching oligonucleotides on the solid matrix sequencing platform. In other cases, DNA fragments are biotin labeled and then captured with streptavidin beads; the beads are subsequently captured on the sequencing platform. Detecting the sequencing reaction is enabled through use of nucleotides A G C T, each labeled with different colors

of fluorescent dyes, and fluorescent images are captured. The flow cells used as sequencing platforms are partitioned into several channels so that a number of samples can be simultaneously analyzed.

Next-generation sequencing is referred to as massively parallel sequencing because thousands of short sequences are read at the same time and each sequence is read optimally about 30 times. Sequence data generated on the platform is submitted to a computer and is subsequently aligned to reference sequence.

Whole-genome sequencing in humans generates a very large amount of data for analysis. In determining disease-causing mutations in humans, capturing specific genomic regions and capturing the human exome represent methods that reduce the complexity of the data analysis.

Data analysis may be further simplified by prioritizing genomic regions or genes to be studied through filtering at the levels of bioinformatic analysis. Roach, et al. (2010), carried out studies on two siblings affected with an autosomal recessive disorder called Miller syndrome and their parents. They applied previous information on polymorphic markers and haplotype analysis to select areas of the genome for computational analysis following whole-genome sequencing. They focused their analysis on 22% of the genome where both affected offspring inherited the same genomic segments from both parents.

In reviewing the application of next-generation sequencing to the discovery of rare gene defects that cause Mendelian diseases, Ng, et al. (2010b), emphasized that linkage information may narrow the genomic region that needs to be sequenced or computationally analyzed.

Additional studies are usually required to definitively establish the significance of sequence alterations that are likely candidates for disease causation. Significant changes include chain termination substitution, deletions, and nonsynonymous nucleotide substitution that cause amino acid substitutions that likely alter the structure or

localization of a gene product. Follow-up studies include applying PCR (polymerase chain reaction amplification) and Sanger sequencing. Downstream follow-up includes biochemical and physiological studies.

### ***Whole-genome sequencing and the discovery of mutation leading to Charcot-Marie-Tooth Neuropathy***

Charcot-Marie-Tooth neuropathies (CMT) are a group of disorders characterized by peripheral motor and sensory neuropathies with different modes of inheritance, including autosomal dominant, autosomal recessive, and X-linked inheritance. They are characterized clinically by symmetric distal polyneuropathy. Progressive muscle weakness and atrophy occur particularly in the peroneal muscles, leading to foot-drop and abnormal gait.

CMT results from mutations in at least 39 different genes. Lupski, et al. (2010), reported that mutation testing is available in the United States for 15 of the 39 genes and costs \$15,000.

Lupski, et al., reported results of whole-genome sequencing and follow-up analysis on a family with CMT. Sequencing yielded 89 gigabytes of sequence data; the depth of coverage was 30, indicating that each base was sequenced 30 times. The sequence derived from the affected proband was compared to the human reference genome sequence, and differences between the two were documented. These differences included single base substitutions, small deletions, and insertions and copy number changes.

Copy number variants were examined by array comparative hybridization and by sequence analysis. No copy number variants were identified that impacted genes known to play roles in CMT.

Lupski, et al., focused attention on the 9,069 single nucleotide substitutions that led to nonsynonymous codon changes. Of these 121 were nonsense mutations. Data was examined to search for single nucleotide substitutions in the proband that impacted genes known

to cause neuropathic conditions. Two nucleotide substitutions were found in the SH3TC2 gene, one missense mutation that led to Y169H and one nonsense mutation R954X.

Lupski, et al., noted that mutations in the SH3TC2 gene were previously described in Eastern European, Turkish, and Spanish gypsy patients and that the R954X mutation was present in some of these patients.

In the family reported by Lupski, et al., the R954X mutation occurred in one parent of the proband and the Y169H mutations occurred in the other parent. The proband had three siblings affected with CMT, and all three carried both the R954X and Y169H mutations. Subclinical phenotypes revealed by neurophysiological studies occurred in heterozygotes for each of the mutations.

### ***Earlier studies on SH3TC2 and Charcot-Marie-Tooth neuropathy***

Demyelinating autosomal recessive CMT was mapped to chromosome 5q23-q33 through homozygosity mapping in consanguineous families. Subsequently, sequence analysis of genes in this region revealed mutations in the SH3TC2 gene (Azzedine, et al., 2006). In ten consanguineous families, eight different mutations were found. Six of the mutations occurred in exon 11. Two cases had R954X mutations. Azzedine, et al., noted that the patients had foot deformities and that spinal abnormalities (kyphoscoliosis) also occurred.

In an analysis of 23 English patients with autosomal recessive CMT, Houlden, et al. (2009), identified 5 patients with SH3TC2 mutations. Affected members in four families were homozygous for the R954X mutation, and in one family, the affected members were compound heterozygotes for the R954X mutation and E657K mutation. Houlden, et al., noted clinical heterogeneity in the families with

respect to the severity of neuropathy. Neuropathology on sural nerve biopsies revealed demyelinating fibers and an abnormal Schwann cell that formed onion bulb–like structures.

The SH3TC2 protein localizes to the cellular plasma membrane and to the membrane of vesicles in the endocytic membrane trafficking pathway (Lupo, et al., 2009).

Disruption in this pathway apparently leads to impaired interactions between Schwann cells and axons.

Roberts, et al. (2010), demonstrated interaction between SH3TC2 protein and the membrane small GTPase Rab 11. Rab11 is known to regulate the recycling of internalized membranes in the endosomal pathway.

### ***Exome sequencing***

Analysis is simplified when exome sequencing rather than whole-genome sequencing is carried out, because the exome constitutes approximately 1% of the genome, approximately 30 megabases (Mb). Ng, et al. (2010a), carried out exome sequencing on four unrelated individuals affected with Miller syndrome. Clinical features in Miller syndrome include micrognathia, cleft lip and palate, and eye and limb anomalies. To derive the sequence, Ng, et al., used array-based capture of exomes. Their study was initiated using DNA from affected siblings, which facilitated a search for changes in regions where siblings had identical polymorphisms and nucleotide substitutions. They identified a mutation in the dihydro-otate dehydrogenase gene DHODH. They subsequently carried out studies on individuals affected by Miller syndrome in three unrelated families. Sequence analysis established that these affected individuals were compound heterozygotes for DHODH mutations. The DHODH gene product plays a role in pyrimidine metabolism.

Next-generation sequencing continues to shed light on DNA sequence changes and their potential roles in diseases. Bioinformatic analysis of sequence data is challenging, and continued development of resources for analysis is important. Equally important will be downstream analysis of the biochemical and physiological effects of sequence changes.

# INDEX

## Numbers

1p13 region, 13  
1p13.5 region, 13  
1p32-p36 chromosome, 66  
1q21.1 region, 4  
2p16-21 chromosome, 162  
3HMGCoA reductase, 106  
3p25-22 chromosome, 162  
4q25-q28 chromosome, 66  
5q23-q33 chromosome, 20  
9p21.3 region, 13  
11q12-q13 chromosome, 15  
11q22.23 chromosome, 60  
15q21.1-q21.2 chromosome, 64  
16p11.2 region, 3  
16p13.3 chromosome, 46, 54  
17p11.2 chromosome, 76  
17q12 region, 3  
19p13 region, 13  
19p13.12-p13.13 chromosome, 160  
22q11.2 chromosome, 98  
22q13 chromosome, 87

## A

Abeta accumulation in Alzheimer's disease, 103  
ABL oncogene, phosphorylation, 34  
acetylation, 34, 37  
ADNFLE (autosomal dominant nocturnal frontal lobe epilepsy), 94  
ADP ribosylation in histone modifications, 38  
adult-onset hemochromatosis, penetrance in, 54-55

## aggregates

analysis in ALS, 121  
formation of, 122  
in late-onset neurodegenerative diseases, cellular transfer, 126-127  
in Parkinson's disease, 123

## aging

DNA methylation and, 51  
effect on penetrance, 53  
neurodegenerative diseases and, 123

## AIP (arylhydrocarbon receptor interacting protein), 15-17

## AKT gene, 72

## AKT1 gene, 99

## ALK-EML4 fusion genes, 140, 142

## alpha secretase activity in Alzheimer's disease treatment, 114-115

## alpha-thalassemia, 45, 57

## alpha2 globin gene, long ncRNA, 29

## ALS (amyotrophic lateral sclerosis), 117

FUS in, 119-120

genotype-phenotype studies, 121

protein aggregate analysis, 121

SOD1 in, 118

TDP43 in, 118-119

## alternate splicing, 26, 30-31

## Alzheimer's disease

alpha and beta secretase activity, 114-115

amyloid aggregates as prions, 113

## APOE

*APOE4 versus APOE3, 104*

*APP and, 103*

*brain lipids and, 105*

*polymorphism, 102*

APP and, 101-102  
 biomarkers in detection of, 108  
 brain cholesterol homeostasis and treatment, 115  
 brain cholesterol, role of, 107-108  
 cholesterol biosynthesis in brain, 105-106  
 clusterin and, 110  
 GWAS and, 109-112  
 molecular-based treatment, 113  
 neurodegeneration in, 108-109  
 prion protein, role of, 112, 116  
 questions regarding treatment, 116-117  
 Amish population studies in genomic medicine, 154-155  
 AMPAR receptors, 82  
 AMPK gene, 72, 76  
 amyloid aggregates as prions, 113  
 amyotrophic lateral sclerosis. *See* ALS  
 Angelman syndrome, UBE3A gene and, 95-96  
 ANTI1 protein, MECP2 interaction, 47  
 antiepilepsy drugs, 95  
 APC gene, 138  
 APOA1 gene, 14  
 APOB gene, 14  
 APOE gene  
   in Alzheimer's disease, 102  
   brain lipids and, 105  
   in coronary heart disease, 14  
 APOE3 protein versus APOE4 protein in Alzheimer's disease, 104  
 APP (amyloid precursor protein), Alzheimer's disease and, 101-103  
 arginine deaminase in cancer treatment, 147  
 ARH (arylhydrocarbon receptor), 17  
 ARID1A gene, 48  
 ARNT protein, 17  
 arylhydrocarbon receptor interacting protein (AIP), 15, 17  
 asparaginase in cancer treatment, 147  
 ataxia telangiectasia (AT), DNA damage and repair, 60  
 ATM gene, 60-61, 149  
 ATM-ATR response, 63  
 ATP-dependent chromatin remodeling, 41-42  
 ATRX gene, 45

autism  
   gene mutations in, 96-97  
   mTOR signaling pathway and, 89  
   in phenotypically discordant monozygotic twins, 51

## B

barrier insulation, 40  
 basal cell nevus syndrome, 133  
 base excision repair, 149  
 BCL2 gene, 51  
 BCL11A oncogene, 160-161  
 BDNF (brain-derived neurotrophic factor), 32  
   in obesity, 12  
   in synaptic plasticity, 91  
 Bernard, Claude, 169  
 beta-catenin signaling pathway  
   hepatocellular carcinoma, 139  
   identifying oncogenes in, 137-139  
 BHD (Birt Hogge Dube) syndrome, 76  
 bioenergetic regulation in cancer treatment, 145-148  
 biology of tumor cells, analysis of, 140-144  
 biomarkers  
   in Alzheimer's disease detection, 108  
   fusion genes as, 140-142  
 bipolar disease, 158  
 Bird, Adrian, 37  
 Birt Hogge Dube syndrome, 76  
 Blackfan Diamond anemia, 78  
 BMAL1 protein, circadian rhythms and, 44  
 BRAF gene, 135  
 brain. *See also* cognitive impairment  
   cholesterol biosynthesis in, 105-106  
   cholesterol in, role in Alzheimer's disease, 107-108  
   lipid metabolism in, 105  
 brain cholesterol homeostasis in Alzheimer's disease treatment, 115  
 brain development, microRNA and, 32  
 BRCA2 gene, 62  
 BRD (bromodomain) proteins, 42  
 BRG1 (Brahma) gene, 42

**C**

C282Y gene, 54

cancer. *See also* renal cancer; tumors

bioenergetic regulation in

treatment, 145-148

biology analysis of tumor cells,

140-144

driver mutations, treatments for,

134-136

epithelial ovarian cancer, 162

fusion genes as biomarkers, 140-142

genome instability and, 149-152

hedgehog signaling pathway,

133-134

metabolic targets in treatment,

146-147

molecular networks, role in

treatment, 136-139

molecular studies and treatments,

131-134

mTOR pathway and, 72

personalized treatments for, 143-144

prostate cancer, ETS transcription

factors in, 143

systems biology modeling, 144-145

telomeres and, 150

cardiovascular system, microRNA  
and, 32-33

CBP (CREB binding protein), 45

CDK8 gene, 138

CDKAL1 gene, 11-12

CDKN2A gene, 11-12

CDKN2B gene, 11-12

cDNA, generating from

polyadenylated RNA, 25

centrosomes, defined, 63

CENP152 gene, 64

CETP gene, 14

chaperone proteins, 79

Charcot-Marie-Tooth neuropathies  
(CMT), 19-21

CHARGE syndrome, 42

CHD proteins, chromatin  
remodeling, 42

cholesterol biosynthesis in brain,  
105-106

cholesterol homeostasis, 105

cholesterol in brain, role in  
Alzheimer's disease, 107-108

cholesterol levels, GWAS on, 13-14

CHRNA4 gene, 94

CHRN2 gene, 94

chromatin

ATP-dependent chromatin  
remodeling, 41-42

in Fanconi anemia, 61-65

insulators and, 39-40

interphase human chromosome  
analysis, 40-41

neuronal stimulation and, 42-43

chromatin modifier genes, 48-49

chromosome instability, cancer and,  
149-152

chronic myelogenous leukemia,  
Philadelphia chromosome in, 132

chronic obstructive pulmonary  
disease (COPD), environmental  
factors in, 58-59

circadian rhythm, metabolism and,  
43-45

circulating tumor cells, biology  
analysis, 140-144

clinical trials of molecular-based  
cancer treatments, 137

CLOCK gene, 43-45

clonal heterogeneity, 145

clusterin, Alzheimer's disease and, 110

CMT (Charcot-Marie-Tooth  
neuropathies), 19-21

cognitive impairment

synaptic activity and, 81

*dendritic spines*, 82

*FMRP gene*, 84-87

*gene expression*, 83-84

*neuroligins*, 82

WDR62 gene mutations and, 92-93

cohesins, defects in, 45

congenital malformations, ribosome  
biogenesis pathway and, 78-79

consciousness, EEG studies on, 93

COPD (chronic obstructive  
pulmonary disease), environmental  
factors in, 58-59

copy number variants, 3-5, 19

core duplicons, 2

Cornelia de Lange Syndrome, 45

coronary heart disease, GWAS on,  
13-14

CpG islands, methylation, 38

CR1 gene, 58

CREB (cyclic AMP response element binding protein), 32, 45, 158  
 Crestor, 165  
 CTCF protein, 40-41  
 Cushing, Harvey, 15  
 CYFIP1 gene, 85  
 CYP46A1 gene, 107  
 cytosine-to-thymidine (C-to-T) transversions, 66

## D

D4Z4 DNA element, gene  
 expression, 47  
 data integration  
 in functional genomics, 161  
 systems biology and medicine, 163  
 deletions, phenotypic abnormalities and, 3-5  
 dendritic spines, synaptic activity and, 82  
 dephosphorylation of FMRP gene, 86  
 DGCR8 gene, 31  
 DHHC8 gene, 98  
 DHHC9 gene, 99  
 DHHC15 gene, 99  
 DHHC17 gene, 98  
 DHHC21 gene, 98  
 DHODH gene, 21  
 diabetes. *See* type 2 diabetes  
 discovery-driven approaches in system analyses, 161  
 disease-specific mutations. *See* genome wide association studies  
 disulfide bond formation, 34  
 DNA damage and repair  
 cancer and, 149-150  
 environmental factors in, 59-61  
 in Fanconi anemia, 61-65  
 in Seckel syndrome, 63-64  
 signatures of environmentally induced damage, 65-66  
 ubiquitin ligases and, 64-65  
 DNA methyl transferases, methylation, 39  
 DNA methylation, aging and, 51  
 donor identification by genotyping, 155  
 double-stranded DNA breakage, ubiquitin ligases and, 64-65

driver mutations in cancer, treatments for, 134-136  
 DROSHA RNAase complex, 31  
 drug metabolism in personalized medicine, 165  
 Duffy blood group, malaria and, 58  
 duplications, segmental, 2-5

## E

E657K gene, 20  
 EEG studies on consciousness, 93  
 EGFR (epidermal growth factor receptor), 132  
 EML4-ALK fusion genes, 140-142  
 ENCODE project, transcription, 23-24  
 alternate splicing and, 26, 30-31  
 gene expression variation, 26-27  
 genes, definition of, 24  
 LINC RNA, 29  
 long ncRNA, 28-29  
 ncRNA, 27-28  
 RNA sequencing and, 25  
 Encyclopedia of DNA Elements. *See* ENCODE project  
 enhancer block insulation, 39  
 enriched environments, epigenetics in, 43  
 environmental factors in gene expression, 45-46, 53  
 DNA damage and repair, 59-61  
 in Fanconi anemia, 61-65  
 penetrance, impact on, 53-59  
 signatures of, 65-66  
 environmental signals, integrating with growth, 73  
 epidermal growth factor receptor (EGFR), 132  
 epigenetics  
 ATP-dependent chromatin remodeling and, 41-42  
 chromatin modifier genes in tumors, 48-49  
 circadian rhythm and, 43-45  
 defined, 37  
 DNA methylation and aging, 51  
 in enriched environments, 43  
 environmental cues and gene expression, 45-46  
 histone modifications, 37-38

insulators and, 39-40  
interphase human chromosome analysis in, 40-41  
MECP2 gene expression, 46-47  
methylation and, 38-39  
neuronal stimulation and, 42-43  
parent of origin allelic expression, 49-50  
in phenotypically discordant monozygotic twins, 50-51  
stem cell biology and, 52  
therapeutic interventions based on, 47-48  
variant histones, 49  
epilepsy, 94-95  
epistasis, 79  
epithelial ovarian cancer, 162  
erasers in chromatin remodeling, 42  
ERBB4 protein, 99-100  
erlotinib, 132  
erythroblasts, gene expression in, 40  
ETS transcription factors, 143  
ETV1 gene, 142  
ETV6–RUNX1 translocation, 145  
excitatory synapses, 82, 85  
exome sequencing, 21-22  
expression of fetal hemoglobin, persistence of, 158, 161

## F

familial pituitary adenoma, GWAS on, 15-17  
FANCA gene, 62  
FANCD1 gene, 62  
FANCD2 gene, 62  
FANCF gene, 62  
FANCI gene, 62  
Fanconi anemia, 61-65, 166  
fascio-scapulo-humeral muscular dystrophy (FSHD), 47  
FAT domains, 88  
fetal hemoglobin, persistence of expression, 158, 161  
Finnish disease heritage, 8, 10  
FKBP domains, 16  
FLCN gene mutations, 76  
FMR1 gene, 69  
FMRP gene, 84-87  
foldback inversion, 152  
folliculin gene mutations, 76

FOX3A gene, 72  
Fragile X mental retardation protein. *See* FMRP gene  
Fragile X syndrome  
  GABA neurotransmitter function in, 86  
  mGluR gene, 86  
  mTOR signaling pathway and, 89  
  palmitoylation and, 99  
  phenocopies in, 68-69  
FRB domain, 88  
French-Canadian Leigh syndrome, 162-163  
FSHD (fascio-scapulo-humeral muscular dystrophy), 47  
FTO gene, 12-13  
fumarate hydratase, germline mutations of, 76  
functional genomics, data integration in, 161. *See also* genomic medicine  
FUS positive inclusions, 121  
FUS protein, 119-121  
fusion genes as biomarkers, 140-142

## G

G6PD enzyme, malaria survival, 56  
GABA neurotransmitter function, 86  
gamma synchrony, 93  
gastrointestinal stromal tumors, treatment of, 142  
GCK gene, 11-12  
GCKR gene, 14  
GDC-0449 molecule, 133  
gefitinib, 132  
Gene Cards, 169  
gene expression  
  environmental factors, 53  
    *DNA damage and repair*, 59-61  
    *in Fanconi anemia*, 61-65  
    *penetrance, impact on*, 53-59  
    *signatures of*, 65-66  
  linking environmental cues to, 45-46  
  MECP2 impact on, 46-47  
  parent of origin allelic expression, 49-50  
  in schizophrenia, 99-100  
  synaptic activity and, 83-84  
  variation in  
    *microRNA and*, 31-33  
    *RNA sequencing and*, 26-27

gene mutations in autism, 96-97  
 genes, definition of, 24  
 Genetests, 169  
 genetics, benefits of studying, 167-169  
 genome instability, cancer and, 149-152  
 genome wide association studies. *See* GWAS (genome wide association studies)  
 genomic integrity, DNA damage response and, 63-64  
 genomic medicine. *See also* functional genomics  
   Amish population studies in, 154-155  
   defined, 153  
   integration with systems biology, 163  
   pharmacologic response to, 158  
   in PKD (polycystic kidney disease), 157  
   therapy development, 156-157  
 genotype-phenotype studies in ALS, 121  
 genotyping, donor identification by, 155  
 germline mutations, 73  
   in Birt Hogge Dube syndrome, 76  
   of fumarate hydratase, 76  
   succinate dehydrogenase mutations, 77  
   in Von Hippel Lindau syndrome, 74-75  
 GKAP gene, 87  
 Gleevec, 132  
 glucose metabolism, 145  
 glutamine in cancer treatment, 147  
 glycolysis pathway, 71  
 glycosylation, 34  
 growth, integrating with environmental signals, 73  
 GSK3 gene, 158  
 GSK3B gene, 73  
 guanine-to-cytosine (G-to-C) transversions, 65  
 guanine-to-thymidine (G-to-T) transversions, 65  
 GWAS (genome wide association studies), 10  
   Alzheimer's disease, 109-112  
   coronary heart disease, 13-14  
   familial pituitary adenoma, 15-17  
   type 2 diabetes and obesity, 10-13

## H

H2AK5ac histone, 40  
 H2AX histone, 49, 61  
 H2AZ histone, 49  
 H3K27me histone, 42  
 H3K27Me3 histone, 40  
 hair root analysis, advantages of, 6  
 HAR1 gene, 28  
 HAT (histone acetyl transferase), 38  
 HbC allele, 56  
 HbE allele, 57  
 HbF (fetal hemoglobin), persistence of expression, 158, 161  
 HbS allele, 56  
 HDAC (histone deacetylase), 38  
 HEAT domains, 88  
 hedgehog signaling pathway in cancer, 133-134  
 hematopoietic stem cell transplants, donor identification for, 155  
 hemochromatosis, penetrance in, 54-55  
 hemoglobinopathies, 159-161  
 hepatocellular carcinoma, 139  
 HERC2 gene, 6, 64-65  
 Herceptin, 132  
 hereditary disorders, population history and, 8-10  
 Heterozygosity in regional populations, 7  
 HHEX gene, 11  
 HIF (hypoxia inducible factor), 74  
 HIF1 gene, 71  
 HIF1b protein, 17  
 high-throughput RNA sequencing, 25  
 histone acetyl transferase (HAT), 38  
 histone acetylation, 48  
 histone deacetylase (HDAC), 38, 47-48  
 histone modifications, 37-39  
 histones, variants, 49  
 HLA complex, P4 medicine and, 164  
 HLA gene variants in diabetes, 12  
 HNF1A gene, 11-12  
 HNF1B gene, 11-12  
 HNF4A gene, 11  
 homologous recombination, 149  
 homozygosity  
   mapping, 8-9  
   in Saqqaq culture, 6

HOTAIR long ncRNA, 28  
 HOXC gene, 28  
 HOXD gene, 28  
 HSP70 chaperone protein, 79  
 HSP90 chaperone protein, 79  
 Hunterian skeleton, familial pituitary adenoma in, 15  
 huntingtin gene, 68  
 Huntington's chorea, palmitoylation and, 98  
 Huntington's disease  
   genetic studies on, 156  
   phenocopies in, 68  
 hydroxylation, 34  
 hyperlipidemia, treatment for, 14  
 hypothesis-driven approaches in system analyses, 161  
 hypoxia inducible factor (HIF), 74

## I

IDH1 mutation, 148  
 IDH2 mutation, 148  
 IL28B gene, 165  
 imatinib, 132, 142  
 imetelstat, 150  
 indisulam, 147  
 induced pluripotent stem cells (iPS cells), 165  
 inducible proteopathies, 127  
 inhibitors  
   in cancer treatment, 147  
   on mTOR protein, 90  
 inhibitory synapses, 82  
 INS gene, 12  
 insulators, 39-40  
 integrative genomics  
   explained, 161  
   in French-Canadian Leigh syndrome research, 162-163  
 interleukin4 receptor gene, 79  
 interleukin13 receptor gene, 79  
 interphase human chromosome analysis, 40-41  
 intrachromosomal duplications, 2  
 IPF gene, 11  
 iPS cells (induced pluripotent stem cells), 165  
 iron levels in hemochromatosis, 54-55  
 IRS1 gene, 11-12  
 isocitrate dehydrogenases IDH1 and IDH2, 148

## J-K

JPH3 gene, 68  
 K288X gene, 160  
 KCNJ11 gene, 11-12  
 KIN domain, 88  
 KIT oncogene activation, 142  
 KLF1 gene, 160-161  
 Kuru disease, 127

## L

lactate in cancer cells, 145  
 large intervening noncoding RNA (LINC RNA), 29  
 late-onset neurodegenerative diseases  
   aggregates, transfer methods of, 126-127  
   ALS. *See* ALS (amyotrophic lateral sclerosis)  
   Alzheimer's disease. *See* Alzheimer's disease  
   FUS and TDP43 in, 120-121  
   inducible proteopathies, 127  
   mitochondria and, 123-124  
   mitochondrial function regulation, 128-129  
   Parkinson's disease, aggregates in, 123  
   prion protein conformations, 127-128  
   protein aggregates, formation of, 122  
   protein modification in, 124-126  
   questions regarding, 129-130  
   SIRT1 gene, role in treatment, 129  
 LDL cholesterol levels, GWAS on, 13-14  
 LDLR gene, 14  
 Leigh syndrome, 162-163  
 LINC RNA, 29  
 linkage disequilibrium, 9-10  
 lipid metabolism, APOE and, 105  
 lipid rafts, 106, 115  
 lithium treatment, 158  
 liver damage, 139  
 living cells, interphase human chromosome analysis in, 40-41  
 long ncRNA, 28-29  
 long Q-T syndrome, 165  
 Lou Gehrig's disease. *See* ALS  
 LPL gene, 14  
 LRPPRC gene, 163  
 LUC7L gene, 29

**M**

malaria, genetic factors in survival, 55-58  
 MAP3K8 kinase, 135  
 maple syrup urine disease, 154  
 massively parallel sequencing. *See* next-generation sequencing  
 maturity-onset diabetes of the young (MODY), 10-12  
 meclizine, 156  
 MECP2 gene, 45-47  
 medicine  
   integration with systems biology, 163  
   P4 medicine, 164-165  
 medulloblastomas, 133  
 melanocortin 4-receptor gene, 11  
 melanomas, treatment of, 135  
 mental retardation syndrome, 45  
 MET proto-oncogene, 75  
 metabolic targets in cancer  
   treatment, 146-147  
 metabolism, circadian rhythm and, 43-45  
 metastatic tumors, biology analysis, 140-144  
 methyl-binding domains, 39  
 methylation, 34, 38-39  
   aging and, 51  
   in histone modifications, 38  
 mGluR gene, 86-87  
 mGluR receptors, 82  
 mGluR1 gene, 85  
 mGluR5 gene, 85  
 microcephaly, WDR62 gene  
   mutations and, 92-93  
 microRNA  
   cardiovascular system and, 32-33  
   gene expression and, 31-33  
 Miller syndrome  
   exome sequencing of, 21  
   next-generation sequencing and, 18  
 miR29 gene, 33  
 miR208b gene, 33  
 miR499 gene, 33  
 miR2081 gene, 33  
 miRNA1 gene, 32  
 miRNA29 gene, 32  
 miRNA-132 gene, 32  
 miRNA-145 gene, 32  
 miRNA-219 gene, 32

mitochondria, neurodegenerative diseases and, 123-124  
 mitochondrial diseases, integrative genomics and, 163  
 mitochondrial function regulation in neurodegenerative disease  
   treatment, 128-129  
 MODY (maturity-onset diabetes of the young), 10-12  
 molecular networks, role in cancer  
   treatment, 136-139  
 molecular studies in cancer  
   treatments, 131-134  
 molecular-based treatment of Alzheimer's disease, 113  
 monogenic epilepsy, 94  
 monozygotic twins, phenotypical discordance in, 50-51  
 motor neuron disease. *See* ALS  
 mRNA  
   FMRP binding, 84  
   repressing translation of, 83  
 MTHFD1L gene, 111-112  
 MTNR1B gene, 11-12  
 mTOR protein  
   function of, 89-90  
   inhibitors on, 90  
   regulating expression by stress, 72-73  
   role in diseases, 69-72  
   signaling pathways on, 87  
   structure of, 88  
   synaptic activity and, 87  
   TSC1 TSC2 complex and, 90  
 mTORC1 gene, 89-90  
 mTORC2 gene, 88-90  
 mutations. *See* variations  
 myocardial infarction, treatment for, 14

**N**

NAD-dependent histone deacetylases  
   in epigenetics, 43-45  
 NAMPT enzyme, 44  
 ncRNA, 27-29  
 Neanderthal fossils, sequence  
   variation in, 5-6  
 NEGRI gene, 12  
 neprilysin in Alzheimer's disease, 114  
 NEUROD1 gene, 11

neurodegeneration in Alzheimer's disease, 108-109  
 neurodegenerative disease. *See* late-onset neurodegenerative diseases  
 neurofibromatosis, mTOR signaling pathway and, 89  
 neuroligins, synaptic activity and, 82  
 neurological diseases, palmitoylation and, 98-99  
 neuronal stimulation, 42-43  
 next-generation sequencing, 17-19.  
*See also* exome sequencing; whole-genome sequencing  
 NFS1 gene, 11  
 Niemann Pick disease type C (NPC), 106  
 NIPBL gene, 45  
 nitrosative stress, protein modification and, 124-126  
 NMDAR receptors, 82  
 non-protein coding RNA (ncRNA), 27-28  
 LINC RNA, 29  
 long ncRNA, 28-29  
 nonhomologous end joining, 149  
 NRG1 gene in schizophrenia, 99-100  
 nucleosome mobility, 41  
 nucleotide excision repair, 59, 149

## O–P

obesity, GWAS on, 10-13  
 OCA2 gene, 6  
 oncogenes  
 activation, targeting, 142  
 addiction, 139  
 identifying, 137-139  
 oxidative stress, protein modification and, 124-126  
 P4 medicine, 164-165  
 p400 gene, 42  
 palmitoylation, 34  
 explained, 97-98  
 in late-onset neurodegenerative diseases, 124-126  
 neurological diseases and, 98-99  
 post-translational protein modification, 33-35  
 pancreatic cancer, genome instability and, 151-152

parent of origin allelic expression, 49-50  
 Parkinson's disease, aggregates in, 123  
 PARP inhibitors, 136  
 pathways  
 in cancer driver mutations, 134-135  
 hedgehog signaling pathway in cancer, 133-134  
 identifying oncogenes in, 137-139  
 mTOR pathway  
*regulating expression by stress*, 72-73  
*role in diseases*, 69-72  
 in renal cancer, 73  
*Birt Hogge Dube syndrome*, 76  
*MET proto-oncogene*, 75  
*succinate dehydrogenase mutations*, 77  
*tricarboxylic acid metabolism*, 76  
*tuberous sclerosis*, 75  
*Von Hippel Lindau syndrome*, 74-75  
 ribosome biogenesis pathway, congenital malformations and, 78-79  
 PDGF gene, 74  
 PDK1 gene, 146  
 PDK1 inhibitors, 147  
 penetrance  
 defined, 53  
 factors affecting, 53-59  
*in COPD*, 58-59  
*in hemochromatosis*, 54-55  
*in malaria*, 55, 57-58  
 Pennisi, Elizabeth, 23  
 pentose phosphate pathway, 71  
 pericentric duplications, 2  
 pericentrin in centrosomes, 63  
 persistence of expression of fetal hemoglobin, 158, 161  
 personalized cancer treatments, 143-144  
 personalized medicine, 153, 164-165  
 pharmacologic response to genomic medicine, 158  
 PharmGKB (Pharmacogenomics knowledge base), 169  
 phenocopies  
 defined, 67  
 in Fragile X syndrome, 68-69  
 in Huntington's disease, 68

phenome-interactome networks, 162  
 phenotypic abnormalities, structural variation and, 3-5  
 phenotypical discordance in monozygotic twins, 50-51  
 Philadelphia chromosome, 132  
 phosphorylation, 34  
   of FMRP gene, 86  
   in histone modifications, 38  
 PI3K pathway, 75  
 PICALM gene, 110  
 PKD (polycystic kidney disease), 157  
 pluripotent stem cells, 52, 165  
 PLX4032 treatment, 135  
 POLR1D gene, 78  
 polyadenylated RNA, conversion to cDNA, 25  
 polyadenylation, alternate splicing and, 30-31  
 polycystic kidney disease (PKD), 157  
 population history, hereditary disorders and, 8-10  
 populations, sequence variation in regional populations, 7-8  
 post-synaptic density (PSD), 82  
 post-translational protein modification, 33-35  
 PPAR $\gamma$  gene, 11-12  
 presenilins in Alzheimer's disease, 101  
 primary tumors, biology analysis, 140-144  
 prion protein  
   amyloid aggregates as, 113  
   conformations, 127-128  
   role in Alzheimer's disease, 112, 116  
 PRNP gene, 68  
 prostate cancer, ETS transcription factors in, 143  
 protein aggregates. *See* aggregates  
 protein modification. *See*  
   palmitoylation  
 proteome, defined, 33  
 proteostasis network, 79  
 PSD (post-synaptic density), 82  
 PSD95 domain, 87  
 PSEN1 gene, 101  
 PSEN2 gene, 101  
 PTCH1 gene, 133  
 PTPN22 gene, 12  
 pyruvate in cancer cells, 145

## Q-R

Q14X gene, 16  
 quercetin, 139  
 R271W gene, 16  
 R954X gene, 20  
 Rab11 membrane, 21  
 rapamycin, 90  
 Raptor protein, 88  
 readers in chromatin remodeling, 42  
 red-cell membrane proteins, malaria and, 58  
 REDD1 gene, 72  
 REDD2 gene, 72  
 regional populations, sequence variation in, 7-8  
 renal cancer, pathways in, 73  
   Birt Hogge Dube syndrome, 76  
   MET proto-oncogene, 75  
   succinate dehydrogenase mutations, 77  
   tricarboxylic acid metabolism, 76  
   tuberous sclerosis, 75  
   Von Hippel Lindau syndrome, 74-75  
 resistance to cancer treatments, 135  
 Rett syndrome, 45-46  
 RhebGTP gene, 71  
 ribosome biogenesis pathway, congenital malformations and, 78-79  
 Rictor protein, 88  
 RNA sequencing, 25-27  
 RNF168 gene, 64-65  
 RNG105 gene, 86  
 RORA gene, 51  
 ROS (reactive oxygen species), 123  
 rosiglitazone, 104

## S

Saqqaq individuals, sequence variation in, 6-7  
 schizophrenia  
   gene expression in, 99-100  
   palmitoylation and, 98  
 SCID (severe combined immune deficiency), 155  
 SCRAP gene, 42  
 Seckel syndrome, DNA damage response and, 63-64  
 secretase levels in Alzheimer's disease treatment, 114-115

segmental duplications, 2-5  
 sequence variation, 4  
   exome sequencing, 21-22  
   GWAS. *See* GWAS (genome wide association studies)  
   in hereditary disorders and population history, 8-10  
   homozygosity mapping, 8-9  
   in Neanderthal fossils, 5-6  
   next-generation sequencing, 17-19  
   in regional populations, 7-8  
   in Saqqaq individuals, 6-7  
   whole-genome sequencing, 18-20  
 serotonin pathways, lithium treatment and, 158  
 severe combined immune deficiency (SCID), 155  
 SH2B1 gene, 12  
 SH3TC2 gene, 20-21  
 SHANK proteins, PSD95 domains and, 87  
 “Shining a Light on the Genome’s Dark Matter” (Pennisi), 23  
 signaling pathways on mTOR protein, 87  
 signatures of environmentally induced DNA damage, 65-66  
 single nucleotide polymorphisms (SNPs), 1, 10  
 SIRT1 deficiency, 91  
 SIRT1 gene  
   circadian rhythms, 43, 45  
   role in neurodegenerative disease treatment, 129  
 SIRT1-7 proteins, 128  
 SLC25A3 gene, 30  
 SMCA1A gene, 45  
 SMC3 gene, 45  
 Smith Laemli Opitz syndrome, 106  
 SMO gene, 133  
 SNPs (single nucleotide polymorphisms), 1, 10  
 SOD1 gene, 118  
 somatic nuclear transfer stem cells, 52  
 SORT1 gene, 13  
 SORTL1 gene, 14  
 spinal muscular atrophy, 166  
 spines. *See* dendritic spines  
 SREBP gene, 71  
 statins in coronary heart disease treatment, 14

stem cell biology, epigenetics and, 52  
 stress, regulating mTOR pathway expression, 72-73  
 structural variation, 2-5  
 subtelomeric duplications, 2  
 succinate dehydrogenase mutations, 77  
 sumoylation in histone modifications, 38  
 SWI SNF protein family, 41  
 synaptic activity  
   in autism, 96-97  
   cognitive impairment and, 81  
     *dendritic spines*, 82  
     *FMRP gene*, 84-87  
     *gene expression*, 83-84  
     *neuroligins*, 82  
   mTOR function in, 89  
   mTOR protein and, 87  
   palmitoylation and, 97-98  
   UBE3A gene and Angelman syndrome, 95-96  
 synaptic plasticity, 43, 91-92  
 synaptophysin protein, 91  
 synthetic lethality, 136  
 synuclein aggregates in Parkinson’s disease, 123  
 system biology, integration with medicine, 163  
 systems biology modeling in cancer research, 144-145

## T

tamoxifen, 132  
 targeted mutational analysis of tumors, 143-144  
 tau phosphorylation, Alzheimer’s disease and, 108-109  
 TCF7L2 gene, 11-12  
 TCOF1 gene, 78  
 TDP43 gene, 117-121  
 telomerase, 150  
 telomere shortening, 150  
 telomeres, 46, 61, 150  
 TGFbeta gene, 59  
 therapeutic interventions based on epigenetics, 47-48  
 therapy development in genomic medicine, 156-157  
 TMPRSS2-ERG fusion genes, 143

tobacco smoke, DNA damage signatures from, 65-66  
**TORC1 gene**, 88  
transcription, ENCODE project, 23-24  
  alternate splicing and, 26, 30-31  
  gene expression variation, 26-27  
  genes, definition of, 24  
  LINC RNA, 29  
  long ncRNA, 28-29  
  ncRNA, 27-28  
  RNA sequencing and, 25  
transcriptomics, RNA sequencing and, 25  
transfer methods for aggregates in late-onset neurodegenerative diseases, 126-127  
translation, mTORC1 in, 89  
translation initiation, triggering, 85  
transversions in environmentally induced DNA damage, 65  
trastuzumab, 132  
Treacher Collins syndrome, 78-79  
treacle, 78  
treatments based on epigenetics, 47-48  
**TRF1 gene**, 61  
**TRF2 gene**, 61  
tricarboxylic acid metabolism, 76  
**TSC1 gene**, 70  
**TSC1 TSC2 complex**, 90  
**TSC2 gene**, 54, 70  
tuberous sclerosis  
  FMRP phosphorylation and, 86  
  mTOR pathway and, 70, 89  
  renal cancer and, 75  
  variable expressivity in, 54

tumors. *See also* cancer  
  chromatin modifier genes in, 48-49  
  DNA damage signatures in, 65-66  
  microRNA and, 32  
  mTOR pathway and, 72  
twins, phenotypical discordance in monozygotic twins, 50-51  
type 2 diabetes, GWAS on, 10-13

## U–V

**UBE3A gene**, 95-96  
ubiquitin ligases, double-stranded DNA breakage and, 64-65  
ubiquitination in histone modifications, 38  
variable expressivity, 54  
variant histones, 49  
variation. *See* sequence variation; structural variation  
**VEGF gene**, 74  
Von Hippel Lindau syndrome, 74-75

## W–Z

Warburg effect, 146  
**WDR62 gene**, 92-93  
whole-genome sequencing, 18-20  
Williams syndrome, 3  
writers in chromatin remodeling, 42  
**Y169H gene**, 20  
**YY1 protein**, 46-47