3.2 The genomic basis of medicine 218

3.2 The genomic basis of medicine 218

ESSENTIALS It is now possible to determine the entire DNA information content of living organisms—the genome. The completion of the human ref- erence DNA sequence has provided an enormous tool for genomic analyses and has enhanced our view of the genetic and genomic variation contributing to the genetic bases of disease. Several human genomic studies in diverse populations (e.g. the International HapMap Project (HapMap), the ENCyclopedia Of DNA Elements Project (ENCODE), and 1000 Genomes Project) have revealed that the tremendous amount of genetic variation in humans consists of two major types: nucleotide sequence variants and genomic structural changes. The contribution of rare variants and de novo mutations to disease, embodied in the Clan Genomics hypothesis, is of great clinical utility, has gathered extensive sup- portive data, and further aligns clinical practice with human biology in the context of evolution. The first phase of the studies on genetic variation in humans has been focused on single nucleotide polymorphisms and common variation. The large number of single nucleotide polymorphisms identified has enabled successful genome-wide association studies for disease susceptibility risk of complex traits (e.g. diabetes and cancer), but for the most part has had limited practical applications in clinical medicine. Technological developments enabling a higher-resolution analysis of the human genome have uncovered extensive submicroscopic structural variation, including copy-number variants. Copy-number variants involving dosage-sensitive genes result in several diseases and contribute to human diversity and evolution. An emerging group of genetic diseases have been described that result from DNA rearrangements (e.g. copy-number variants and other structural variations including copy-number neutral inversions and translocations), rather than from single nucleotide changes. Such conditions have been referred to as genomic disorders. Recurrent rearrangements of the human genome, or those of common size that contain the same genomic interval in different individual personal genomes and have clustered breakpoints, most frequently result from a mechanism of nonallelic homologous re- combination between region-specific low-copy repeats, or seg- mental duplications. Nonrecurrent rearrangements, or those for which breakpoints do not cluster and that are generally different in size and genome content among families, can result from non- homologous end-joining recombination mechanism. More recently, DNA replication mechanisms involving template switching have been shown to play a major role in

the origin of nonrecurrent re- arrangements; template switching also plays an important role in many Alu–Alu mediated events. Iterative template switches during replicative repair can result in complex genomic rearrangements. The development of array-based comparative genomic hybridiza- tion and single nucleotide polymorphism arrays have enabled high- resolution screening of genomic imbalances throughout the entire genome with the level of resolution depending only on the size and dis- tance between the arrayed interrogating probes. This has had tremen- dous clinical applications for both postnatal and prenatal diagnosis. Advances in massively parallel next-generation sequencing tech- nologies have led to development and research and clinical imple-mentation of exome sequencing and whole-genome sequencing, revolutionizing medical genetic diagnostics. These studies document a role for new mutations (either copy-number variants or single nu- cleotide variant) in sporadic disease traits. Such genome-wide vari- ation studies are beginning to yield insights into multilocus effects on disease trait manifestations. In the current postgenomic era both high-resolution genome analysis by chromosome microarray analyses and personalized dip- loid genomic sequencing applied to the study of inherited and com- plex traits promise a continued revolution in our understanding of normal physiology and the pathophysiology of disease heralding the genomic basis of medicine and the precision medicine initiative.
Introduction The elucidation of the DNA double helix establishing the chem- ical basis of heredity in 1953 and the determination of the correct number of human chromosomes 3 years later laid the fundamentals for the development of two major fields in human and medical gen- etics: clinical molecular genetics and clinical cytogenetics. Although developing independently for the first four decades, these two fields have contributed enormously to the molecular diagnosis of disease and provide a better understanding of the genetic bases of human 3.2 The genomic basis of medicine Paweł Stankiewicz and James R. Lupski

3.2 The genomic basis of medicine 219 physiology and pathophysiology. Around 25 years ago, techno- logical advances, mainly in fluorescence microscopy, have led to the development of molecular cytogenetics techniques that by enabling the identification of submicroscopic chromosome rearrangements, bridged the genetic variation ‘resolution’ gap between molecular genetics and clinical cytogenetics. As a consequence, since the early 1990s, the genomic aspects of inheritance have come to be recog- nized, as elucidated, for example, through studies of the submicro- scopic genomic duplication at chromosome 17p12 causing the common autosomal dominant adult onset Charcot–Marie–Tooth type 1A distal symmetric polyneuropathy (CMT1A). The beginning of human genetics can be traced to the rediscovery of Gregor Mendel’s observations on the inheritance of phenotypic traits in the garden pea Pisum sativum and Archibald Garrod’s elu-cidation of the genetics of biochemical traits such as alkaptonuria. Mendel found that during gamete formation, each member of the diploid allelic pair separates from the other one to form the gen- etic constitution of the haploid gamete. This phenomenon of inde- pendent segregation is now known as Mendel’s first law. Mendel’s second law states that the segregation of two alleles (corresponding DNA loci on homologous chromosomes) during gamete formation is independent from the segregation of the alleles of other allelic pairs. We now know that linkage, the physical proximity of two gen- etic loci on a linear map, results in exceptions to Mendel’s second law. Such linkage information has been used to map disease traits in humans. Mendel’s ‘inheritance factors’ encoding the genetic in- formation were further defined and termed ‘genes’ many years later. During the last two decades it has become possible to deter- mine the sequence and variation of the entire DNA content of the living organism—the genome. The first human haploid reference genome sequence became available at the turn of this century. In the current postgenomic era,

both high-resolution genome ana- lysis by chromosome microarray analyses (CMA), and personal- ized diploid genomic sequencing using exome sequencing (ES) or whole-genome sequencing (WGS) applied to the study of in- herited and complex traits promise a continued revolution in our understanding of the genetic bases of human biology and the pathophysiology of disease. Genes, chromosomes, and our genome A gene is defined as a set of segments of DNA (deoxyribonucleic acid) that carries the information necessary to produce (transcribe) a functional RNA (ribonucleic acid). Despite the completion of the Human Genome Project (HGP), the exact number of protein-coding genes in the human genome is still unknown; the current estimate is between 20 000 and 25 000. The DNA double helix is a three-dimensional polymer com- posed of units called nucleotides. A combination of two purine bases, adenine (A) and guanine (G), and two pyrimidine bases, thymine (T) and cytosine (C), with deoxyribose sugars and linked by phosphodiester bonds (base + sugar + phosphate = nucleotide) constitute a single strand. The two strands are held together and stabilized by hydrogen bonds that enable Watson–Crick base pairs to form. The base A forms two hydrogen bonds with T while C forms three hydrogen bonds with G. A combination of three nucleotides constitutes a triplet codon that encodes for an individual amino acid by a specific universal genetic code. Different triplets can encode the same amino acids or stop codons during translation; there are 43 = 64 different codon combinations possible, but only 20 amino acids, making the genetic code degenerate. Most genes consist of coding regions termed exons that are sep- arated by intervening introns. The exonic and intronic portions of a gene are transcribed by RNA polymerase II into messenger RNA, or mRNA, that usually begins with a cap on the 5' end and terminates with a polyadenylated (polyA) tail on the 3' end. The in- trons are deleted in a process called splicing and the resulting ma- ture transcript, or spliced mRNA, is translated into a polypeptide chain starting with a methionine encoded by the AUG triplet (in RNA, thymine is replaced by uracil, U) at the 5' end (N-terminal, NH_2, or amino end of polypeptide) and terminated by the stop co- dons: UAA (also known as ochre), UAG (amber), or UGA (opal or umber) at the 3' end (C-terminal, COOH, or carboxyl end of the polypeptide). The human haploid genome consists of approximately 3×109 bp and the normal diploid human genome in each cell is composed of approximately 6×109 bp. Most of the human genome consists of repetitive DNA elements. These can be divided into tandem re- peats represented by satellites (e.g. in centromeres), telomeric re-peats, micro- (di-, tri-, and tetranucleotide repeats), mini-, and macrosatellites, and interspersed repeats derived from transposable elements (e.g. Alu elements and L1 elements), which together com- prise about 60% of the human genome. It has been estimated that greater than 6% of the haploid human genome is present in two or more copies, which have been termed low-copy repeats (LCRs) or segmental duplications. They are defined as DNA fragments larger than 1 kb in size and of more than 90% DNA sequence identity in the haploid reference genome. The unique DNA sequence portion of the human genome in- cludes genes, regulatory elements, and other nongenic sequences. It has been shown that most of this DNA might be transcribed, but protein-coding sequences occupy only about 1.5% of the human genome. For every human, it is important to inherit the proper amount of genomic information, with contributions from both parents and the correct copy-number of each genetic locus for proper function. The genes in a human genome are distributed along 46 chromo- somes. There are 22 pairs of autosomal chromosomes and two sex chromosomes—X and Y in males and two X chromosomes in fe- males. In a conventional clinical cytogenetic analysis using a light microscope, chromosomes can be recognized and distinguished from each other when their chromatin is condensed (arrested in metaphase of the cell cycle) and specifically stained (e.g. with Giemsa), revealing a characteristic G-banding pattern. Each human metaphase chromosome consists of two chromatids forming the chromosome arms connected by a

centromere. Centromeres in the human genome consist of α-satellite DNA (arranged by monomers of approximately 171 bp) and occupy about 2 to 3% of the human genome. Depending on the relative location of the centromere, chromosomes have been divided into three types: metacentric (with similar-sized arms), submetacentric (with one arm significantly longer that the other, the shorter arm referred to as p and the longer as q), and acrocentric (with a centromere located very close to one end of a chromosome— chromosomes 13, 14, 15, 21, and 22).

220 SECTION 3 Cell biology Human chromosome ends are capped by telomeres that con- tain thousands of copies of a telomeric repeat sequence TTAGGG. Telomeres are synthesized by the ribonucleoprotein telomerase. Based on the size and relative centromere position, human chromo- some pairs have been enumerated and arranged in a karyogram that is routinely applied in clinical cytogenetics. A clinical karyotype al- ways designates the number and chromosomal sex: e.g. normal fe- male: 46,XX; normal male: 46,XY; male with trisomy 21 associated with Down syndrome: 47,XY,+21. Pathogenic genetic variants The normal flow of the genetic information is susceptible to perturb- ation at different levels. Changes in the base pairs are called variants or mutations and can arise as a result of replication, recombin- ation, and repair errors, or by exposure to external environmental factors (e.g. radiation or chemical mutagens). To prevent a disease connotation of the term ‘mutation’, ‘variant’ is now used more com- monly. Structurally, small variants can be divided into point muta- tions (substitutions) and insertions or deletions (indels). The most common variants involving exchange of pyrimidine for pyrimidine (e.g. C to T) or purine for purine (e.g. A to G) are called transitions. The rarer transversions substitute purine by pyrimidine (e.g. A to C) or the reciprocal (e.g. G to T). The CpG dinucleotide is particularly prone to transition variants (about tenfold relative to other bases) because methylated C (after CpG island methylation) becomes T if deaminated, and now pairs with A. DNA alterations that do not lead to a change in an amino acid, because of the degenerate code, are called silent (synonymous) vari- ants. These do not change an amino acid but can have functional consequences, e.g. if they create a cryptic splice site or affect an exon splice enhancer. Missense mutations result in an amino acid change and nonsense mutations introduce stop codons that truncate the protein prematurely. Small insertions and deletions called indels that are not a multiple of three nucleotides, which can shift a reading frame and thus alter the protein primary sequence structure, are called frameshift mutations. Abnormally truncated or erroneous transcripts with a premature termination codon (PTC) due to nonsense, frameshift, or splice mu- tations are eliminated from cells by a surveillance mechanism called nonsense mediated decay (NMD). NMD is usually triggered by a PTC in any exon except the last and a portion of the penultimate exon; PTCs in the last 50 to 55 bp of the penultimate exon or in the final exon escape NMD presumably because of the inability of the machinery to distinguish such a PTC from the normal stop codon. About one-third of all human disease-associated point mutations result from PTCs due to nonsense or frameshift alleles. Mutations have been categorized also on the basis of their pheno- typic outcomes. Loss-of-function mutations (hypomorphic if the loss is partial, amorphic if it is complete) manifest phenotypically when a decreased amount of protein is insufficient for the normal cell function (e.g. in haploinsufficient genes). Gain-of-function mu- tations (neomorphic) enhance the normal or take on a new protein function, and dominant negative mutations (antimorphic) result in a protein that acts antagonistically with the normal product from the other allele or another subunit of a protein complex. A genetic locus is said to be homozygous when two alleles have the same status (e.g. both alleles are mutated) and heterozygous when one allele is mutated and the second is normal (wild type). Compound heterozygotes have different mutations in both alleles of one gene. Double

heterozygotes have two mutant alleles, but each is at a different gene locus. The status in which one of the alleles is ab- sent (e.g. for most of the X chromosome genes in males) is described as hemizygous. Typically, different mutations in a gene manifest with the same phenotype, a phenomenon described as allelic heterogeneity. However, different mutations in the same gene can sometimes lead to varied phenotypes. Such a situation is described as allelic affinity. Finally, if the same phenotype is caused by mutations in different genes, this is described as genetic or locus heterogeneity. Patterns of inheritance Mendelian inheritance Genetic traits can show mendelian or nonmendelian inheritance patterns. Mendelian traits involve a single locus, are usually mono- genic, and segregate in autosomal dominant, autosomal recessive, or X-linked fashion. Autosomal dominant inheritance In autosomal dominant inheritance, the mutated allele is trans- mitted to 50% of the gametes and thus is expected to be present in one-half of the progeny. However, if the trait is lethal, incompletely penetrant, age-dependent, or results in variation in expressivity, the proportion with manifestations of disease may vary from 0 to 50%. In pedigree analysis, autosomal dominant inheritance is observed as a vertical transmission of the trait. Autosomal recessive inheritance In autosomal recessive trait the affected individuals, representing one-fourth of the progeny, carry two mutant alleles at a locus as compound heterozygous or homozygous variants, each one usu- ally inherited from carrier parents. In families with healthy siblings, two-thirds are expected to be carriers of the mutant allele and one- third (one-fourth of all progeny) have two wild-type alleles. When the mutated alleles in the affected subject are the same, the family is usually consanguineous. In pedigree analysis, autosomal recessive inheritance is revealed as horizontal transmission of the trait. Of note, for a few autosomal recessive traits it has been shown that the heterozygous carriers of the mutated allele may have an increased susceptibility to complex or multifactorial traits (Table 3.2.1). Moreover, at some loci carrier states convey selective advantage; for example, haemoglobin B gene (HBB) and protection from mal- aria, CFTR and protection from cholera death, hence in some world populations the carrier state can reach a relatively high frequency. The X chromosome In females, the vast majority of genes on the X chromosome undergo random inactivation and thus represent structural disomy but func- tional monosomy. If one of the X chromosomes harbours a mutated recessive allele, X inactivation is usually nonrandomly skewed with the X chromosome harbouring the mutant allele being preferentially inactivated. Therefore, X-linked recessive diseases are not present in females but affect all males since they have only one X chromo- some. However, females with an incomplete X inactivation (e.g. the efficiency of X inactivation decreases significantly with age) or

 skewed X inactivation (e.g. 5–10% of females have a 80:20 ratio of X inactivation), females with Turner syndrome and a 45,X karyotype, or females carrying a balanced translocation between the X chromo- some and an autosome (X material on the derivative chromosomes is not inactivated) can manifest the X-linked recessive disease. For example, some carriers of mutations in the PHF6 gene in families with Börjeson-Forssman-Lehmann (BFL) syndrome manifest mild features or different phenotypes, likely due to escaping X inactiva- tion in some tissues or cell types. However, classical X-inactivation studies cannot explain these findings. In contrast, X-linked dominant diseases are present in both males and females and twice as many females as males are affected. However, the phenotype is usually milder in females than in males. Occasionally, if the disease is lethal in males, the trait can be visible only in females (e.g. Rett syndrome). In the X-linked diseases, no male-to-male transmission is observed and all daughters of affected fathers are obligate carriers of the mutated allele. Penetrance, expressivity, and age of onset The determination of the mendelian segregation pattern can be challenged in

pedigree analysis by incomplete penetrance, wherein a phenotypic feature can be present or absent (e.g. in Marfan syn- drome); variable expressivity, when the same mutation leads to dif- ferent severity or pattern of the phenotype (e.g. in cystic fibrosis); or manifestations depending on age (e.g. in Huntington disease). Nonmendelian inheritance There are many genetic abnormalities, which show familial recur- rence, but do not demonstrate mendelian segregation patterns. Such distortions from mendelian expectations, or nonmendelian traits, can be due to multiple aetiologies, including: genomic imprinting, uniparental disomy, mosaicism, mitochondrial DNA mutations, compound inheritance, digenic or triallelic inheritance, or muta- tional burden. Genomic imprinting If a phenotypic trait is transmitted through only one gender (parent- of-origin effect), genomic imprinting should be considered. During the passage through meiosis, several genes are silenced (imprinted) in a sex-specific manner. For example, the paternal copy of the UBE3A gene on chromosome 15q11.2 is imprinted during spermatogenesis. In the progeny, only the allele inherited from the mother is expressed in a neuronal tissue specific manner and is sufficient to produce enough RNA for normal cell function. When this single active allele is mu- tated or deleted, the individual is affected (in this case with Angelman syndrome). The sex-specific imprint of UBE3A is erased upon the en- trance of chromosome 15 into meiosis and, depending on the sex, a new imprint is established; again, only one allele is expressed. Uniparental disomy The active allele of the imprinted gene will not be transmitted to the progeny if both homologous chromosomes harbouring the im- printed allele are inherited from one parent. Such lack of normal biparental inheritance of homologous chromosomes has been de- fined as uniparental disomy. Two major types of uniparental disomy have been described, heterodisomy and isodisomy. In heterodisomy, two homologous chromosomes from one parent are transmitted to the child and in isodisomy, both homologous chromosomes in an offspring originate from only one of the parental homologues. If a recessive disease gene is present on the isodisomic chromosome, the disease will manifest even though only one parent is a car- rier for the mutation. The most frequent mechanism responsible for uniparental disomy is chromosome nondisjunction in meiosis I followed by trisomy-to-disomy-rescue in an early postzygotic stage of the embryo. In humans, among autosomes only trisomies 13, 18, and 21 are compatible with life. In certain tissues, the cells carrying an extra (trisomic) chromosome can survive only if one of the three copies of the homologous chromosomes (or a large portion thereof) is lost. This process of elimination of the extra chromosome is called trisomy rescue. Since this is a random event, in one-third of cases the remaining chromosomes will be from the same parent, thus rep- resenting uniparental disomy. Because most cases result from an initial nondisjunction event, uniparental disomy is associated with advanced maternal age. Mitochondrial inheritance If a phenotypic trait is inherited only from mothers and never from fathers, then mitochondrial disease should be considered. Table 3.2.1 Recessive disorders and heterozygous predisposition to multifactorial disease Monogenic disease OMIM Gene Risk for multifactorial disease OMIM Aicardi–Goutières syndrome 1 225750 TREX1 Chilblain lupus 610448 Ataxia-telangiectasia 208900 ATM Breast cancer 114480 α1-Antitrypsin deficiency 107400 AAT Chronic obstructive lung disease 606963 Cystic fibrosis 219700 CFTR Pancreatic insufficiency, Chronic rhinosinusitis, Idiopathic bronchiectasis 167800, 211400 Familial hypercholesterolemia 143890 LDLR Coronary artery disease 108725 Gaucher disease 230800, 230900, 231000 GBA Parkinson disease, late-onset 168600 Hyperlipoproteinemia 238600 LPL Ischaemic heart disease 612030 Parkinson disease 600116 PARK2 Lung cancer; ovarian cancer 211980; 167000 Progressive familial intrahepatic cholestasis 602347 ABCB4 Intrahepatic cholestasis of pregnancy 147480 Tay-Sachs disease 272800 HEXA GM2-gangliosidosis, several forms 272800 Stargardt disease 248200 ABCR (ABCA4) Age-related macular degeneration 153800

 Mitochondrial DNA (mtDNA) is present in multiple copies in the cell cytoplasm and is transmitted to progeny only through the oo- cytes; the sperm carries a negligible amount of mtDNA. A broad spectrum of disease phenotypes may be caused by defects in the mitochondrial genome. In some cases, both the mother and her child may present with varying severity of the phenotype due to different proportions of mutated mtDNA in the cytoplasm, a phe- nomenon called heteroplasmy. In mitochondrial disease, the most energy-dependent tissues (e.g. skeletal muscle, brain, heart, eyes) may be the first to reveal clinical signs or symptoms.

Digenic inheritance Several diseases that are not complex traits and are not inherited as simple single-gene mendelian disorders have been shown to be caused by mutations in two different genes, in which the other two alleles are normal. Such double heterozygotes that interact genetic- ally to manifest the phenotype have been described for example in retinitis pigmentosa (ROM1 and RDS encode interacting gene prod- ucts) and deafness (GJB6 and GJB2) (Table 3.2.2).

Triallelic inheritance Bardet–Biedl syndrome, a pleiotropic mendelian disorder character- ized by postnatal obesity, postaxial polydactyly, and progressive ret- inal dystrophy, can be caused in some families by mutations in at least two genes. Mutation analyses have revealed that in some patients with Bardet–Biedl syndrome, three mutant alleles in two different genes segregated with expression of the disease. This phenomenon has been described as triallelic inheritance and has been observed in other diseases also (e.g. familial hypercholesterolemia and cortisone reductase deficiency). Based on this observation, an oligogenic type of inheritance (i.e. mutations in a small number of genes combined rather than a single locus mutation) was proposed to explain some other phenotypes, such as Hirschsprung disease (see Table 3.2.2). Multiple copy-number variants (CNVs) and compound inheritance Sporadic disease can also potentially result from a combination of two CNVs at a single locus (e.g. analogous to the autosomal reces- sive neuromuscular disease spinal muscular atrophy or the renal disease nephronophthisis type I) or theoretically from two or more CNVs at different loci from two normal parents. Systematic evalu- ations of patients with the same-sized CNVs and variable pheno- typic expressivity showed that additional CNVs could modify the severity of the phenotypic manifestations. This phenomenon has been referred to as two-hit (or second-hit) model. More recently, a compound inheritance model, consisting of a rare null mutation and a common, noncoding, haplotype that acted as a hypomorphic allele T-C-A (rs2289292, rs3809624, and rs3809627) of TBX6, with a prevalence of 44% among Han Chinese, was found to account for up to 11% of congenital scoliosis cases in a Chinese population. These findings were elegantly recapitulated in a mouse model of TBX6 null and hypomorphic mutations. Moreover, complex compound inheritance of coding and noncoding regulatory variants, e.g. in tissue-specific enhancer or evolutionarily conserved regions, have been described recently also for congenital anomalies of the kidney and urinary tract (CAKUT), lethal lung developmental diseases, and neurodevelopmental disorders. These studies help unravel the complex molecular bases of incomplete disease penetrance. Mosaicism Another distortion of mendelian inheritance can be caused by mo- saicism. Two or more cell lines can be present either in the gonads only (germline mosaicism) or in somatic cells (somatic mosaicism). Combined somatic and germline (gonosomal) mosaicism has been identified in parents of patients with several genetic conditions, thus raising the possibility that mosaic individuals might be detected by routine blood tests rather than requiring direct examination of germ cells. Mosaicism should be suspected when healthy parents have two or more children with a dominant disease. It has been shown that mutations that exist only in the mosaic state, presumably because constitutional mutations are embryonic lethal (e.g. Proteus syn- drome, OMIM 176920; hemimegalencephaly, OMIM 615937), can have profound effects on the phenotype of an individual. Pleiotropy and epistasis Pleiotropy occurs when a variant

in a single gene has more than one distinguishable phenotypic effect. Epistasis refers to interaction between genes, in which a phenotypic effect is different from what would be expected if mutations of the genes were expressed inde- pendently. In both situations, the inheritance pattern in pedigree analysis can appear as nonmendelian. Clan genomics Analyses of the CMA, ES, WGS, and 1000 Genomes Project data, also targeted genomic sequencing derived from very large sample sizes, reveal that an abundance of rare and private single nucleotide variants (SNVs) and CNVs with large effect have arisen recently in the population history. Such rare and novel variants including new mutations contribute significantly to sporadic clinical traits. Clan Genomics posits that recent mutation may have a greater influence on disease susceptibility or protection and account for many more medically actionable variants, than is conferred by common vari- ations that arose in distant ancestors. Genetic and genomic variation The Human Genome Project has provided a haploid reference human genome enabling an enormous amount of DNA sequence data gener- ation and personal genome analyses for disease-associated CNV and SNV, but it did not assess the scale of genetic and genomic polymorphic variation between different individuals and among populations. HapMap was developed in 2002 to determine the common pat- terns of human DNA sequence variation and to generate a haplotype Table 3.2.2 Models of disease allele transmission Locus Allele Example (disease/gene) Monogenic Monoallelic Angelman (UBE3A) Biallelic Cystic fibrosis (CFTR) Triallelic CMT1A (PMP22) Digenic Biallelic RP (ROM1-RDS) Triallelic BBS (BBS1–19) Oligogenic Hirschsprung disease

 map (i.e. a linked set of genetic markers) of the human genome that in turn would help to identify genes affecting health and responses to drugs and environmental factors. Concurrently, the Human Genome Diversity Project (HGDP) was established to help to understand the diversity and unity of the entire human species, and the ENCODE Project was launched to help interpret this informa- tion, elucidate the functional consequences of noncoding variation, and to better understand the biology of human health and disease. By using high-throughput methods, the ENCODE Project has generated a comprehensive catalogue of the structural and func- tional components encoded in the human genome sequence, including protein-coding genes, nonprotein-coding genes, tran- scriptional regulatory elements, and sequences that mediate chromosome structure and dynamics. HGP, HapMap, HGDP, ENCODE, and several additional studies have revealed that human genetic variation is tremendous and con- sists of two major types: nucleotide sequence variations and gen- omic structural changes. Human genetic and genomic studies have resulted in the proliferation of several databases required to in- terpret the potential clinical significance or relevance of variation (Table 3.2.3). Types of variations Single nucleotide polymorphisms (SNPs) A genetic polymorphism is defined as a heterozygous DNA vari- ation found in more than 1% of the general population. A first phase of the studies on genetic variation in humans has focused on SNPs. A SNP is a nucleotide change that results from a base substitution during DNA replication. The frequency of such events has been es- timated as 10–8 per base pair per generation. The vast majority of SNPs represent inherited changes that have accumulated over thou- sands of human generations. Depending on random genetic drift or natural selection models, the frequency of a particular SNP in the population can significantly change over generations. Typically, there are two major alleles for a SNP locus (i.e. biallelic); however, any base position can potentially be altered to one of three nu- cleotides (i.e. multiallelic), and as more personal genomes are sequenced, evidence indeed supports that all base positions are multiallelic. SNPs are predominantly localized in the noncoding portion of the human genome; only c.15% of over 650 million currently known SNPs map within genes. A SNP that does not change the

polypeptide sequence is termed synonymous (or silent mutation) and if it leads to a polypep- tide sequence change it is described as nonsynonymous. It has been shown that SNPs that are not in protein-coding regions may still have functional consequences (e.g. they can generate splicing mutations, modify transcription factor binding sites or gene regulatory elements, or change the sequence of noncoding RNAs). A combination of closely linked SNPs is defined as a haplotype. Haplotypes result from reduced recombination (crossing-over) events between closely linked genetic markers during meiosis and are generally shared between different populations; however, their frequency can differ widely. The nonrandom association of SNPs is described as linkage disequilibrium. The identification of a large number of SNPs has enabled successful genome-wide association studies for risk alleles for susceptibility to common complex disease traits like diabetes and cancer (Table 3.2.4). Often such suscepti- bility variants map to noncoding regulatory regions.
Repetitive DNA elements Variable number of tandem repeats (VNTR);
short tandem repeats (STRs) Di-, tri-, and tetra-nucleotide repeats such as (GT)n, (CAA)n, or (GATA)n, respectively, have been referred to as microsatellites. They are very unstable and polymorphic genomic loci, thus useful Table 3.2.3 Summary of useful genomic websites Name URL Description 1000 Genomes http://www.1000genomes.org/ The 1000 Genomes Project is the first project to sequence the genomes of a large number of people, to provide a comprehensive resource on human genetic variation. ClinVar http://www.ncbi.nlm.nih.gov/clinvar/ ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence. DECIPHER, DatabasE of Chromosomal Imbalance and Phenotype in Humans http://www.sanger.ac.uk/PostGenomics/ decipher Database of submicroscopic chromosomal imbalance describes clinical phenotype associated with submicroscopic rearrangements Database of Genomic Variants http://projects.tcag.ca/variation/ A curated catalogue of structural variation in the human genome dbSNP http://www.ncbi.nlm.nih.gov/projects/SNP/ A public-domain archive for a broad collection of simple genetic polymorphisms Ensembl http://www.ensembl.org/ Wellcome Trust funded software system which produces and maintains automatic annotation on selected eukaryotic genomes GeneTests/GeneReview http://www.genetests.org/ A publicly funded medical genetics information resource developed for physicians, other healthcare providers, and researchers Human Gene Mutation Database http://www.hgmd.cf.ac.uk/ac/index.php? HGMD constitutes a comprehensive core collection of data on germline mutations in nuclear genes underlying or associated with human inherited disease OMIM, Online Mendelian Inheritance in Man http://www.ncbi.nlm.nih.gov/sites/ entrez?db = OMIM A catalogue of human genes and genetic disorders authored and edited by Dr Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere UCSC Genome Bioinformatics http://genome.ucsc.edu/ This site contains the reference sequence and working draft assemblies for a large collection of genomes

 in DNA fingerprinting and personal identification, population gen- etics, pedigree analysis, recombination, and linkage studies, as well as in determining paternity or parental origin of chromosomes. Minisatellites are DNA segments that consist of a short series (10–100 bp) of GC-rich tandem repeats and are present at more than 1000 locations in the human genome. Retrotransposons Alu elements are approximately 300 bp in size, present in about 1 mil- lion copies, and occupy approximately 10% of the human genome. The pathogenic function of Alu elements has been demonstrated to be exerted by two major mechanisms: insertional mutagenesis, util- izing an RNA intermediate to move or transpose Alu into exons or near spice junctions, and postinsertional ‘activity’ whereby they act as substrates to mediate rearrangements.

Alu elements that share high sequence identity can serve as po- tential homologous recombination substrates. However, most Alu elements share less than 97–98% sequence identity but can act as potential substrates for microhomology-mediated DNA replica- tion errors via template switching, as in Fork Stalling and Template Switching (FoSTeS), and microhomology-mediated break induced replication (MMBIR). Often times, recombination occurs be- tween imperfectly matched substrates, a phenomenon described as homeology. Examples of diseases caused by Alu element insertions include haemophilia A and B, and breast cancer due to disruption of the BRCA1 gene and by Alu–Alu rearrangements in the LDLR, FOXF1, and SPAST genes, and C1 inhibitor locus. L1 elements are approximately 6 kb long, present in about 500 000 copies, and account for some 17% of the human genome. They are an important source of genomic variation. Like Alu repetitive elem- ents, using an RNA intermediate, L1 elements can be mutagenic by insertions into genes. Due to their abundance in the human genome, L1 elements that have high sequence identity can also stimulate and mediate nonallelic homologous recombination (NAHR). L1 elem- ents have been shown to mutate the genes responsible, for example, for Alport syndrome, colon cancer, haemophilia, β-thalassemia, neurofibromatosis type 1, and Duchenne muscular dystrophy. Dynamic mutations Unlike the aforementioned DNA changes, which are usually trans- mitted through many generations without any change, more than 20 diseases have been described to date that are caused by unstable dynamic mutations occurring during DNA replication, repair, or re-combination. Most of these mutations are represented by an expan- sion of a simple triplet or trinucleotide repeat sequence (e.g. CAG, CGG, CTG, and AAG) in either coding regions (e.g. in Huntington disease) or noncoding regions such as introns (e.g. in Friedreich ataxia), or either 5' untranslated regions (e.g. in fragile X syndrome) or 3' untranslated regions (e.g. in myotonic dystrophy). These triplet repeat diseases have been shown to be inherited as autosomal dom-inant (e.g. in myotonic dystrophy), autosomal recessive (e.g. in Friedreich ataxia), or X-linked (e.g. in fragile X syndrome) traits due to gain- or loss-of-function mutations. The minimal number of disease-causing triplet repeats varies among different disorders, with 36 in Huntington disease and about 200 in fragile X syndrome. The intermediate number of repeats, lower than in affected individuals but greater than normal, is called premutation. It has been shown that premutations also can have phenotypic effects. For example, an increased incidence of ovarian failure in females and a late-onset neurological disorder in males have been reported in indi- viduals carrying premutations in the fragile X syndrome FMR1 gene. Premutations have a potential to expand during meiosis and thus manifest the disease in the next generation. The nucleotide expan- sion often occurs in a sex-specific manner and can be observed in a pedigree as a parent-of-origin effect. For example, expansions in fragile X syndrome arise during oogenesis and not in spermatogen- esis. The number of the pathogenic repeats can correlate inversely with the onset and severity of the disease; this provides a molecular explanation for the clinical phenomenon referred to as anticipation. Anticipation is observed in pedigree analysis as reduced age of onset in successive generations. In Huntington disease, the expandable triplet repeat CAG en- codes for the amino acid glutamine. Individuals with Huntington disease have 36 or more CAG repeats, which leads to polyglutamine expansion with subsequent huntingtin protein misfolding, ag- gregation, and degradation that exert toxic effects upon neurons. Similar polyglutamine expansions have been reported in several other neurological diseases (e.g. in spinocerebellar ataxia type 1). Expansions of polyalanine tracts beyond a certain threshold have been described as pathogenic, for example in congenital malfor- mations, skeletal dysplasia, and nervous system anomalies. Other pathogenic expansions have been shown to involve tetranucleotides Table 3.2.4 Genome-wide association studies. SNPs and complex traits Disease Locus Reference Breast cancer FGFR2 Easton et al.

(2007), Nature, 447, 1087–93 Hunter et al. (2007), Nat Genet, 39, 870–4 Coronary heart disease SNP, rs1333049, 9p21.3 Samani et al. (2007), N Engl J Med, 357, 443–53 Crohn disease IRGM Parkes et al. (2007), Nat Genet, 39, 830–2 Diabetes 12q24 (HNF1A), 12q13, 16p13 (CLEC16A), and 18p11 Todd et al. (2007), Nat Genet, 39, 857–64 Macular degeneration CFH Li et al. (2006), Nat Genet, 38, 1049–54 Maller et al. (2006), Nat Genet, 38, 1055–9 Obesity FTO Dina et al. (2007), Nat Genet, 39, 724–6 Frayling et al. (2007), Science, 316, 889–94 Prostate cancer 8q24 (MYC, rs1447295, rs16901979, and rs6983267) Gudmundsson et al. (2007), Nat Genet, 39, 631–7 Yeager et al. (2007), Nat Genet, 39, 645–9 Rheumatoid arthritis SNP, rs10499194, 6q23 Plenge et al. (2007), Nat Genet, 39, 1477–82

 (e.g. CCTG in myotonic dystrophy type 2), pentanucleotides (e.g. ATTCT in spinocerebellar ataxia type 10), and even dodecamers (e.g. CCCCGCCCCGCG in progressive myoclonus epilepsy of the Unverricht–Lundborg type). Secondary DNA structures Abnormal secondary DNA structures can also be mutagenic. Several DNA conformations, different from the canonical right-handed B- form, have been described. The best-known non-B DNA structures include triplexes, left-handed DNA, bent DNA, cruciforms, nodule DNA, flexible and writhed DNA, G4 tetrad (quadruplexes), slipped structures, and sticky DNA. Some of these structures have been de- scribed as pathogenic for more than 20 neurological and psychiatric diseases. One of the best-known examples of the pathogenic role of non-B DNA structures are AT-rich cruciforms in the proximal chromo- some 22q11.2, responsible for genomic instability and susceptibility to the most common recurrent non-Robertsonian translocation t(11;22)(q11.2;q23.3) in humans. Copy-number variants In contrast to c. 500 000 insertion or deletion polymorphisms less than 1 kb in size that have been well-studied and annotated, little was known about the polymorphic changes larger than this. The appli- cation of CMA to analyse the genomes of normal humans has led to the discovery of extensive genomic structural variation, ranging in size from thousands to millions of bases, which are not recognizable by chromosomal banding. These changes have been termed CNVs and result in deviation from the normal diploid state at a given locus. Deletions, duplications, triplications, quadruplications, insertions, or translocations can all result in CNVs. The total number, position, size, gene content, and population distribution of CNVs remain elu- sive. Data are still evolving but even several years ago, estimates have suggested approximate figures of 6000 CNVs in 4000 regions overlap- ping 1500 genes; most of these represent common variant CNV and thus are not associated with disease. However, they may contribute to pathology as recessive alleles. CNVs may account for as much as 360 to 500 Mb and represent 12 to 20% of the human genome. These num- bers can still represent a conservative estimate because CNVs ranging in size from 50 bp to 200 bp, those involving Alu and L1 variation at a single locus, and single exon drop out alleles resulting from error prone DNA replicative mechanisms, have not been well ascertained on a genome-wide scale in different populations. It is anticipated that with the wider application of higher-resolution CMA techniques, and next-generation sequencing to determine in- dividual diploid genomes, the amount of structural variation iden- tified will increase significantly. The genomic distribution of CNVs has been shown to be nonrandom and correlates with exons, seg- mental duplications, and the mobile elements such as Alu repetitive elements, probably reflecting their ongoing evolutionary role. Like many other genomic rearrangements, CNVs can be inherited or sporadic. A commonly used and useful standard is to assume that de novo CNVs in association with sporadic clinical phenotypes are more likely to be disease causative. However, the phenotypic effects of CNVs are sometimes unclear and depend mainly on whether dosage-sensitive genes are affected by the genomic rearrangement.

Some CNVs have been shown to be responsible for mendelian dis- eases, nonmendelian traits such as complex diseases, and common traits (including behavioural traits), or to represent benign poly- morphic variation (Fig. 3.2.1; Table 3.2.5). CNVs have been proposed also to be a major factor responsible for human diversity and evolution. CNVs have been catalogued in public databases such as the Toronto Database of Genomic Variants and 1000 Genomes phase III studies. Clinically relevant CNVs can be found in: DECIPHER (see Table 3.2.3). Chromothripsis Next-generation DNA sequencing of human tumours has led to discovery of chromothripsis, a phenomenon of complex rearrange- ments in one or a few chromosomal loci that arose in a single cata- strophic event. One proposed mechanism for chromothripsis is chromosome shattering with random reassembly in the subsequent interphase by nonhomologous end joining (NHEJ). A similar phe- nomenon has been observed in constitutional rearrangements as- sociated with developmental disorders. Errors of replicative repair in which DNA replication initiates serial, microhomology-mediated template switching is proposed to produce such rearrangements in a process termed chromoanasynthesis. Genomic disorders Over the past two decades it has become evident that higher-order genomic architectural features can confer susceptibility to DNA rearrangements that are a frequent cause of diseases in humans. Conditions that result from such rearrangements of the human genome have been referred to as genomic disorders. Many genomic disorders occur sporadically, and these frequent events are often caused by de novo rearrangements. Various cal- culations have shown that the de novo locus-specific mutation rates for genomic rearrangements are between 10–4 and 10−5; this is at least 100- to 10 000-fold more frequent than point mutations (Table 3.2.6). Genomic rearrangements can cause mendelian diseases (e.g. CMT1A, MODY5) or complex traits such as behaviours and intel- lectual disability, or may represent benign polymorphic changes. The major mechanism by which rearrangements convey pheno- types is altered gene dosage due to a variation in gene copy-number. When the deleted or duplicated region harbours a dosage-sensitive gene, the rearrangement will lead to an abnormal phenotype. Other mechanisms include gene interruptions, gene fusions, position ef- fects, and unmasking of variants in coding region or other func- tional SNVs in the second allele (Fig. 3.2.1). For a few genomic disorders, significant differences in incidences have been observed in different world populations. In some of them, structural variations of the genomic region in the patients' parents have been found, demonstrating that the variation of genomic archi- tecture is a significant factor for disease susceptibility. For instance, submicroscopic genomic inversions can result in haplotype blocks (due to reduced recombination) and generate an architecture with directly oriented LCRs that can act as NAHR substrates. This can lead to the susceptibility to deletion/duplication rearrangements only in the individuals within the population who harbour the in- version variant with the rearrangement-prone genome architec- ture (e.g. in Williams–Beuren syndrome or 17q21.31 microdeletion syndrome).

 Genomic alteration Nonallelic homologous recombination Many LCRs have a complex structure and have arisen during pri- mate speciation during the last 25 to 40 million years as a result of serial segmental duplications. LCRs longer than 10 kb and of more than about 97% sequence identity can lead to local genomic instability. LCRs have been shown to stimulate and/or mediate constitutional (both recurrent and nonrecurrent), evolutionary, and somatic genomic rearrangements. When located at a distance less than 5 to 10 Mb from each other, LCRs can mediate NAHR, and potentially result in unequal crossing-over. NAHR between directly oriented LCRs leads to deletions or reciprocal duplications of the genomic re- gion located between them, and NAHR between the inverted LCRs results in an inversion of the intervening

genomic segment. In LCRs A) gene dosage D) position effect E) unmasking recessive allele or functional polymorphism * * or B) gene interruption C) gene fusion Neuropathy ID Infertility Ptosis Bleeding Anemia Color blindness Blood hypertension ID & deafness Overgrowth & bleeding Pigmentation Trait NAHR substrate Disease CMT1A-REP SMS-REP AZFc REP int22h-1 in Factor VIII and int22h-2 or int22h-3 α-globin RCP and GCP CYP11B1 and CYP11B2 SMS-REP & (mutation in MYO15A) Sos-REP & (mutation in Factor XII) PWS-REPs CMT1A/HNPP SMS/PTLS Azoospermia FOXL2 Haemophilia A α-thalassemia Deuteranopia, protanopia Glucocorticoid-remediable aldosteronism SMS & DFNB3
SoS & Factor XII deficiency PWS Dosage sensitive gene PMP22 RAI1 ? F8 RAI1 NSD1 and F12 P locus α-globin RCP and GCP CYP11B1 and CYP11B2 Blepharophimosis Fig. 3.2.1 Schematic models for molecular mechanisms of genomic disorders. For each model, examples of trait, nonallelic homologous recombination (NAHR) substrate, and disease are shown. (a) Gene dosage, where there is a dosage-sensitive gene within the rearrangement; (b) gene interruption, wherein the rearrangement breakpoint disrupts a gene; (c) gene fusion, whereby a fusion gene is created at the breakpoint that either fuses coding sequences or a novel regulatory sequence to the gene. For example, two genes encoding cytochrome P450 enzymes CYP11B2 (aldosterone synthase) and ACTH-regulated CYP11B1 (11-β-hydroxylase, cortisol biosynthesis) located on chromosome 8q21 are 45 kb apart and have 10 kb segments of 95% sequence identity. NAHR between these two genes results in a chromosome deletion, yielding a fusion hybrid CYP11B1/CYP11B2 gene. CYP11B1/CYP11B2 is under the regulation of ACTH and leads to glucocorticoid-remediable aldosteronism (GRA, MIM 103900). All symptoms of the disease can be normalized by the administration of glucocorticoid analogues and are exacerbated by administration of ACTH; (d) position effect, in which the rearrangement has effects on expression/regulation of a gene near the breakpoint, potentially by removing or altering a regulatory sequence; and (e) unmasking recessive allele, where a deletion results in hemizygous expression of a recessive mutation or further uncovers/exacerbates effects of a functional polymorphism. In each model, both chromosome homologues are depicted as horizontal lines. The rearranged genomic interval is enclosed by brackets. Dashed lines indicate genomic regions either deleted or duplicated, an absent line indicates deletion with phenotypic effects from the remaining allele unmasked because of the rearrangement, and a dotted line represents deletion but where phenotypic effects result from the absence of interactions between alleles. Gene is depicted by filled horizontal rectangle, while regulatory region is shown as a hatch-marked square. Asterisks denote point mutations. CMT1A, Charcot–Marie–Tooth disease type 1A; DFNB3, deafness, neurosensory, autosomal recessive 3; HNPP, hereditary neuropathy with liability to pressure palsies; ID, intellectual disability; PWS, Prader–Willi syndrome; SMS, Smith–Magenis syndrome; PMD, Pelizaeus–Merzbacher syndrome; PTLS, Potocki–Lupski syndrome; SoS, Sotos syndrome. Adapted from Lupski JR, Stankiewicz P (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet, 1, e49. Table 3.2.5 CNVs and complex traits Disease Gene Reference Alzheimer disease APP Rovelet-Lecrux et al. (2006), Nat Genet, 38, 24–6 Chronic pancreatitis PRSS1 Le Maréchal et al. (2006), Nat Genet, 38, 1372–4 Crohn disease IRGM McCarrol et al. (2008), Nat genet, 40, 1107–12 Lupus with glomerulonephritis FCGR3B Aitman et al. (2006), Nature, 439, 851–5 Parkinson disease SNCA Singleton et al. (2003), Science, 302, 841 Systemic lupus erythematosus Complement C4 component Yang et al. (2007), Am J Hum Genet, 80, 1037–54

 of a more complex genomic structure consisting of both direct and inverted subunits, distinct portions can serve as NAHR substrates leading to

deletions/duplications or inversions, respectively. Recombination hot spots Interestingly, the strand exchanges for NAHR sites are not scattered throughout the length of homology within LCRs, but cluster in re- combination hot spots. Normal allelic homologous recombination, like NAHR, is characterized by hot spots and cold spots throughout the genome. A meiosis-specific histone methyltransferase PR do- main zinc finger protein 9 (PRDM9) has been shown to recognize a 13-mer motif CCNCCNTNNCCNC enriched at human hotspots and cause histone H3 lysine 4 trimethylation. This reorganization is thought to be associated with increased probability of recombination. Interestingly, sequence variation in PRDM9 recognition sequence may be responsible for hotspot differences between species. Moreover, variation of PRDM9 in human populations has been directly linked to recombination rates at NAHR generated duplications and deletions. Microdeletion and microduplication syndromes Two common autosomal dominant peripheral neuropathies, CMT1A and hereditary neuropathy with liability to pressure palsies (HNPP), are among the first and best-characterized genomic dis- orders. CMT1A and HNPP are caused in the vast majority of cases by copy-number change of a dosage-sensitive myelin gene PMP22 as a result of reciprocal duplication and deletion, respectively, of an ap- proximately 1.4 Mb genomic fragment within 17p12. This genomic segment is flanked by two LCRs of about 24 kb, approximately 98.7% identical, termed the proximal CMT1A-REP and the distal CMT1A- REP, which serve as substrates for NAHR. The proximal chromosome 17p also harbours another meiotically unstable genomic region with a haploinsufficient RAI1 gene. Deletions and point mutations of RAI1 result in Smith–Magenis syndrome (SMS), a disorder with multiple congenital anomalies and intellec- tual disability characterized by minor craniofacial and skeletal anom- alies such as brachycephaly, frontal bossing, synophrys, midfacial hypoplasia, short stature, and brachydactyly, neurobehavioral ab- normalities such as aggressive and self-injurious behaviour and sleep disturbances, and ophthalmic, otolaryngological, cardiac, and renal anomalies. A genomic segment of approximately 4 Mb encompassing RAI1 and flanked by large, complex, highly identical, and directly oriented, proximal (approximately 256 kb) and distal (approximately 176 kb) LCRs termed SMS-REPs is deleted in 70 to 80% of patients with Smith–Magenis syndrome via the NAHR mech- anism (i.e. common recurrent deletion). The remainder of CNVs in these patients are mediated by NAHR using alternate flanking LCRs as substrates (uncommon, recurrent) or due to nonrecurrent re- arrangements mediated by template switching DNA replication error mechanisms. The latter CNVs have been recently shown to be often accompanied by megabase long hypermutation clusters of SNVs. The reciprocal duplication dup(17)(p11.2p11.2) of this region has been described in patients with Potocki–Lupski syndrome (PTLS). Clinical features observed in patients with this syndrome are dis- tinct from those seen with Smith–Magenis syndrome and include infantile hypotonia, failure to thrive, intellectual disability, autistic features, sleep apnoea, and structural cardiovascular anomalies. Other well-characterized microdeletion syndromes include Williams–Beuren syndrome (7q11.23), Prader–Willi and Angelman syndromes (15q11.2q12), DiGeorge/velocardiofacial syndrome (22q11.2), microdeletion 17q21.31 syndrome, and Sotos syndrome (5q35). For all these microdeletions, the reciprocal microduplications predicted by the NAHR model have been reported. Typically, the phenotypic manifestation of microduplication syndromes is milder than their reciprocal microdeletion counterpart. In chromosome duplications, the increase of 2 to 3 in gene copy-number results in a 1.5-fold increase (50% change) of the protein amount, vs. the 2 to 1 decrease in gene copy-number leading to twofold reduction (100% change) of the protein amount in the reciprocal deletions. Mirror traits For some genomic regions in humans, deletion and reciprocal du- plication CNVs have been found in patients with opposing pheno- types. 1q21.1 deletion was associated with microcephaly and schizophrenia, whereas

1q21.1 duplication was found in patients with macrocephaly and a trend towards autism. Conversely, dele- tion in 16p11.2 was identified in association with macrocephaly and autism, while duplication of 16p11.2 was associated with micro- cephaly and schizophrenia. Moreover, patients with SMS were ob- served to be overweight with high body mass index (BMI), whereas those with PTLS due to dup17p11.2 are usually underweight, which Table 3.2.6 New mutation rates for genomic rearrangements Rearrangement hot spot Mutation rate direct measure Method Mutation rate indirect estimate Method Deletion Duplication Deletion Duplication CMT1A-REP 4.2 × 10−5 1.73 × 10−5 Real-time PCR on sperm DNA 1.7–2.6 × 10−5 Prevalence + molecular AZFa-HERV 2.16 × 10−5 5.26 × 10−6 Real-time PCR on sperm DNA LCR17p 1.87 × 10−6 8.74 × 10−7 Real-time PCR on sperm DNA WBS-LCR 9.55 × 10−6 4.54 × 10−6 Real-time PCR on sperm DNA 2.0–12.5 × 10−5 Prevalence DGS/VCFS; SMS 2.0–12.5 × 10−5 Prevalence DMD 1.0 × 10–4 1.0 × 10−4 Prevalence + molecular α-Globin 4.2 × 10–5 Sperm PCR t(11;22) 1.2–9.5 × 10–5 (translocation) Sperm PCR Normal controls 1.7 × 10–6 1.7 × 10–6 Array CGH of trios Adapted from Turner DJ, et al. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat Genet, 40, 90–5 and Lupski JR (2007). Genomic rearrangements and sporadic disease. Nat Genet, 39, S43–7.

228 SECTION 3 Cell biology was recapitulated in the mouse models of SMS and PTLS. These mirror trait findings are consistent with a theory regarding evolution of the social brain that posited that autism and schizophrenia repre- sented opposing phenotypic extremes of normal human behaviour. Nonhomologous end joining and Fork Stalling and Template Switching (FoSTeS) The NAHR/low-copy repeat mechanism is most prevalent and responsible for the common recurrent deletions, duplications, or inversions. Nonrecurrent rearrangements have been shown in selected cases to arise by the NHEJ mechanism, where the low- copy repeats, if present, stimulate but do not mediate the recom- bination events. DNA replication errors have been shown to play an important role in the origin of some genomic disorders due to nonrecurrent rearrangements. These models all incorporate the concept of tem- plate switching during DNA replication—short distance, within replication fork, and long-distance template switching embodied in the FoSTeS, Fork Stalling Template Switching, model. The MMBIR model was initially proposed based on data from E. coli, yeast, and human genome rearrangements not readily ex- plained by the current models for generating rearrangements. This model appears to account for complex genomic rearrangements due to iterative template switches. Chromosome aberrations Variations of the human genome larger than about 5 Mb in size can be visible in the light microscope and are referred to as chromo- some aberrations. Chromosome aberrations are frequent events, with the total incidence estimated as 1 in approximately 160 live births. They can be categorized as numerical or structural abnor- malities. Numerical abnormalities (1 in approximately 250 new- borns) are observed more frequently than structural ones (1 in approximately 375 newborns). Numerical aberrations Deviations from the normal chromosome number are usually un- balanced and defined as aneuploidy. Triploidies and tetraploidies Triploid (3n) complements of chromosomes, 69,XXX, 69,XXY, or 69,XYY, typically result from an egg being fertilized by two sperms. Tetraploid (4n) sets, 92,XXYY or 92,XXXX, are caused by a failure in zygote division. Both triploidies and tetraploidies states are lethal. Trisomies, monosomies The more commonly observed aneuploidies that result from numer- ical changes of a single chromosome, trisomies and monosomies, are caused by chromosome nondisjunction in meiosis I, or some- times in meiosis II, and in some cases (particularly those involving the sex chromosomes) are compatible with life. Although the most frequent aneuploidy in

humans is trisomy 21 in patients with Down syndrome (1 in 670 newborns), aneuploidies of sex chromosomes are more frequent (1 in 440 newborns) than those involving auto- somes (1 in 700 newborns). This is due to the fact that, in addition to trisomy 21, only trisomies of chromosome 18 (Edwards syndrome, 1 in 7500 newborns) and chromosome 13 (Patau syndrome, 1 in 22 700 newborns) are compatible with life. Although approximately 99% of fetuses with monosomy X are spontaneously aborted, patients with Turner syndrome account for 1 in 4000 female newborns. Very often, however, the 45,X cell line is mosaic, accompanied by another cell line with either a normal cell chromosome complement or structural rearrangements of chromosome X (e.g. deletion of the short arm, ring chromosome, or isochromosome of the long or short arms). The most frequent aneuploidy during fetal life, trisomy 16 (one-third of all trisomies), leads to early miscarriages and is not identified in live newborns. The karyotype 47,XXY is found in patients with Klinefelter syndrome (1 in 1000 male newborns). Marker chromosomes Marker chromosomes (SMCs) are small supernumerary chromo- somes and are detected with a frequency of 0.24/1000 in newborns, 0.4 to 1.5/1000 in prenatal studies, 2 to 3/1000 among phenotyp- ically abnormal individuals, and 0.5/1000 in the general popula- tion. Marker chromosomes are usually derived from acrocentric autosomes (c.85%), and particularly from chromosome 15 (some 40–50%). The risk of an abnormal phenotype in de novo cases has been estimated to be about 28%. The severity of the phenotype de- pends on the size of the marker chromosome and the extent of mosaicism. Structural chromosome aberrations Deletions and duplications Chromosome deletions involving autosomes lead to structural and functional monosomies of the missing genomic material. In XY males, deletions of sex chromosomes result in structural and func- tional nullisomies. The phenotypic manifestation of a deletion is caused by the haploinsufficient gene(s) located in the deleted frag- ment or disrupted by the deletion breakpoint (Fig. 3.2.1). If more than one haploinsufficient gene is present in the deleted region, the abnormality is referred to as a contiguous gene deletion syndrome, as in the Potocki–Shaffer syndrome (11p11.2) or Langer–Giedion syndrome (8q23q24). Many of the smaller deletions in the unstable genomic regions have been shown to have the same size and are re- current. These microdeletion genomic disorders are usually caused by NAHR between directly oriented low-copy repeats and are fre- quent events (Table 3.2.6). Reciprocal translocations Reciprocal translocation is defined as an exchange of the chromo- some segments between two chromosomes (homologous or non-homologous). Balanced reciprocal translocations are found in approximately 1 in 600 individuals; hence, approximately 1 in 300 couples are at risk of unbalanced progeny. In most cases, balanced reciprocal translocations are not associated with an abnormal phenotype; however, it has been shown that up to 40% of the ap- parently balanced reciprocal chromosome translocations in pa- tients with abnormal phenotype are accompanied by a chromosome imbalance either at the translocation breakpoint or elsewhere in the genome. Balanced translocations can also have clinical conse- quences for normal individuals. Depending on the type of meiotic segregation and the size of the translocated chromosome material, the unbalanced meiotic products of the segregating translocation chromosomes can result in chromosome imbalance and be associ- ated with either spontaneous abortions or births of affected children.

 The products of reciprocal chromosome translocation can be trans- mitted to progeny in a balanced or unbalanced form as a conse- quence of alternate or adjacent segregation. In the vast majority of cases, reciprocal translocations appear to be random events. However, two of the most common constitutional non-Robertsonian translocations in humans have been shown to result from a specific genomic architectural features predisposing to -

recurrent events; the breakpoints of translocation t(11;22)(q11.2;q23.3) utilize AT-rich cruciforms whereas low-copy repeats on 4p, 8p, 11p, and 12p mediate the translocations t(4;8)(p16;p23) (olfactory receptor-gene clusters), t(4;11)(p16.2;p15.4), and t(8;12)(p23.1;p13.31). Genomic architecture involving low-copy repeats has also been shown to play a role in the formation of the most frequent somatic chromosome abnormality found in chronic myeloid leukaemia; translocation der(22)t(9;22)(q34;q11)—Philadelphia chromosome. Robertsonian translocations Translocation between two acrocentric chromosomes (13, 14, 15, 21, or 22), with breakpoints occurring in the short arms within or close to the centromere, is defined as Robertsonian translocation or centric fusion. Inverted repeats in acrocentric short arms have been proposed to mediate Robertsonian translocation. One in approximately 900 newborns carries a Robertsonian translocation, making it the most common chromosome rearrange- ment in humans. In some cases, the rearrangement involving long arms of one chromosome is not a product of the centric fu- sion between two homologous chromosomes but a consequence of replication of one chromosome arm, and thus represents an isochromosome. The karyotype of the carrier of Robertsonian translocation is balanced and consists of 45 chromosomes (the acentric heterochro- matic short arms contain no genes and are lost during cell division). All combinations of acrocentric chromosomes have been found; however, translocations between chromosomes 13 and 14 or 14 and 21 are most prevalent, with the Robertsonian translocation 13;14 being the most common chromosome aberration in humans (1 in 1300). Carriers of Robertsonian translocation have a significantly increased risk of abnormal progeny; for example, carriers of trans- location 21q21q have an almost 100% chance of having a child with Down syndrome. The carriers of Robertsonian translocation are also at increased risk of having offspring with uniparental disomy for the acrocentrics involved in the rearrangement due to the trisomy rescue mechanism (see earlier). Uniparental disomy has clinical consequences for car- riers of Robertsonian translocations involving acrocentric chromo- somes 14 and 15 that are known to contain imprinted genes. Insertions A nonreciprocal translocation of DNA material from one chromo- some arm into another arm is described as an insertion or insertional translocation. The carrier of a balanced insertion has up to a 50% chance of an abnormal progeny. Inversions An inversion is defined as a double-break chromosome rearrange- ment, in which a segment of a chromosome is reversed and rein- serted back into the chromosome. Some inversions (particularly those on chromosome 8p) have been shown to be mediated by a specific genomic architecture involving low-copy repeats in an in- verted orientation. When the inverted fragment contains the centromere, the inver- sion is described as pericentric. The recombination products of the pericentric inversion are a chromosome with a terminal deletion of one chromosome arm and a terminal duplication of the second arm. Paracentric inversions do not include the centromere; both breaks occur in one arm of the chromosome. The product of the paracen- tric inversion is either an acentric or dicentric chromosome; in both cases it is unstable and usually a lethal event. Typically, inversions are balanced; however, occasionally imbalances are found at their break- points. In addition, an inversion breakpoint can disrupt a dosage- sensitive gene (e.g. the most common cause of severe haemophilia A, representing over 40% of cases), resulting in an abnormal phenotype, or convey a phenotype because of a position effect. Complex chromosome rearrangements When more than two breakpoints involve two or more chromosomes the resulting aberration is referred to as complex chromosome re- arrangement. These usually arise in spermatogenesis but are more often transmitted to subsequent generations through oogenesis. Ring chromosomes Ring chromosomes are usually formed when two chromosome arms break and fuse, thus forming a circular structure. Rings are often associated with abnormal phenotypes because of loss of genomic material at one or both chromosome ends. In rare cases,

the breaks occur on one chromosome arm and the resulting ring chromo- somes do not contain alphoid centromeres. Such acentric rings can generate neocentromeres from a euchromatic material and can be transmitted to the daughter cells. Rings are mitotically unstable, are often found in a mosaic state, and can form double ring structures as a result of crossing-over events. Isochromosomes When one chromosome arm is lost and the other is duplicated, the resulting mirror-image chromosome is called an isochromosome. When the breakpoint is within the centromere (centromere misdivision), the resulting isochromosome is monocentric and stable. If the original chromosome breaks outside the centromere, the derivative chromosome product is dicentric and thus unstable. To stabilize such a chromosome, one of the centromeres becomes inactive. Such chromosomes are then called pseudodicentric (pseudoisodicentric). The clinically relevant isochromosomes are, for example, isochromosomes of the long arms of chromosome X found in patients with Turner syndrome. Moreover, an isodicentric chromosome idic(17)(p11.2) occurring as a somatic event is fre- quently found in chronic myeloid leukaemia and in childhood primitive neuroectodermal tumours. The idic(17)(p11.2) is recur- rently formed utilizing large cruciform structures containing some 38 to 49 kb low-copy repeats of approximately 99.8% identity lo- calized in the Smith–Magenis syndrome common deletion region in chromosome 17p11.2. The specific mechanism is a NAHR using inverted LCR located on non-sister chromatids Centromere fission Very rarely, as a result of centromere misdivision, the short arms of a chromosome are separated from its long arms and after replication

 form two isochromosomes, representing a balanced rearrangement. Such events are known as centromeric fission. Heterochromatin variants In addition to aberrations involving euchromatin, nonpathogenic variations of heterochromatin are often seen in karyotype ana- lysis. The most common polymorphisms involve differences in size of satellite DNA of the short arm of acrocentric chromosomes and size or location of heterochromatin in 1qhet, 9qhet, 16qhet, and Yqhet. Chromosome mosaicism The presence of two or more different chromosome complements in one individual is defined as chromosomal mosaicism. Somatic chromosomal mosaicism is a well-known cause for birth defects, intellectual disability, and, in some instances, specific genetic syn- dromes such as hypomelanosis of Ito and Pallister–Killian syndrome (tetrasomy 12p). Chromosomal mosaicism is found in up to 50% of embryos at the eight-cell stage and up to 75% in blastocysts. The most common cause of chromosomal mosaicism is chromosome nondisjunction followed by trisomy rescue in a subpopulation of cells. Routine clinical G-banded karyotype analysis is performed in peripheral blood T lymphocytes stimulated to divide by phytohaem- agglutinin. Thus, only a subpopulation of nucleated cells, and only those healthy enough to respond to stimulation, are expanded and examined. Applications of array comparative genomic hybridiza- tion (array CGH) technology on genomic DNA extracted directly from uncultured peripheral blood has enabled the identification of mosaic chromosome abnormalities that were undetected by conven- tional karyotype analysis. Thus, array CGH has enabled better detec- tion of mosaicism of unbalanced chromosome abnormalities than traditional cytogenetic techniques. Genetic and genomic analyses The pathogenic abnormalities in the human genome vary in size from SNV (locus-specific mutation rates approximately 10–6 to 10–8) to CNV involving entire genes (mutation rate 10–4 to 10–5) to micro- scopically visible chromosome aberrations (found in 1 in 160 new- borns). Despite the broad spectrum of available techniques that have been developed recently to analyse the human genome, there is no single method that can identify all types of genetic and genomic variation (Fig. 3.2.2). Single nucleotide changes and next-generation sequencing Point mutations are commonly analysed using conventional DNA

sequencing with polymerase chain reaction (PCR) amplifica- tion followed by chain termination with fluorescently labelled dideoxynucleotides. However, this method is low-throughput and relatively expensive. A large number of SNPs analysed in genome-wide association studies are currently analysed using hybridization-based oligo- nucleotide microarrays (Table 3.2.4). The available technologies (Affymetrix, Illumina) enable analysis of more than 1 million SNPs in one experiment. Point mutations SNPs LCRs Retrotransposons Down Edwards Turner Charcot–Marie–Tooth disease type 1A Prader Willi DiGeorge Smith–Magenis Potocki–Lupski Friedrich ataxia Huntington Haemophilia A Apert β-Thalassemia Colon cancer Breast cancer Fragile X Cystic fibrosis Recurrent translocation t(11;22)(q23;q11) Disease Repetitive DNA DNA repeats 100 101 102 103 104 105 106 107 108 109 bp 3×109 Mutation size Mutation type Analysis method DNA sequencing PCR Southern analysis Array CGH Dynamic mutations CNVs Non-B DNA Chromosome banding PFGE FISH Fig. 3.2.2 Genomic rearrangements, phenotypic traits, and methods used to assay. Above are shown the traits that can be due to DNA rearrangements. Below are ranges of DNA changes, descriptions of rearrangements, and the methods of assaying different sized changes.

3.2 The genomic basis of medicine 231 Development of massively parallel next-generation sequencing technologies along with the bioinformatic pipelines to analyse large data sets have enabled successful research and clinical implemen- tation of ES. This enables robust, accurate, fast, and cost-effective DNA sequencing of the entire coding portion of the genome in one assay. The amount of the next-generation sequencing data being generated and the characterization of the numerous variants iden- tified are challenging to interpret. However, data sharing among large databases and the 1000 Genomes Project have substantially fa- cilitated appropriate classification of variants and discovery of new disease-causing genes. As a result, ES, which has a diagnostic rate of 20–30%, has revolutionized the diagnosis of mendelian diseases. Most of the identified pathogenic variants are in autosomal dom- inant traits, demonstrating an important role of de novo germline point mutations in both rare and common genetic disorders associ- ated with reduced fitness. WGS has revealed that the average genome harbours 50 to 100 de novo point variants; in ES trio analyses, on average, up to five apparent de novo SNVs (1–2 nonsynonymous and 2–3 synonymous) have been identified per exome. The fre- quency of de novo variants increases with paternal age at a rate of about one new paternally derived variant per every 2 years past the age of 30 years. These variants are readily identified using a parent- child family trio-based exome sequencing approach. Next-generation sequencing technologies also allow for analysis of the entire genome from single cell as well as cell-free DNA from maternal serum. The latter is extremely useful in noninvasive pre- natal screening of most common aneuploidies. Mutational burden and dual molecular diagnoses In addition to the aforementioned digenic or triallelic inheritance, and the two-hit (or second-hit) model, severity of a disease mani- festation can be caused by the abnormal copy-number variation of dosage-sensitive genes. Intrafamilial increase of copy-number of PMP22 due to an NAHR-mediated change from duplication to triplication has been associated with a more severe muscle atrophy of the lower leg and hand muscles, and severe pes cavus deformity due to decreased nerve conduction velocity. Similar phenotypic se- verity phenomenon has been reported for CHRNA7 triplications on chromosome 15q13.3, STS triplications on Xp22.31, as well as homozygous duplication (tetrasomy) of the DiGeorge syndrome critical region on chromosome 22q11.2. Moreover, systematic aggregate ES analyses in multiple unrelated families with CMT-like peripheral neuropathy refractory to pre- vious molecular diagnosis revealed a significantly increased number of rare variants across 58 neuropathy-associated genes in subjects versus controls, suggesting that the combinatorial effect of rare vari- ants contributes to disease burden

and variable expressivity. In contrast to conventional genetic analyses, genomic approaches to disease, such as ES of a large cohort of subjects, have revealed that two or more pathogenic variants can be found in as many as 5% of patients when compared with unrelated control individuals. These studies also facilitated dual molecular diagnoses of ‘blended mendelian phenotypes’ as well as explanation of intra-familial clin- ical variability by multilocus variation among affected siblings. By deconvolution of the complicated phenotypic presentations due to coexistence of multiple genetic conditions, they demonstrated how combinatorial effects of rare variants contribute to disease burden. In dual molecular diagnoses, each disease will segregate according to its known associated trait. This is in contrast with digenic inher- itance whereby both heterozygous variants are required for disease manifestation. Detecting genome structural changes Genomic rearrangements such as deletions, duplications, or inver- sions that are up to 30 kb in size can be detected using the poly- merase chain reaction or Southern blot hybridization. Recently, small genomic rearrangements are detected using next-generation sequencing and digital droplet PCR (ddPCR). Large visible chromosome rearrangements can be analysed using the light microscope by conventional banding techniques (most often G-banding). The detection of genomic changes between 30 kb and 5 Mb in size had remained beyond the level of resolution of available methods until the development of the fluorescent in situ hybridization techniques. Likewise, pulsed-field gel electrophoresis also enabled the resolution of genomic changes of similar magnitude. However, both these technologies are still limited to the examination of specific genomic regions (i.e. they represent locus-specific tests). The development of array CGH and SNP arrays have enabled high-resolution screening of genomic imbalances throughout the entire genome. The level of resolution is dependent on the size and distance between the arrayed interrogating probes. Initially, large genomic clones (bacterial or P1 artificial chromosomes) were im- mobilized and arrayed on glass slides and used as interrogating probes. Such microarrays enabled detection of CNVs throughout the entire human genome with a resolution of approximately 100 kb. The bacterial clones have been replaced by oligonucleotide probes. The currently commercially available arrays have several hundred thousands or millions of oligonucleotide probes. This technology has revolutionized clinical cytogenetics and may re- place much of chromosome analysis with high-resolution genome analysis (Fig. 3.2.3). As an alternative approach to genome-wide screening for the detection of specific large deletions or duplications, a quantitative technique called multiplex ligation-dependent probe amplification based on the polymerase chain reaction, has been developed. This technique relies on sequence-specific probe hybridization to gen- omic DNA, followed by amplification of the hybridized probe and semi-quantitative analysis of the resulting polymerase chain reac- tion products. The relative peak heights or band intensities from each target indicate their initial concentration. This has proven to be an inexpensive, simple, rapid, and sensitive tool to detect dosage alterations in selected genomic regions. More recently, for pre- cise quantitative measurement, ddPCR has been used. The analysed DNA sample is separated into a large number of partitions and the reaction is carried out in each partition individually, providing an absolute quantification of target DNA molecules with previously un- achievable accuracy and sensitivity. Human genetics approaches to drug development Mendelian diseases are a good model to study critical pathways and the knowledge gained could be of benefit to the understanding and therapeutic developments for multifactorial diseases. Hypercholesterolemia and coronary atherosclerosis have been shown to result from elevated levels of low-density lipoprotein (LDL) cholesterol or reduced number of LDL receptors (LDLR).

232 SECTION 3 Cell biology In addition to alterations in LDLR and its ligand, ApoB, hypercholesterolemia has also been shown to be caused by missense gain- of-function mutations in PCSK9 that encodes a serine protease in the secretory pathway. Conversely, nonsense/truncating variants in PCSK9 have been found in individuals with low LDL cholesterol and a substantial reduction in the incidence of coronary events. Thus, inhibitors of PCSK9 became a potential target for therapeutic ap- proaches preventing coronary atherosclerosis. Intravenous or sub- cutaneous administration of monoclonal antibodies to PCSK9 (alirocumab or evolocumab) significantly lowered LDL choles terol levels in clinical studies of healthy subjects and in subjects with familial or nonfamilial hypercholesterolemia. Currently, this approach is considered as a treatment in adults whose cholesterol levels are not controlled by diet and treatment with statins. Recent technological advances in developing antisense oligo- nucleotides (ASOs) have opened a promising and unparalleled po- tential for treatment of monogenic diseases. Some of the therapeutic approaches using ASOs have demonstrated successful phenotypic amelioration both in animal models and in humans. For example, application of ASOs successfully corrected defective pre-mRNA spli- cing of transcripts from the Ush1c gene in a mouse model of human hereditary deafness. Importantly, these effects were sustained for several months, demonstrating the therapeutic potential of ASOs in the treatment of deafness. ASO targeting of long noncoding RNA (Ube3a-ATS) in the mouse model of the genomic imprinting disorder Angelman syn- drome led to specific reduction of Ube3a-ATS and sustained un- silencing of paternal Ube3a in neurons both in vitro and in vivo. As a result, partial restoration of UBE3A protein enabled reversal of imprinting ameliorating some cognitive deficits associated with the disease. Humans have two near identical copies of the survival motor neuron gene: SMN1 and SMN2. C to T transition (C6T) within exon 7 of SMN2 disrupts a modulator of splicing, leading to the exclusion of exon 7 from c.90% of the mRNA transcript. Deletion or mutation of SMN1 combined with the inability of SMN2 to com- pensate for the loss of SMN1 results in spinal muscular atrophy (SMA), a severe lethal neurodegenerative disease. Diverse treat- ment strategies aimed at improving the function of SMN2 have been envisioned (e.g. manipulation of transcription, correction of aberrant splicing, and stabilization of SMN mRNA). Several studies applying ASOs targeting the splicing site in SMN2 demon- strated successful elevation of SMN protein from SMN2 and im- proved survival and function. 100 101 102 103 104 105 106 107 108 109 bp DNA sequencing PFGE / FISH Oligonucleotide microarrays bp 1 Human male G-bands 6 13 19 20 21 22 X Y 14 15 16 17 18 7 8 9 10 11 12 2 3 4 5 Chromosome banding Fig. 3.2.3 Genome architecture and methods to resolve structure of varying DNA. Above is shown a scale of the human genome from 1(100) bp to 3 × 109 bp and the size ranges (colour coded) in which the different methods can physically resolve differences. Chromosomal banding (green) examines the whole genome at once, but cannot resolve changes of more than c.5 Mb (106–107 bp) in size. DNA sequencing (purple) can resolve single nucleotide changes and changes of several bases, but cannot identify CNVs. Pulsed-field gel electrophoresis (PFGE) and FISH (yellow) extend the reach of conventional karyotyping and resolve changes from 104 to 106 bp in size. Array CGH can resolve changes causing genomic imbalance from 103 to 108 bp (including aneuploidies), simultaneously performing thousands of locus-specific FISH procedures as well as detecting imbalances seen by chromosome analysis. Adapted from Lupski (2003). 2002 Curt Stern Award Address. Genomic disorders recombination-based disease resulting
from genomic architecture. Am J Hum Genet, 72, 246–52; Lupski JR (2007). Genomic rearrangements and sporadic disease. Nat Genet, 39, S43–7.

3.2 The genomic basis of medicine 233 Conclusions In a classical mendelian monogenic model of a disease, Watson– Crick DNA base-pair changes in a single gene are recognized as a mechanism affecting the structure, function, or regulation of the en- coded protein. Completion of the human reference DNA sequence and recent advances in novel technologies that enable us to study the entire human genome of a given patient have extended our view of the genetic bases of disease in humans. It has become apparent that many disease traits are caused by genomic alterations rather than by single nucleotide changes. The genetic heterogeneity of several complex traits is being resolved. Also, the contributions of variant alleles at more than one locus in a given personal genome to disease manifestations are being better understood. Genome-wide studies have led to important discoveries of large- scale CNVs in the human genome. The clinical consequences of the overwhelming majority of CNVs are not known. Many, if not most, CNVs are likely benign but some have been shown to be responsible for mendelian traits and others lead to increased susceptibility for complex traits. Personalized and precision genomic medicine The concept of personalized medicine has been developed with the Human Genome Project. In contrast to conventional medicine, where the patients' diagnoses and treatments are based on disease signs and symptoms, personalized medicine refers to the genetic bases of the patient's traits and susceptibility to traits. The hypoth- esis underlying personalized genomic medicine is that personalized medical care can be guided by the unique genomic content of an in- dividual patient. The aim of personal genomic medicine is the inter- pretation of unique information encoded in the individual patient's genome to be able to anticipate genetic risks and liability and ad- just personal lifestyle changes, diet, medications, prevention, and therapy to mitigate the consequences of genetic risk. More recently, to avoid misinterpretations that unique treatments can be designed for each individual, the term precision medicine has been coined. In precision medicine, individuals can be classified into subpopulations that differ in their biology, susceptibility, prog- nosis, development, or in their response to a specific treatment for a particular disease. The increasing ability to assay an individual's DNA poly morphisms (both SNPs and CNVs) will continue to further en- able prediction of personal responses to different drugs depending on an individual's genetic background (i.e. pharmacogenomics). With the clinical implementation of new technologies, including massive parallel sequencing and high-resolution oligonucleotide array CGH and SNP arrays that offer analysis of the individual diploid human genome (DNA sequence and CNVs) within a rela- tively short time, the information content of entire genomes of in- dividuals is expected to become affordable. Recent whole-genome studies, however, suggest that interpretation of the complexity of the genetic load of an individual or selected patients will require better understanding of genotype/phenotype correlations to pro- vide clinically relevant information in a format commensurate with clinical implementation. Such an approach will potentially revolu- tionize clinical diagnostics and therapy and may provide tremen- dous benefits for the patients' health. FURTHER READING Akawi N, et al. (2015). Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat Genet, 47, 1363–9. Audano PA, et al. (2019). Characterizing the major structural variant alleles of the human genome. Cell, 176, 663–75.e19. Badano JL, et al. (2006). Dissection of epistasis in oligogenic Bardet– Biedl syndrome. Nature, 439, 326–30. Badano JL, Katsanis N (2002). Beyond Mendel: an evolving view of human genetic disease transmission. Nat Rev Genet, 3, 779–89. Ballif BC, et al. (2006). Detection of low-level mosaicism by array CGH in routine diagnostic specimens. Am J Med Genet A, 140, 2757–67. Barbouti A, et al. (2004). The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a 220 kb region containing palindromic low-copy repeats. Am J Hum Genet, 74, 1–10. Baudat F, et al. (2010).

PRDM9 is a major determinant of meiotic re- combination hotspots in humans and mice. Science, 327, 836–40. Beck CR, et al. (2019). Megabase length hypermutation accompanies human structural variation at 17p11.2. Cell, 176, 1310–24.e10. Bentley DR (2006). Whole-genome re-sequencing. Curr Opin Genet Dev, 16, 545–52. Berg IL, et al. (2010). PRDM9 variation strongly influences recombin- ation hot-spot activity and meiotic instability in humans. Nat Genet, 42, 859–63. Carvalho MBC, Lupski JR (2016). Mechanisms underlying struc- tural variant formation in genomic disorders. Nat Rev Genet, 17, 224–38. Chance PF, et al. (1994). Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromo- some 17. Hum Mol Genet, 3, 223–8. Cheung SW, et al. (2007). Microarray-based CGH detects chromo- somal mosaicism not revealed by conventional cytogenetics. Am J Med Genet, 143, 1679–86. Chong JX, et al. (2015). The genetic basis of mendelian pheno- types: discoveries, challenges, and opportunities. Am J Hum Genet, 97, 199–215. Cohen JC, et al. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med, 354, 1264–72. Cooper DN, Youssoufian H (1998). The CpG dinucleotide and human genetic disease. Hum Genet, 78, 151–5. Costanzo M, et al. (2019). Global genetic networks and the genotype- to-phenotype relationship. Cell, 177, 85–100. Coulondre C, et al. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature, 274, 775–80. Crespi B, Stead P, Elliot M (2010). Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizo- phrenia. Proc Natl Acad Sci U S A, 107 Suppl 1, 1736–41. Deciphering Developmental Disorders Study (2015). Large-scale dis- covery of novel genetic causes of developmental disorders. Nature, 519, 223–8. Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature, 542, 433–8. Dharmadhikari AV, et al. (2019). Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,020 clinical exome cases. Genome Med, 11, 30.

 Dipple KM, McCabe ER (2000). Phenotypes of patients with ‘simple’ Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. Am J Hum Genet, 66, 1729–35. Dumas L, et al. (2007). Gene copy-number variation spanning 60 million years of human and primate evolution. Genome Res, 17, 1266–77. Edelmann L, et al. (2001). AT-rich palindromes mediate the constitu- tional t (11;22) translocation. Am J Hum Genet, 68, 1–13. Eichers ER, et al. (2004). Triallelic inheritance: a bridge between Mendelian and multifactorial traits. Ann Med, 36, 262–72. Eldomery MK, et al. (2017). Lessons learned from additional research analyses of unsolved clinical exome cases. Genome Med, 9, 26. ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 306, 636–40. Firth HV, et al. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet, 84, 524–33. Gabriel SB, et al. (2002). Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat Genet, 31, 89–93. Giglio S, et al. (2002). Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. Am J Hum Genet, 71, 276–85. Girirajan S, et al. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet, 42, 203–9. Girirajan S, et al. (2012). Phenotypic heterogeneity of genomic dis-orders and rare copy-number variants. N Engl J Med, 367, 1321–31. Gonzaga-Jauregui C, et al. (2015). Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and com- plex neuropathy. Cell Rep, 12, 1169–83. Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012). Human genome sequencing in health and disease. Annu Rev Med, 63, 35–61. Hastings PJ, et al. (2009). Mechanisms of change in gene copy-number. Nat Rev Genet, 10, 551–64. Iafrate

AJ, et al. (2004). Detection of large-scale variation in the human genome. Nat Genet, 36, 949–51. Inoue K, et al. (2004). Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. Nat Genet, 36, 361–9. International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature, 409, 860–921. International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature, 431, 931–45. Karaca E, et al. (2018). Phenotypic expansion illuminates multilocus pathogenic variation. Genet Med, 20, 1528–37. Karolak J, et al. (2019). Complex compound inheritance of lethal lung developmental disorders due to disruption of the TBX-FGF pathway. Am J Hum Genet, 104, 213–28. Kato T, et al. (2006). Genetic variation affects de novo translocation frequency. Science, 311, 971. Katsanis N, et al. (2001). Triallelic inheritance in Bardet–Biedl syn- drome, a Mendelian recessive disorder. Science, 293, 2256–9. Kong A, et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature, 488, 471–5. Kurahashi H, et al. (2000). Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). Hum Mol Genet, 9, 1665–70. Lappalainen T, et al. (2019). Genomic analysis in the age of human genome sequencing. Cell, 177, 70–84. Lee H, et al. (2014). Clinical exome sequencing for genetic identifica- tion of rare Mendelian disorders. JAMA, 312, 1880–7. Lee JA, Carvalho CMB, Lupski JR (2007). A DNA replication mech- anism for generating nonrecurrent rearrangements associated with genomic disorders. Cell, 131, 1235–47. Lee JA, Lupski JR (2006). Genomic rearrangements and gene copy- number alterations as a cause of nervous system disorders. Neuron, 52, 103–21. Lentz JJ, et al. (2013). Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. Nat Med, 19, 345–50. Levy S, et al. (2007). The diploid genome sequence of an individual human. PLoS Biol, 5, e254. Lifton RP, et al. (1992). A chimaeric 11-beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. Nature, 355, 262–5. Lindhurst MJ, et al. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med, 365, 611–19. Liu J, et al. (2019). TBX6-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and TBX6 gene dosage model. Genet Med, 21, 1548–58. Liu P, et al. (2011). Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell, 146, 889–903. Liu P, et al. (2014). Mechanism, prevalence, and more severe neur- opathy phenotype of the Charcot–Marie–Tooth type 1A triplication. Am J Hum Genet, 94, 462–9. Liu P, et al. (2017). An Organismal CNV Mutator Phenotype Restricted to Early Human Development. Cell, 168, 830-42.e7. Liu P, et al. (2019). Reanalysis of clinical exome sequencing data.
N Engl J Med, 380, 2478–80. Lupski JR (2006). Genome structural variation and sporadic disease traits. Nat Genet, 38, 974–6. Lupski JR (2007). Genomic rearrangements and sporadic disease. Nat Genet, 39, S43–7. Lupski JR (2007). Structural variation in the human genome. N Engl J Med, 356, 1169–71. Lupski JR, et al. (1991). DNA duplication associated with Charcot– Marie–Tooth disease type 1A. Cell, 66, 219–32. Lupski JR, et al. (1992). Gene dosage is a mechanism for Charcot–Marie–Tooth disease type 1A. Nat Genet, 1, 29–33. Lupski JR, et al. (2010). Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. N Engl J Med, 362, 1181–91. Lupski JR, et al. (2011). Clan genomics and the complex architecture of human disease. Cell, 147, 32–43. Lupski JR, Stankiewicz P (2005). Genomic disorders: molecular mech- anisms for rearrangements and conveyed phenotypes. PLoS Genet, 1, e49. Lupski JR, Stankiewicz P (eds) (2006). Genomic disorders: the genomic basis of disease. Humana Press, Totowa. Lupski JR (2015). Structural variation mutagenesis of the human genome: impact on disease and evolution. Environ Mol

Mutagen, 56, 419–36. Männik K, et al. (2015). Copy-number variations and cognitive pheno- types in unselected populations. JAMA, 313, 2044–54. Martin HC, et al. (2018). Quantifying the contribution of recessive coding variation to developmental disorders. Science, 362, 1161–4. Meng L, et al. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. Nature, 518, 409–12. Myers S, et al. (2010). Drive against hotspot motifs in primates im- plicates the PRDM9 gene in meiotic recombination. Science, 327, 876–9.

 Niemi MEK, et al. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature, 562, 268–71. Pentao L, et al. (1992). Charcot–Marie–Tooth type 1A duplication ap- pears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. Nat Genet, 2, 292–300. Posey I, et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. N Engl J Med, 376, 21–31. Posey J, et al. (2019). Insights into genetics, human biology and dis- ease gleaned from family based genomic studies. Genet Med, 21, 798–812. Posey JE, et al. (2015). Molecular diagnostic experience of whole- exome sequencing in adult patients. Genet Med, 18, 678–85. Redon R, et al. (2006). Global variation in copy-number in the human genome. Nature, 444, 444–54. Rosenfeld JA, et al. (2013). Estimates of penetrance for recurrent pathogenic copy-number variations. Genet Med. 15, 478–81. Schmickel RD (1986). Contiguous gene syndromes: a component of recognizable syndromes. J Pediatr, 109, 231–41. Scriver CR, Waters PJ (1999). Monogenic traits are not simple: lessons from henylketonuria. Trends Genet, 15, 267–72. Sebat J, et al. (2004). Large-scale copy-number polymorphism in the human genome. Science, 305, 525–8. Short PJ, et al. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. Nature, 555, 611–16. Sifrim A, et al. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet, 48, 1060–5. Song X, et al. (2018). Predicting human genes susceptible to gen- omic instability associated with Alu/Alu-mediated rearrangements. Genome Res, 28, 1228–42. Spence JE, et al. (1988). Uniparental disomy as a mechanism for human genetic disease. Am J Hum Genet, 42, 217–26. Stankiewicz P, Beaudet AL (2007). Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idio- pathic mental retardation. Curr Opin Genet Dev, 17, 182–92. Stefansson H, et al. (2005). A common inversion under selection in Europeans. Nat Genet, 37, 129–37. Stein EA, et al. (2012). Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. N Engl J Med, 366, 1108–18. Sudmant PH, et al. (2015). An integrated map of structural variation in 2,504 human genomes. Nature, 526, 75–81. The International HapMap Consortium (2003). The International HapMap Project. Nature, 426, 789–96. Tijo JH, Levan A (1956). The chromosome number of man. Hereditas, 42, 1–6. Todd JA, et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet, 39, 857–64. Turner DJ, et al. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat Genet, 40, 90–5. Verbitsky M, et al. (2019). The copy number variation landscape of con- genital anomalies of the kidney and urinary tract. Nat Genet, 51, 117–27. Watson JD, Crick FH (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171, 737–8. Wells RD (2007). Non-B DNA conformations, mutagenesis and dis- ease. Trends Biochem Sci, 32, 271–8. Wheeler DA, et al. (2008). The complete genome of a single individual by massively parallel DNA sequencing. Nature, 452, 872–6. Willingham AT, Gingeras TR (2006). TUF love for ‘junk’ DNA. Cell, 125, 1215–20. Wright CF, et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet, 385, 1305–14. Wu N, et al. (2015). TBX6 null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med, 372, 341–50. Yang Y, et al.

(2013). Clinical whole-exome sequencing for the diag- nosis of mendelian disorders. N Engl J Med, 369, 1502–11. Yang Y, et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. JAMA, 312, 1870–9. Yang N, et al. (2019). TBX6 compound inheritance leads to congenital vertebral malformations in humans and mice. Hum Mol Genet, 28, 539–47. Zhang F, Carvalho CM, Lupski JR (2009). Complex human chromo- somal and genomic rearrangements. Trends Genet, 25, 298–307. Zhang F, et al. (2009). Copy-number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet, 10, 451–81.

Revision #1
Created 2026-01-22 16:44:12 UTC by Omar Ayman
Updated 2026-01-22 16:44:12 UTC by Omar Ayman