

03-3 Clinical genetics

3 Clinical genetics

Clinical genetics K Tatton-Brown DR FitzPatrick The fundamental principles of genomics 38 The packaging of genes: DNA, chromatin and chromosomes 38 From DNA to protein 38 Non-coding RNA 40 Cell division, differentiation and migration 40 Cell death, apoptosis and senescence 41 Genomics, health and disease 42 Classes of genetic variant 42 Consequences of genomic variation 44 Normal genomic variation 45 Constitutional genetic disease 46 Somatic genetic disease 50 Interrogating the genome: the changing landscape of genomic technologies 51 Looking at chromosomes 51 Looking at genes 52 Genomics and clinical practice 56 Genomics and health care 56 Treatment of genetic disease 58 Ethics in a genomic age 59

38 • CLINICAL GENETICS via the production of messenger ribonucleic acid (mRNA) to the production of proteins. The human genome contains over 20 000 genes, although many of these are inactive or silenced in different cell types, reflecting the variable gene expression responsible for cell-specific characteristics. The central dogma is the pathway describing the basic steps of protein production: transcription, splicing, translation and protein modification (Fig. 3.2). Although this is now recognised as an over-simplification (contrary to this linear relationship, a single gene will often encode many different proteins), it remains a useful starting point to explore protein production. Transcription: DNA to messenger RNA Transcription describes the production of ribonucleic acid (RNA) from the DNA template. For transcription to commence, an enzyme called RNA polymerase binds to a segment of DNA at the start of the gene: the promoter. Once bound, RNA polymerase moves along one strand of DNA, producing an RNA molecule complementary to the DNA template. In protein-coding genes this is known as messenger RNA (mRNA). A DNA sequence close to the end of the gene, called the polyadenylation signal, acts as a signal for termination of the RNA transcript (Fig. 3.3). We have entered a genomic era. Powerful new technologies are driving forward transformational change in health care. Genetic sequencing has evolved from the targeted sequencing of a single gene to the parallel sequencing of multiple genes. In addition to improving the chances of identifying a genetic cause of rare diseases, these technologies are increasingly directing therapies and, in the future, are likely to be used in the diagnosis and prevention of common diseases such as diabetes. In this chapter we explore the fundamentals of genomics, the basic principles underlying these new genomic technologies and how the data generated can be applied safely for patient benefit. We will review the use of genomic technology across a breadth of medical specialties, including obstetrics, paediatrics, oncology and infectious disease, and consider how health care is likely to be transformed by technology over the coming decade. Finally, we will consider the ethical impact that these technologies are likely to have, both for the individual and for their wider family. The fundamental principles of genomics The packaging of genes: DNA, chromatin and chromosomes Genes are

functional units encoded in double-stranded deoxyribonucleic acid (DNA), packaged as chromosomes and located in the nucleus of the cell: a membrane-bound compartment found in all cells except erythrocytes and platelets (Fig. 3.1). DNA consists of a linear sequence of just four bases: adenine (A,) cytosine (C), thymine (T) and guanine (G.) It forms a 'double helix', a twisted ladder-like structure formed from two complementary strands of DNA joined by hydrogen bonds between bases on the opposite strand that can form only between a C and a G base and an A and a T base. It is this feature of DNA that enables faithful DNA replication and is the basis for many of the technologies designed to interrogate the genome: when the DNA double helix 'unzips', one strand can act as a template for the creation of an identical strand. A single copy of the human genome comprises approximately 3.1 billion base pairs of DNA, wound around proteins called histones. The unit consisting of 147 base pairs wrapped around four different histone proteins is called the nucleosome. Sequences of nucleosomes (resembling a string of beads) are wound and packaged to form chromatin: tightly wound, densely packed chromatin is called heterochromatin and open, less tightly wound chromatin is called euchromatin. The chromatin is finally packaged into the chromosomes. Humans are diploid organisms: the nucleus contains two copies of the genome, visible microscopically as 23 chromosome pairs (known as the karyotype). Chromosomes 1 through to 22 are known as the autosomes and consist of identical chromosomal pairs. The 23rd 'pair' of chromosomes are the two sex chromosomes: females have two X chromosomes and males an X and Y chromosome. A normal female karyotype is therefore written as 46,XX and a normal male is 46,XY. From DNA to protein Genes are functional elements on the chromosome that are capable of transmitting information from the DNA template Fig. 3.1 The packaging of DNA, genes and chromosomes. From bottom to top: the double helix and the complementary DNA bases; chromatin; and a normal female chromosome pattern – the karyotype. DNA helix Histones Chromatin Chromosome Normal female karyotype Nucleosome A T G A C G G A T T A C T G C C T A

The fundamental principles of genomics • 39

Fig. 3.2 The central dogma of protein production. Double-stranded DNA as a template for single-stranded RNA, which codes for the production of a peptide chain of amino acids. Each of these chains has an orientation. For DNA and RNA, this is 5' to 3'. For peptides, this is N-terminus to C-terminus. DNA 5'CGATTC3' 3'GCTAAG5' 5'CGAUUC3' N_ArgPhe_C RNA Protein Transcription Translation Fig. 3.3 RNA synthesis and its translation into protein. Gene transcription involves binding of RNA polymerase II to the promoter of genes being transcribed with other proteins (transcription factors) that regulate the transcription rate. The primary RNA transcript is a copy of the whole gene and includes both introns and exons, but the introns are removed within the nucleus by splicing and the exons are joined to form the messenger RNA (mRNA). Prior to export from the nucleus, a methylated guanosine nucleotide is added to the 5' end of the RNA ('cap') and a string of adenine nucleotides is added to the 3' ('polyA tail'). This protects the RNA from degradation and facilitates transport into the cytoplasm. In the cytoplasm, the mRNA binds to ribosomes and forms a template for protein production. (tRNA = transfer RNA; UTR = untranslated region) Protein product N-term C-term cap Messenger RNA (mRNA) Messenger RNA (mRNA) AAAAA tRNAs Ribosome Nuclear membrane Primary RNA transcript Spliceosome RNA export to cytoplasm cap cap AAAAA PolyA tail 3'UTR 5'UTR Nuclear pore AAAAA Splicing Transcription 3' 5' 3' Sense strand Enhancer Transcription factors RNA polymerase II Exon 1 Intron 2 Intron 1 Exon 2 Exon 3 Exon 1 Exon 2 Exon 3 Promoter Nucleolus Nuclear membrane Gene A Gene B Gene C Active gene RNA Nucleus Translation RNA differs from DNA in three main ways: • RNA is single-stranded. • The

sugar residue within the nucleotide is ribose, rather than deoxyribose. • It contains uracil (U) in place of thymine (T). The activity of RNA polymerase is regulated by transcription factors. These proteins bind to specific DNA sequences at the promoter or to enhancer elements that may be many thousands of base pairs away from the promoter; a loop in the chromosomal DNA brings the enhancer close to the promoter, enabling the bound proteins to interact. The human genome encodes more than 1200 different transcription factors. Mutations within transcription factors, promoters and enhancers can cause disease. For example, the blood disorder alpha-thalassaemia is usually caused by gene deletions (see p. 954 and Box 3.4). However, it can also result from a mutation in an enhancer located more than 100 000 base pairs (bp) from the α -globin gene promoter, leading to greatly reduced transcription. Gene activity, or expression, is influenced by a number of complex interacting factors, including the accessibility of the gene promoter to transcription factors. DNA can be modified by the addition of a methyl group to cytosine molecules (methylation). If DNA methylation occurs in promoter regions, transcription is silenced, as methyl cytosines are usually not available for transcription factor binding. A second mechanism determining promoter accessibility is the structural configuration of chromatin. In open chromatin, called euchromatin, gene promoters are accessible to RNA polymerase and transcription factors; therefore it is transcriptionally active. This contrasts with heterochromatin, which is densely packed and transcriptionally silent. The chromatin configuration is determined by modifications (such as methylation or acetylation) of specific amino acid residues of histone protein tails. Modifications of DNA and histones are termed epigenetic ('epi-' meaning 'above' the genome), as they do not alter the primary sequence of the DNA code but have biological significance in chromosomal function. Abnormal epigenetic changes are increasingly recognised as important events in the progression of cancer, allowing expression of normally silenced genes that result in cancer cell de-differentiation and proliferation. They also afford therapeutic targets. For instance, the histone deacetylase inhibitor vorinostat has been successfully used to treat cutaneous T-cell lymphoma, due to the re-expression of genes that had

40 • CLINICAL GENETICS into the cytoplasm or packaged into vesicles for secretion. The clinical importance of post-translational modification of proteins is shown by the severe developmental, neurological, haemostatic and soft tissue abnormalities that are associated with the many different congenital disorders of glycosylation. Post-translational modifications can also be disrupted by the synthesis of proteins with abnormal amino acid sequences. For example, the most common mutation in cystic fibrosis ($\Delta F508$) results in an abnormal protein that cannot be exported from the ER and Golgi (see Box 3.4). Non-coding RNA Approximately 4500 genes in humans encode non-coding RNAs (ncRNA) rather than proteins. There are various categories of ncRNA, including transfer RNA (tRNA), ribosomal RNA (rRNA), ribozymes and microRNA (miRNA). The miRNAs, which number over 1000, have a role in post-translational gene expression: they bind to mRNAs, typically in the 3'UTR, promoting target mRNA degradation and gene silencing. Together, miRNAs affect over half of all human genes and have important roles in normal development, cancer and common degenerative disorders. This is the subject of considerable research interest at present. Cell division, differentiation and migration In normal tissues, molecules such as hormones, growth factors and cytokines provide the signal to activate the cell cycle: a controlled programme of biochemical events that culminates in cell division. In all cells of the body, except the gametes (the sperm and egg cells, also known as the germ line), mitosis completes cell division, resulting in two diploid daughter cells. In contrast, the sperm and eggs cells complete cell division with meiosis, resulting in four haploid daughter cells (Fig. 3.4). The stages of cell division in the non-germ-line,

somatic cells are shown below:

- Cells not committed to mitosis are said to be in G₀.
- Cells committed to mitosis must go through the preparatory phase of interphase consisting of G₁, S and G₂:
- G₁ (first gap): synthesis of the cellular components necessary to complete cell division
- S (synthesis): DNA replication producing identical copies of each chromosome called the sister chromatids
- G₂ (second gap): repair of any errors in the replicated DNA before proceeding to mitosis.
- Mitosis (M) consists of four phases:
- Prophase: the chromosomes condense and become visible, the centrioles move to opposite ends of the cell and the nuclear membrane disappears.
- Metaphase: the centrioles complete their migration to opposite ends of the cell and the chromosomes – consisting of two identical sister chromatids – line up at the equator of the cell.
- Anaphase: spindle fibres attach to the chromosome and pull the sister chromatids apart.
- Telophase: the chromosomes decondense, the nuclear membrane reforms and two daughter cells – each with 46 chromosomes – are formed. The progression from one phase to the next is tightly controlled by cell-cycle checkpoints.

For example, the checkpoint between previously been silenced in the tumour. These genes encode transcription factors that promote T-cell differentiation as opposed to proliferation, thereby causing tumour regression.

RNA splicing, editing and degradation

Transcription produces an RNA molecule that is a copy of the whole gene, termed the primary or nascent transcript. This nascent transcript then undergoes splicing, whereby regions not required to make protein (the intronic regions) are removed while those segments that are necessary for protein production (the exonic regions) are retained and rejoined. Splicing is a highly regulated process that is carried out by a multimeric protein complex called the spliceosome. Following splicing, the mRNA molecule is exported from the nucleus and used as a template for protein synthesis. Many genes produce more than one form of mRNA (and thus protein) by a process termed alternative splicing, in which different combinations of exons are joined together. Different proteins from the same gene can have entirely distinct functions. For example, in thyroid C cells the calcitonin gene produces mRNA encoding the osteoclast inhibitor calcitonin (p. 634), but in neurons the same gene produces an mRNA with a different complement of exons via alternative splicing that encodes a neurotransmitter, calcitonin-gene-related peptide (p. 772).

Translation and protein production

Following splicing, the segment of mRNA containing the code that directs synthesis of a protein product is called the open reading frame (ORF). The inclusion of a particular amino acid in the protein is specified by a codon composed of three contiguous bases. There are 64 different codons with some redundancy in the system: 61 codons encode one of the 20 amino acids, and the remaining three codons – UAA, UAG and UGA (known as stop codons) – cause termination of the growing polypeptide chain. ORFs in humans most commonly start with the amino acid methionine. All mRNA molecules have domains before and after the ORF called the 5' untranslated region (UTR) and 3'UTR, respectively. The start of the 5'UTR contains a cap structure that protects mRNA from enzymatic degradation, and other elements within the 5'UTR are required for efficient translation. The 3'UTR also contains elements that regulate efficiency of translation and mRNA stability, including a stretch of adenine bases known as a polyA tail (see Fig. 3.3). The mRNAs then leave the nucleus via nuclear pores and associate with ribosomes, the sites of protein production (see Fig. 3.3). Each ribosome consists of two subunits (40S and 60S), which comprise non-coding rRNA molecules (see Fig. 3.9, p. 50) complexed with proteins. During translation, a different RNA molecule known as transfer RNA (tRNA) binds to the ribosome. The tRNAs deliver amino acids to the ribosome so that the newly synthesised protein can be assembled in a stepwise fashion. Individual tRNA molecules bind a specific amino acid and 'read' the mRNA ORF via an 'anticodon' of three nucleotides that is complementary to the codon in mRNA (see Fig. 3.3). A proportion of ribosomes is bound to the membrane of the endoplasmic reticulum (ER), a complex

tubular structure that surrounds the nucleus. Proteins synthesised on these ribosomes are translocated into the lumen of the ER, where they undergo folding and processing. From here, the protein may be transferred to the Golgi apparatus, where it undergoes post-translational modifications, such as glycosylation (covalent attachment of sugar moieties), to form the mature protein that can be exported

The fundamental principles of genomics • 41

(prophase, metaphase, anaphase and telophase) but differs from mitosis in the following ways: • It consists of two separate cell divisions known as meiosis I and meiosis II. • It reduces the chromosome number from the diploid to the haploid number via a tetraploid stage, i.e. from 46 to 92 (MI S) to 46 (MI M) to 23 (MII M) chromosomes, so that when a sperm cell fertilises the egg, the resulting zygote will return to a diploid, 46, chromosome complement. This reduction to the haploid number occurs at the end of meiosis II. • The 92 chromosome stage consists of 23 homologous pairs of sister chromatids, which then swap genetic material, a process known as recombination. This occurs at the end of MI prophase and ensures that the chromosome that a parent passes to his or her offspring is a mix of the chromosomes that the parent inherited from his or her own mother and father. The individual steps in meiotic cell division are similar in males and females. However, the timing of the cell divisions is very different. In females, meiosis begins in fetal life but does not complete until after ovulation. A single meiotic cell division can thus take more than 40 years to complete. As women become older, the separation of chromosomes at meiosis II becomes less efficient. That is why the risk of trisomies (p. 44) due to non-disjunction grows greater with increasing maternal age. In males, meiotic division does not begin until puberty and continues throughout life. In the testes, both meiotic divisions are completed in a matter of days. Cell death, apoptosis and senescence With the exception of stem cells, human cells have only a limited capacity for cell division. The Hayflick limit is the number of divisions a cell population can go through in culture before division stops and enters a state known as senescence. This 'biological clock' is of great interest in the study of the normal ageing process. Rare human diseases associated with premature ageing, called progeric syndromes, have been very helpful in identifying the importance of DNA repair mechanisms in senescence (p. 1034). For example, in Werner's syndrome, a DNA helicase (an enzyme that separates the two DNA strands) is mutated, leading to failure of DNA repair and premature ageing. A distinct mechanism of cell death is seen in apoptosis, or programmed cell death. Apoptosis is an active process that occurs in normal tissues and plays an important role in development, tissue remodelling and the immune response. The signal that triggers apoptosis is specific to each tissue or cell type. This signal activates enzymes, called caspases, which actively destroy cellular components, including chromosomal DNA. This degradation results in cell death, but the cellular corpse contains characteristic vesicles called apoptotic bodies. The corpse is then recognised and removed by phagocytic cells of the immune system, such as macrophages, in a manner that does not provoke an inflammatory response. A third mechanism of cell death is necrosis. This is a pathological process in which the cellular environment loses one or more of the components necessary for cell viability. Hypoxia is probably the most common cause of necrosis. Fig. 3.4 Meiosis and gametogenesis: the main chromosomal stages of meiosis in both males and females. A single homologous pair of chromosomes is represented in different colours. The final step is the production of haploid germ cells. Each round of meiosis in the male results in four sperm cells; in the female, however, only one egg cell is produced, as the other divisions are sequestered on the periphery of the mature egg as peripheral

polar bodies. Egg Sperm Father Mother Meiotic cell divisions 1st polar bodies 2nd polar body DNA replication Sister chromatids Homologous pairing Swapping of genetic material between homologues: Recombination Individual chromosome pair (homologues) Non-disjunction of chromosomes is a common error in human meiosis, resulting in trisomy of individual chromosomes or uniparental disomy (both chromosomes from single parent) G2 and mitosis ensures that all damaged DNA is repaired prior to segregation of the chromosomes. Failure of these control processes is a crucial driver in the pathogenesis of cancer, as discussed on page 1316. • Meiosis is a special, gamete-specific, form of cell division (Fig. 3.4). Like mitosis, meiosis consists of four phases

42 • CLINICAL GENETICS mutation (Box 3.1 and Fig. 3.5). If a multiple of three nucleotides is involved, this is in-frame. If an indel change affects one or two nucleotides within the ORF of a protein-coding gene, this can have serious consequences because the triple nucleotide sequence of the codons is disrupted, resulting in a frameshift mutation. The effect on the gene is typically severe because the amino acid sequence is totally disrupted. Fig. 3.5 Different types of mutation affecting coding exons. A Normal sequence. B A synonymous nucleotide substitution changing the third base of a codon; the resulting amino acid sequence is unchanged. C A missense mutation in which the nucleotide substitution results in a change in a single amino acid from the normal sequence (AAG) encoding lysine to glutamine (CAG). D Insertion of a G residue (boxed) causes a frameshift mutation, completely altering the amino acid sequence downstream. This usually results in a loss-of-function mutation. E A nonsense mutation resulting in a single nucleotide change from a lysine codon (AAG) to a premature stop codon (TAG). Normal Silent polymorphism (no amino acid change) Missense mutation causing Lys-Gln amino acid change 'G' insertion causing frameshift mutation Nonsense mutation causing premature termination codon A B C D E

3.1 Classes of genetics variant The classes of genetic variant can be illustrated using the sentence 'THE FAT FOX WAS ILL COS SHE ATE THE OLD CAT' Synonymous Silent polymorphism with no amino acid change THE FAT FOX WAS ILL COS SHE ATE THE OLD KAT where the C is replaced with a K but the meaning remains the same Non-synonymous Causing an amino acid change THE FAT BOX WAS ILL COS SHE ATE THE OLD CAT where the F of FOX is replaced by a B and the original meaning of the sentence is lost Stop gain (also called a nonsense mutation) Causing the generation of a premature stop codon THE CAT where the F of FAT is replaced by a C generating a premature stop codon Indel Where bases are either inserted or deleted; disruption of the reading frame is dependent on the number of bases inserted or deleted THE FAT FOX WAS ILL ILL COS SHE ATE THE OLD CAT where the insertion of three bases results in maintenance of the reading frame THE FAT FOX WAW ASI LLC OSS HEA TET HEO LDC AT where the insertion of two bases results in disruption of the reading frame

Genomics, health and disease Classes of genetic variant There are many different classes of variation in the human genome, categorised by the size of the DNA segment involved and/or by the mechanism giving rise to the variation. Nucleotide substitutions The substitution of one nucleotide for another is the most common type of genomic variation. Depending on their frequency and functional consequences, these changes are known as point mutations or single nucleotide polymorphisms (SNPs). They occur by misincorporation of a nucleotide during DNA synthesis or by chemical modification of the base. When these substitutions occur within ORFs of a protein-coding gene, they are further classified into: • synonymous - resulting in a change in the codon without altering the amino acid • non-synonymous (also known as a missense mutation) - resulting in a change in the codon and the encoded amino acid • stop gain (or nonsense mutation) - introducing a premature stop codon and resulting in truncation of the protein • splicing - taking place at splice

sites that most frequently occur at the junction between an intron and an exon. These different types of mutation are illustrated in Box 3.1 and examples are shown in Figures 3.5 and 3.6. Insertions and deletions One or more nucleotides may be inserted or lost in a DNA sequence, resulting in an insertion/deletion (indel) polymorphism or

Genomics, health and disease • 43

size of the original repeat, in that longer repeats tend to be more unstable. Many microsatellites and minisatellites occur in introns or in chromosomal regions between genes and have no obvious adverse effects. However, some genetic diseases are caused by microsatellite repeats that result in duplication of amino acids within the affected gene product or affect gene expression (Box 3.2). Simple tandem repeat mutations Variations in the length of simple tandem repeats of DNA are thought to arise as the result of slippage of DNA during meiosis and are termed microsatellite (small) or minisatellite (larger) repeats. These repeats are unstable and can expand or contract in different generations. This instability is proportional to the Fig. 3.6 Splice site mutations. A The normal sequence is shown, illustrating two exons, and intervening intron (blue) with splice donor (AG) and splice acceptor sites (GT) underlined. Normally, the intron is removed by splicing to give the mature messenger RNA that encodes the protein. B In a splice site mutation the donor site is mutated. As a result, splicing no longer occurs, leading to read-through of the mRNA into the intron, which contains a premature termination codon downstream of the mutation. Normal Splice site mutation Splice donor site Exon Exon Intron Exon Exon Intron Intron removed by splicing Splice acceptor site mRNA 'reads through' intron Abnormal protein with premature stop codon A B 3.2 Diseases associated with triplet and other repeat expansions* Repeat No. of repeats Gene Gene location Inheritance Normal Mutant Coding repeat expansion Huntington's disease [CAG] 6-34

35 Huntingtin 4p16 AD Spinocerebellar ataxia (type 1) [CAG] 6-39 40 Ataxin 6p22-23 AD Spinocerebellar ataxia (types 2, 3, 6, 7) [CAG] Various Various Various Various AD Dentatorubral-pallidoluysian atrophy [CAG] 7-25 49 Atrophin 12p12-13 AD Machado-Joseph disease [CAG] 12-40 67 MJD 14q32 AD Spinobulbar muscular atrophy [CAG] 11-34 40 Androgen receptor Xq11-12 XL recessive Non-coding repeat expansion Myotonic dystrophy [CTG] 5-37 50 DMPK-3'UTR 19q13 AD Friedreich's ataxia [GAA] 7-22 200 Frataxin-intronic 9q13 AR Progressive myoclonic epilepsy [CCCCGCCCGCG]4-8 2-3 25 Cystatin B-5'UTR 21q AR Fragile X mental retardation [CGG] 5-52 200 FMR1-5'UTR Xq27 XL dominant Fragile site mental retardation 2 (FRAXE) [GCC] 6-35 200 FMR2 Xq28 XL, probably recessive *The triplet repeat diseases fall into two major groups: those with disease stemming from expansion of [CAG]_n repeats in coding DNA, resulting in multiple adjacent glutamine residues (polyglutamine tracts), and those with non-coding repeats. The latter tend to be longer. Unaffected parents usually display 'pre-mutation' allele lengths that are just above the normal range. (AD/AR = autosomal dominant/recessive; UTR = untranslated region; XL = X-linked)

44 • CLINICAL GENETICS Copy number variations Variation in the number of copies of an individual segment of the genome from the usual diploid (two copies) content can be categorised by the size of the segment involved. Rarely, individuals may gain (trisomy) or lose (monosomy) a whole chromosome. Such numerical chromosome anomalies most commonly occur by a process known as non-disjunction, where pairs of homologous chromosomes do not separate at meiosis II (p. 40). Common trisomies include Down's syndrome (trisomy 21), Edward's syndrome (trisomy 18) and Patau's syndrome (trisomy 13). Monosomy of the autosomes (present in all the cells, as opposed to in a mosaic distribution) does not occur but Turner's syndrome, in which there is monosomy for the X chromosome, affects approximately 1 in 2500 live births (Box 3.3). Large insertions or deletions of chromosomal DNA also occur and are usually associated with a learning disability and/or congenital malformations. Such structural chromosomal anomalies usually arise as the result of one of two different processes: • non-homologous end-joining • non-allelic homologous recombination. Random double-stranded breaks in DNA are a necessary process in meiotic recombination and also occur during mitosis at a predictable rate. The rate of these breaks is dramatically increased by exposure to ionising radiation. When such breaks take place, they are usually repaired accurately by DNA repair mechanisms within the cell. However, in a proportion of breaks, segments of DNA that are not normally contiguous will be joined ('non-homologous end-joining'). If the joined fragments are from different chromosomes, this results in a translocation. If they are 3.3 Chromosome and contiguous gene disorders Disease Locus Incidence Clinical features Numerical chromosomal abnormalities Down's syndrome (trisomy 21) 47,XY,+21 or 47,XX,+21 1 in 800 Characteristic facies, IQ usually < 50, congenital heart disease, reduced life expectancy Edwards' syndrome (trisomy 18) 47,XY,+18 or 47,XX,+18 1 in 6000 Early lethality, characteristic skull and facies, frequent malformations of heart, kidney and other organs Patau's syndrome (trisomy 13) 47,XY,+13 or 47, XX,+13 1 in 15 000 Early lethality, cleft lip and palate, polydactyly, small head, frequent congenital heart disease Klinefelter's syndrome 47,XXY 1 in 1000 Phenotypic male, infertility, gynaecomastia, small testes (p. 660) XYY 47,XYY 1 in 1000 Usually asymptomatic, some impulse control problems Triple X syndrome 47,XXX 1 in 1000 Usually asymptomatic, may have reduced IQ Turner's syndrome 45,X 1 in 5000 Phenotypic female, short stature, webbed neck, coarctation of the aorta, primary amenorrhoea (p. 659) Recurrent deletions, microdeletions and contiguous gene defects Di George/velocardiofacial syndrome 22q11.2 1 in 4000 Cardiac outflow tract defects, distinctive facial appearance, thymic hypoplasia, cleft palate and hypocalcaemia. Major gene seems to be TBX1 (cardiac defects and cleft palate) Prader-Willi syndrome 15q11-q13 1 in 15 000 Distinctive facial appearance, hyperphagia, small hands and feet, distinct behavioural phenotype. Imprinted region, deletions on paternal allele in 70% of cases Angelman's syndrome 15q11-q13 1 in 15 000 Distinctive facial appearance, absent speech, electroencephalogram (EEG) abnormality, characteristic gait. Imprinted region, deletions on maternal allele encompassing UBE3A Williams' syndrome 7q11.23 1 in 10 000 Distinctive facial appearance, supravalvular aortic stenosis, learning disability and infantile hypercalcaemia. Major gene for supravalvular aortic stenosis is elastin Smith-Magenis syndrome 17p11.2 1 in 25 000 Distinctive facial appearance and behavioural phenotype, self-injury and rapid eye movement (REM) sleep abnormalities. Major gene seems to be RAI1 from the same chromosome, this will result in either inversion, duplication or deletion of a chromosomal fragment (Fig. 3.7). Large insertions and deletions may be cytogenetically visible as chromosomal deletions or duplications. If the anomalies are too small to be detected by microscopy, they are termed microdeletions and microduplications. Many microdeletion syndromes have been described and most result from nonallelic homologous recombination between repeats of highly similar DNA sequences, which leads to recurrent

chromosome anomalies – and clinical syndromes – occurring in unrelated individuals (Fig. 3.7 and Box 3.3). Consequences of genomic variation The consequence of an individual mutation depends on many factors, including the mutation type, the nature of the gene product and the position of the variant in the protein. Mutations can have profound or subtle effects on gene and cell function. Variations that have profound effects are responsible for ‘classical’ genetic diseases, whereas those with subtle effects may contribute to the pathogenesis of common disease where there is a genetic component, such as diabetes.

- Neutral variants have no effect on quality or type of protein produced.
- Loss-of-function mutations result in loss or reduction in the normal protein function. Whole-gene deletions are the archetypal loss-of-function variants but stop-gain or indel mutations (early in the ORF), missense mutations affecting a critical domain and splice-site mutations can also result in loss of protein function.

Genomics, health and disease • 45

as a common polymorphism. However, the most frequent is the single nucleotide polymorphism, or SNP (pronounced ‘snip’), describing the substitution of a single base. Polymorphisms and common disease The protective and detrimental polymorphisms associated with common disease have been identified primarily through genome-wide association studies (GWAS, p. 56) and are the basis for many direct-to-consumer tests that purport to determine individual risk profiles for common diseases or traits such as cardiovascular disease, diabetes and even male-pattern baldness! An example is the polymorphism in the gene SLC2A9 that not only explains a significant proportion of the normal population variation in serum urate concentration but also predisposes ‘high-risk’ allele carriers to the development of gout. However, the current reality is that, until we have a more comprehensive understanding of the full genomic landscape and knowledge of the complete set of detrimental and protective polymorphisms, we cannot accurately assess risk.

Evolutionary selection Genetic variants play an important role in evolutionary selection, with advantageous variants resulting in positive selection via improved reproductive fitness, and variations that decrease

- Gain-of-function mutations result in a gain of protein function. They are typically non-synonymous mutations that alter the protein structure, leading to activation/ alteration of its normal function through causing either an interaction with a novel substrate or a change in its normal function.
- Dominant negative mutations are the result of nonsynonymous mutations or in-frame deletions/duplications but may also, less frequently, be caused by triplet repeat expansion mutations. Dominant negative mutations are heterozygous changes that result in the production of an abnormal protein that interferes with the normal functioning of the wild-type protein.

Normal genomic variation We each have 5–50 million variants in our genome, occurring approximately every 300 bases. These variants are mostly polymorphisms, arising in more than 1% of the population; they have no or subtle effects on gene and cell function, and are not associated with a high risk of disease. Polymorphisms can occur within exons, introns or the intergenic regions that comprise 98–99% of the human genome. Each of the classes of genetic variant discussed on page 42 is present in the genome

Fig. 3.7 Chromosomal analysis and structural chromosomal disorders. A Human chromosomes can be classed as metacentric if the centromere is near the middle, or acrocentric if the centromere is at the end. The bands of each chromosome are given a number, starting at the centromere and working out along the short (p) arm and long (q) arm.

Translocations and inversions are balanced structural chromosome anomalies where no genetic material is missing but it is in the wrong order. Translocations can be divided into reciprocal (direct swap of chromosomal material between nonhomologous chromosomes) and Robertsonian (fusion

of acrocentric chromosomes). Deletions and duplications can also occur due to non-allelic homologous recombination (illustrated in part B). Deletions are classified as interstitial if they lie within a chromosome, and terminal if the terminal region of the chromosome is affected. Duplications can be either in tandem (where the duplicated fragment is inserted next to the region that is duplicated and orientated in the correct direction) or inverted (where the duplicated fragment is in the wrong direction). (N = normal; A = abnormal) B A common error of meiotic recombination, known as non-allelic homologous recombination, can occur (right panel), resulting in a deletion on one chromosome and a duplication in the homologous chromosome. The error is induced by tandem repeats in the DNA sequences (green), which can misalign and bind to each other, thereby 'fooling' the DNA into thinking the pairing prior to recombination is correct.

Chromosome
Cen p

21.1 21.3 21.2

34.2 q Cen p N N A N A N A N A N A N A A 13 (sa) (st) 11.2

24.2

32.2 q Metacentric Acrocentric Chromosome

Reciprocal translocation Robertsonian translocation Inversions Interstitial Terminal Tandem Inverted Mechanism underlying recurrent deletions and duplication: non-allelic homologous recombination How structural chromosomal anomalies are described Deletions Duplications B A Recombination Normal pairing Abnormal pairing between DNA repeats Deletion Duplication

DNA repeat Maternal chromosome Paternal chromosome

46 • CLINICAL GENETICS history, on both sides of the family, enquiring about details of all medical conditions in family members, consanguinity, dates of birth and death, and any history of pregnancy loss or infant death. The basic symbols and nomenclature used in drawing a pedigree are shown in Figure 3.8. Patterns of disease inheritance Autosomal dominant inheritance Take some time to draw out the following pedigree: Anne is referred to Clinical Genetics to discuss her personal history of colon cancer (she was diagnosed at the age of 46 years) and family history of colon/endometrial cancer: her mother was diagnosed with endometrial cancer at the age of 60 years and her cousin through her healthy maternal aunt was diagnosed with colon cancer in her fifties. Both her maternal grandmother and grandfather died of 'old age'. There is no family history of note on her father's side of the family. He has one brother and both his parents died of old age, in their eighties. Anne has two healthy daughters, aged 12 and 14 years, and a healthy full sister. This family history is typical of an autosomal dominant condition (Fig. 3.8): in this case, a colon/endometrial cancer susceptibility syndrome known as Lynch's syndrome, associated with disruption of one of the mismatch repair genes: MSH2, MSH6, MLH1 and PMS2 (see p. 830 and Box 3.11, p. 57). reproductive fitness becoming excluded through evolution. Given this simple paradigm, it would be tempting to assume that common mutations are all advantageous and all rare mutations are pathogenic. Unfortunately, it is often difficult to classify any common mutation as either advantageous or deleterious - or, indeed, neutral. Mutations that are advantageous in

early life and thus enhance reproductive fitness may be deleterious in later life. There may be mutations that are advantageous for survival in particular conditions (e.g. famine or pandemic) that may be disadvantageous in more benign circumstances by causing a predisposition to obesity or autoimmune disorders. Constitutional genetic disease Familial genetic disease is caused by constitutional mutations, which are inherited through the germ line. However, different mutations in the same gene can have different consequences, depending on the genetic mechanism underlying that disease. About 1% of the human population carries constitutional mutations that cause disease. Constructing a family tree The family tree – or pedigree – is a fundamental tool of the clinical geneticist, who will routinely take a three-generation family Fig. 3.8 Drawing a pedigree and patterns of inheritance. A The main symbols used to represent pedigrees in diagrammatic form. B The main modes of disease inheritance (see text for details). SB Male Clinically affected Deceased individual (with age at death) Separated Consanguinity Clinically affected, several diagnoses Carrier Positive presymptomatic test Monozygotic twins Dizygotic twins Stillbirth (with gestation) Termination Miscarriage (with gestation) Unknown sex Female Partners Recessive inheritance Dominant inheritance Mitochondrial DNA disorder X-linked recessive inheritance Transmission to 50% of offspring independent of gender Consanguinity Affected males related through unaffected females Both sexes affected but only inherited through female meiosis I II III IV I II III IV I II III IV V I II III IV

d. 50 y 30 wk SB 39 wk 16 wk A B

Genomics, health and disease • 47

- Males and females are usually affected in roughly equal numbers (unless the clinical presentation of the condition is gender-specific, such as an inherited susceptibility to breast and/or ovarian cancer). The offspring risk for an individual affected with an autosomal dominant condition is 1 in 2 (or 50%). This offspring risk is true for each pregnancy, since half the affected individual gametes (sperm or egg cells) will contain the affected chromosome/gene and half will contain the normal chromosome/gene. There is a long list of autosomal dominant conditions, some of which are shown in Box 3.4. Features of an autosomal dominant pedigree include:
 - There are affected individuals in each generation (unless the mutation has arisen de novo, i.e. for the first time in an affected individual). However, variable penetrance and expressivity can influence the number of affected individuals and the severity of disease in each generation. Penetrance is defined as the proportion of individuals bearing a mutated allele who develop the disease phenotype. The mutation is said to be fully penetrant if all individuals who inherit a mutation develop the disease. Expressivity describes the level of severity of each aspect of the disease phenotype.

3.4 Genetic conditions dealt with by clinicians in other specialties

Name of condition	Gene	Reference
Autosomal dominant conditions		
Autosomal dominant polycystic kidney disease (ADPKD)	PKD1 (85%), PKD2 (15%)	p. 405 Box 15.28, p. 415
Tuberous sclerosis	TSC1 TSC2	p. 1264 p. 1264
Marfan's syndrome	FBN1	p. 508
Long QT syndrome	KCNQ1	p. 476
Brugada's syndrome	SCN5A	p. 477
Neurofibromatosis type 1	NF1	p. 1131 Box 25.77, p. 1132
Neurofibromatosis type 2	NF2	p. 1131 Box 25.77, p. 1132
Hereditary spherocytosis	ANK1	p. 947
Vascular Ehlers–Danlos syndrome (EDS type 4)	COL3A1	p. 970
Hereditary haemorrhagic telangiectasia	ENG, ALK1, GDF2	p. 970
Osteogenesis imperfecta	COL1A1, COL1A2	p. 1055
Charcot–Marie–Tooth disease	PMP22, MPZ, GJB1	p. 1140
Hereditary neuropathy with liability to pressure palsies	PMP22	
Autosomal recessive conditions		
Familial Mediterranean fever	MEFV	p. 81
Mevalonic aciduria (mevalonate kinase deficiency)	MVK	p. 81

recessive polycystic kidney disease (ARPKD) PKHD1 Box 15.28, p. 415 Kartagener's syndrome (primary ciliary dyskinesia) DNAI1 Box 17.30, p. 578 Cystic fibrosis CFTR1 p. 580 Box 17.30, p. 578 p. 842 Pendred's syndrome SLC26A4 p. 650 Congenital adrenal hyperplasia-21 hydroxylase deficiency CYP21A p. 676 Box 18.27, p. 658 Haemochromatosis HFE p. 895 Wilson's disease ATP7B p. 896 Alpha1-antitrypsin deficiency SERPINA1 p. 897 Gilbert's syndrome UGT1A1 p. 897 Benign recurrent intrahepatic cholestasis ATP8B1 p. 902 Alpha-thalassaemia HBA1, HBA2 p. 951 p. 954 Beta-thalassaemia HBB p. 951 p. 953 Sickle cell disease HBB p. 951 Spinal muscular atrophy SMN1 p. 1117 X-linked conditions Alport's syndrome COL4A5 Box 15.28, p. 415 p. 403 Primary agammaglobulinaemia BTK p. 78 Haemophilia A (factor VIII deficiency) F8 p. 971 Haemophilia B (factor IX deficiency) F9 p. 973 Duchenne muscular dystrophy DMD p. 1143 and Box 25.91

48 • CLINICAL GENETICS Autosomal recessive inheritance As above, take some time to draw a pedigree representing the following: Mr and Mrs Kent, a non-consanguineous couple, are referred because their son, Jamie, had severe neonatal liver disease. Included among the many investigations that the paediatric hepatologist undertook was testing for α 1-antitrypsin deficiency (Box 3.5). Jamie was shown to have the PiZZ phenotype. Testing confirmed both parents as carriers with PiMZ phenotypes. In the family, Jamie has an older sister who has no medical problems. Mr Kent is one of four children with two brothers and a sister and Mrs Kent has a younger brother. Both sets of grandparents are alive and well. There is no family history of α 1-antitrypsin deficiency. This family history is characteristic of an autosomal recessive disorder (Fig. 3.8), where both alleles of a gene must be mutated before the disease is manifest in an individual; an affected individual inherits one mutant allele from each of their parents, who are therefore healthy carriers for the condition. An autosomal recessive condition might be suspected in a family where:

- Males and females are affected in roughly equal proportions.
- Parents are blood related; this is known as consanguinity. Where there is consanguinity, the mutations are usually homozygous, i.e. the same mutant allele is inherited from both parents.
- Individuals within one sibship in one generation are affected and so the condition can appear to have arisen 'out of the blue'. Approximately 1 in 4 children born to carriers of an autosomal recessive condition will be affected. The offspring risk for carrier parents is therefore 25% and the chances of an unaffected child, with an affected sibling, being a carrier is 2/3. Examples of some autosomal recessive conditions, discussed elsewhere in this book, are shown in Box 3.4.

X-linked inheritance The following is an exemplar of an X-linked recessive pedigree (Fig. 3.8): Edward has a diagnosis of Duchenne muscular dystrophy (DMD, Box 3.6). His parents had suspected the diagnosis when he was 3 years old because he was not yet walking and there was a family history of DMD: Edward's maternal uncle had been affected and died at the age of 24 years. Edward's mother has no additional siblings. After Edward demonstrated a very high creatinine kinase level, the paediatrician also requested genetic testing, which identified a deletion of exons 2-8 of the dystrophin gene. Edward has a younger, healthy sister and grandparents on both sides of the family are well, although the maternal grandmother has recently developed a cardiomyopathy. Edward's father has an older sister and an older brother who are both well. Genetic diseases caused by mutations on the X chromosome have specific characteristics:

- X-linked diseases are mostly recessive and restricted to males who carry the mutant allele. This is because males

3.5 Alpha1-antitrypsin deficiency Inheritance pattern • Autosomal recessive Genetic cause • Two common mutations in the SERPINA1 gene: p.Glu342Lys and p.Glu264Val Prevalence • 1 in 1500-3000 of European ancestry Clinical presentation • Variable presentation from neonatal period through to adulthood • Neonatal period: prolonged jaundice with conjugated hyperbilirubinaemia or (rarely) liver disease • Adulthood: pulmonary emphysema

and/or cirrhosis. Rarely, the skin disease, panniculitis, develops

Disease mechanism • SERPINA1 encodes α 1-antitrypsin, which protects the body from the effects of neutrophil elastase. The symptoms of α 1-antitrypsin deficiency result from the effects of this enzyme attacking normal tissue

Disease variants • M variant: if an individual has normal SERPINA1 genes and produces normal levels of α 1-antitrypsin, they are said to have an M variant • S variant: p.Glu264Val mutation results in α 1-antitrypsin levels reduced to about 40% of normal • Z variant: p.Glu342Lys mutation results in very little α 1-antitrypsin • PiZZ: individuals who are homozygous for the p.Glu342Lys mutation are likely to have α 1-antitrypsin deficiency and the associated symptoms • PiZS: individuals who are compound heterozygous for p.Glu342Lys and p.Glu264Val are likely to be affected, especially if they smoke, but usually to a milder degree

3.6 Duchenne muscular dystrophy*

Inheritance pattern • X-linked recessive Genetic cause • Mutations or deletions encompassing/within the DMD (dystrophin) gene located at Xp21 Prevalence • 1 in 3000–4000 live male births Clinical presentation • Delayed motor milestones • Speech delay • Grossly elevated creatine kinase (CK) levels (in the thousands) • Ambulation is usually lost between the ages of 7 and 13 years • Lifespan is reduced with a mean age of death, usually from respiratory failure, in the mid-twenties • Cardiomyopathy affects almost all boys with Duchenne muscular dystrophy and some female carriers Disease mechanism • DMD encodes dystrophin, a major structural component of muscle • Dystrophin links the internal cytoskeleton to the extracellular matrix Disease variants • Becker muscular dystrophy, although a separate disease, is also caused by mutations in the dystrophin gene • In Duchenne muscular dystrophy, there is no dystrophin protein, whereas in Becker muscular dystrophy there is a reduction in the amount or alteration in the size of the dystrophin protein *See also page 1143.

Genomics, health and disease • 49

dinucleotide (NADH) and the reduced form of flavine adenine dinucleotide (FADH₂). Both NADH and FADH₂ then donate electrons to the respiratory chain. Here these electrons are transferred via a complex series of reactions, resulting in the formation of a proton gradient across the inner mitochondrial membrane. The gradient is used by an inner mitochondrial membrane protein, ATP synthase, to produce ATP, which is then transported to other parts of the cell. Dephosphorylation of ATP is used to produce the energy required for many cellular processes. Each mitochondrion contains 2–10 copies of a 16-kilobase (kB) double-stranded circular DNA molecule (mtDNA). This mtDNA contains 13 protein-coding genes, all involved in the respiratory chain, and the ncRNA genes required for protein synthesis within the mitochondria (Fig. 3.9). The mutational rate of mtDNA is relatively high due to the lack of protection by chromatin. Several mtDNA diseases characterised by defects in ATP production have been described. Mitochondria are most numerous in cells with high metabolic demands, such as muscle, retina and the basal ganglia, and these tissues tend to be the ones most severely affected in mitochondrial diseases (Box 3.7). There are many other mitochondrial diseases that are caused by mutations in nuclear genes, which encode proteins that are then imported into the mitochondrion and are critical for energy production, e.g. most forms of Leigh's syndrome (although Leigh's syndrome may also be caused by a mitochondrial gene mutation). The inheritance of mtDNA disorders is characterised by transmission from females, but males and females generally are equally affected (see Fig. 3.8). Unlike the other inheritance patterns mentioned above, mitochondrial inheritance has nothing to do with meiosis but reflects the fact that mitochondrial DNA is transmitted by oocytes: sperm do not contribute mitochondria to the zygote. Mitochondrial disorders tend to be variable in penetrance and

expressivity within families, and this is mostly accounted for by the fact that only a proportion of multiple mtDNA molecules within mitochondria contain the causal mutation (the degree of mtDNA heteroplasmy). Imprinting Several chromosomal regions (loci) have been identified where gene expression is inherited in a parent-of-origin-specific manner; have only one X chromosome, whereas females have two (see Fig. 3.1). However, occasionally, female carriers may exhibit signs of an X-linked disease due to a phenomenon called skewed X-inactivation. All female embryos, at about 100 cells in size, stably inactivate one of their two X chromosomes in each cell. Where this inactivation is random, approximately 50% of the cells will express the genes from one X chromosome and 50% of cells will express genes from the other. Where there is a mutant gene, there is often skewing away from the associated X chromosome, resulting in an unaffected female carrier. However, if, by chance, there is a disproportionate inactivation of the normal X chromosome with skewing towards the mutant allele, then an affected female carrier may be affected (albeit more mildly than males).

- The gene can be transmitted from female carriers to their sons: in families with an X-linked recessive condition, there are often a number of affected males related through unaffected females.
- Affected males cannot transmit the condition to their sons (but all their daughters would be carriers). The risk of a female carrier having an affected child is 25% or half of her male offspring.

Mitochondrial inheritance The mitochondrion is the main site of energy production within the cell. Mitochondria arose during evolution via the symbiotic association with an intracellular bacterium. They have a distinctive structure with functionally distinct inner and outer membranes. Mitochondria produce energy in the form of adenosine triphosphate (ATP). ATP is mostly derived from the metabolism of glucose and fat (Fig. 3.9). Glucose cannot enter mitochondria directly but is first metabolised to pyruvate via glycolysis. Pyruvate is then imported into the mitochondrion and metabolised to acetyl-co-enzyme A (acetyl-CoA). Fatty acids are transported into the mitochondria following conjugation with carnitine and are sequentially catabolised by a process called β -oxidation to produce acetyl-CoA. The acetyl-CoA from both pyruvate and fatty acid oxidation is used in the citric acid (Krebs) cycle – a series of enzymatic reactions that produces CO₂, the reduced form of nicotinamide adenine 3.7

The structure of the respiratory chain complexes and the diseases associated with their dysfunction

Complex	Enzyme	nDNA subunits	mtDNA subunits	Diseases
I	NADH dehydrogenase	1	2	MELAS, MERRF, bilateral striatal necrosis, LHON, myopathy and exercise intolerance, Parkinsonism, Leigh's syndrome, exercise myoglobinuria, leucodystrophy/myoclonic epilepsy
II	Succinate dehydrogenase	1	0	Phaeochromocytoma, Leigh's syndrome
III	Cytochrome bc ₁ complex	1	0	Leigh's syndrome
IV	Cytochrome c oxidase	1	2	Parkinsonism/MELAS, cardiomyopathy, myopathy, exercise myoglobinuria, Leigh's syndrome
V	ATP synthase	1	2	Sideroblastic anaemia, myoclonic ataxia, deafness, myopathy, MELAS, MERRF, mitochondrial encephalomyopathy, motor neuron disease-like, exercise myoglobinuria, Leigh's syndrome

MELAS, MERRF, bilateral striatal necrosis, LHON, myopathy and exercise intolerance, Parkinsonism, Leigh's syndrome, exercise myoglobinuria, leucodystrophy/myoclonic epilepsy

Phaeochromocytoma, Leigh's syndrome

Parkinsonism/MELAS, cardiomyopathy, myopathy, exercise myoglobinuria, Leigh's syndrome

Sideroblastic anaemia, myoclonic ataxia, deafness, myopathy, MELAS, MERRF, mitochondrial encephalomyopathy, motor neuron disease-like, exercise myoglobinuria, Leigh's syndrome

Leigh's syndrome, NARP, bilateral striatal necrosis

1nDNA subunits. 2mtDNA subunits = number of different protein subunits in each complex that are encoded in the nDNA and mtDNA, respectively. (ATP = adenosine triphosphate; LHON = Leber hereditary optic neuropathy; MELAS = myopathy,

encephalopathy, lactic acidosis and stroke-like episodes; MERRF = myoclonic epilepsy and ragged red fibres; mtDNA = mitochondrial DNA; NADH = the reduced form of nicotinamide adenine dinucleotide; NARP = neuropathy, ataxia and retinitis pigmentosa; nDNA = nuclear DNA)

50 • CLINICAL GENETICS Somatic genetic disease Somatic mutations are not inherited but instead occur during post-zygotic mitotic cell divisions at any point from embryonic development to late adult life. An example of this phenomenon is polyostotic fibrous dysplasia (McCune-Albright syndrome), in Fig. 3.9 Mitochondria. A Mitochondrial structure. There is a smooth outer membrane surrounding a convoluted inner membrane, which has inward projections called cristae. The membranes create two compartments: the inter-membrane compartment, which plays a crucial role in the electron transport chain, and the inner compartment (or matrix), which contains mitochondrial DNA and the enzymes responsible for the citric acid (Krebs) cycle and the fatty acid β -oxidation cycle. B Mitochondrial DNA. The mitochondrion contains several copies of a circular double-stranded DNA molecule, which has a non-coding region, and a coding region that encodes the genes responsible for energy production, mitochondrial transfer RNA (tRNA) molecules and mitochondrial ribosomal RNA (rRNA) molecules. (ATP = adenosine triphosphate; NADH = the reduced form of nicotinamide adenine dinucleotide) C Mitochondrial energy production. Fatty acids enter the mitochondrion conjugated to carnitine by carnitine-palmitoyl transferase type 1 (CPT I) and, once inside the matrix, are unconjugated by CPT II to release free fatty acids (FFA). These are broken down by the β -oxidation cycle to produce acetyl-co-enzyme A (acetyl-CoA). Pyruvate can enter the mitochondrion directly and is metabolised by pyruvate dehydrogenase (PDH) to produce acetyl-CoA. The acetyl-CoA enters the Krebs cycle, leading to the production of NADH and flavine adenine dinucleotide (reduced form) (FADH₂), which are used by proteins in the electron transport chain to generate a hydrogen ion gradient across the inter-membrane compartment. Reduction of NADH and FADH₂ by proteins I and II, respectively, releases electrons (e), and the energy released is used to pump protons into the inter-membrane compartment. Coenzyme Q10/ubiquinone (Q) is an intensely hydrophobic electron carrier that is mobile within the inner membrane. As electrons are exchanged between proteins in the chain, more protons are pumped across the membrane, until the electrons reach complex IV (cytochrome oxidase), which uses the energy to reduce oxygen to water. The hydrogen ion gradient is used to produce ATP by the enzyme ATP synthase, which consists of a proton channel and catalytic sites for the synthesis of ATP from ADP. When the channel opens, hydrogen ions enter the matrix down the concentration gradient, and energy is released that is used to make ATP. L

s t r a n d H

s t r a n d Outer membrane Inner membrane NADH NAD I II III Q Cyt C IV NADH FADH₂ Fatty acid β -oxidation cycle Citric acid (Krebs) cycle H⁺ e 2e FADH₂ FADH₂ Lactate Pyruvate PDH Acetyl-CoA Glucose 22 tRNAs NADH dehydrogenase 7 subunits Cytochrome B/C oxidase 4 subunits 2 ribosomal RNA subunits 2 ATP synthase subunits Intragenic DNA Inner membrane Cristae Matrix Outer membrane FFA CPT I CPT II Carnitine Carnitine-FA ester C A B FFA FAD 2H⁺ H₂O O₂ ATP ADP

- Pi H⁺ H⁺ ATP synthase Carnitine e e e these are called imprinted loci. Within these loci the paternally inherited gene may be active while the maternally inherited may be silenced, or vice versa. Mutations within imprinted loci lead to an unusual pattern of inheritance where the phenotype is manifest only if inherited from the parent who

contributes the transcriptionally active allele. Examples of imprinting disorders are given in Box 3.8.

Interrogating the genome: the changing landscape of genomic technologies • 51

light or cigarette smoke, or if the cell has defects in DNA repair systems. Cancer is thus a disease that affects the fundamental processes of molecular and cell biology.

Interrogating the genome: the changing landscape of genomic technologies

Looking at chromosomes

The analysis of metaphase chromosomes by light microscopy was the mainstay of clinical cytogenetic analysis for decades, the aim being to detect gain or loss of whole chromosomes (aneuploidy) or large chromosomal segments (> 4 million bp). More recently, genome-wide microarrays (array comparative genomic hybridisation or array CGH) have replaced chromosome analysis, allowing rapid and precise detection of segmental gain or loss of DNA throughout the genome (see Box 3.3). Microarrays consist of grids of multiple wells containing short DNA sequences (reference DNA) that are complementary to known sequences in the genome. Patient and reference DNA are each labelled with a coloured fluorescent dye (generally, patient DNA is labelled with a green fluorescent dye and reference DNA with a red fluorescent dye) and added to the microarray grid. Where there is an equal quantity of patient and reference DNA bound to the spot, this results in yellow fluorescence. Where there is too much patient DNA (representing a duplication of a chromosome region), the spot will be greener; it will be more red (appears orange) where there is 2 : 1 ratio of the control:patient DNA (representing heterozygous deletion of a chromosome region; Fig. 3.10). Array CGH and other array-based approaches can detect small chromosomal deletions and duplications. They are also generally more sensitive than conventional karyotyping at detecting mosaicism (where there are two or more populations of cells, derived from a single fertilised egg, with different genotypes).

which a somatic mutation in the *GS alpha* gene causes constitutive activation of downstream signalling, resulting in focal lesions in the skeleton and endocrine dysfunction (p. 1055). The most important example of human disease caused by somatic mutations is cancer (see Ch. 33). Here, 'driver' mutations occur within genes that are involved in regulating cell division or apoptosis, resulting in abnormal cell growth and tumour formation. The two general categories of cancer-causing mutation are gain-of-function mutations in growth-promoting genes (oncogenes) and loss-of-function mutations in growth-suppressing genes (tumour suppressor genes). Whichever mechanism is acting, most tumours require an initiating mutation in a single cell that can then escape from normal growth controls. This cell replicates more frequently or fails to undergo programmed death, resulting in clonal expansion. As the size of the clone increases, one or more cells may acquire additional mutations that confer a further growth advantage, leading to proliferation of these subclones, which may ultimately result in aggressive metastatic cancer. The cell's complex self-regulating machinery means that more than one mutation is usually required to produce a malignant tumour (see Fig. 33.3, p. 1318). For example, if a mutation results in activation of a growth factor gene or receptor, then that cell will replicate more frequently as a result of autocrine stimulation. However, this mutant cell will still be subject to normal cell-cycle checkpoints to promote DNA integrity in its progeny. If additional mutations in the same cell result in defective cell-cycle checkpoints, however, it will rapidly accumulate further mutations, which may allow completely unregulated growth and/ or separation from its matrix and cellular attachments and/or resistance to apoptosis. As cell growth becomes increasingly dysregulated, cells de-differentiate, lose their response to normal tissue environment and cease to ensure appropriate mitotic chromosomal segregation. These processes combine to generate the

classical malignant characteristics of disorganised growth, variable levels of differentiation, and numerical and structural chromosome abnormalities. An increase in somatic mutation rate can occur on exposure to external mutagens, such as ultraviolet

3.8 Imprinting disorders Disorder Locus Genes Notes Beckwith–Wiedemann syndrome 11p15 CDKN1C, IGF2, H19 Increased growth, macroglossia, hemihypertrophy, abdominal wall defects, ear lobe pits/creases and increased susceptibility to developing childhood tumours Prader–Willi syndrome 15q11–q13 SNRPN, Necdin and others Obesity, hypogonadism and learning disability. Lack of paternal contribution (due to deletion of paternal 15q11–q13, or inheritance of both chromosome 15q11–q13 regions from the mother) Angelman’s syndrome (AS) 15q11–q13 UBE3A Severe mental retardation, ataxia, epilepsy and inappropriate laughing bouts. Due to loss-of-function mutations in the maternal UBE3A gene. The neurological phenotype results because most tissues express both maternal and paternal alleles of UBE3A, whereas the brain expresses predominantly the maternal allele

Pseudohypoparathyroidism (p. 664) 20q13 GNAS1 Inheritance of the mutation from the mother results in hypocalcaemia, hyperphosphataemia, raised parathyroid hormone (PTH) levels, ectopic calcification, obesity, delayed puberty and shortened 4th and 5th metacarpals (the syndrome known as Albright’s hereditary osteodystrophy, AHO). When the mutation is inherited from the father, PTH, calcium and phosphate levels are normal but the other features are present (pseudopseudohypoparathyroidism, p. 664). These differences are due to the fact that, in the kidney (the main target organ through which PTH regulates serum calcium and phosphate), the paternal allele is silenced and the maternal allele is expressed, whereas both alleles are expressed in other tissues.

52 • CLINICAL GENETICS cycle of heating/cooling and denaturation/replication is repeated many times, resulting in the exponential amplification of DNA between primer sites (Fig. 3.11). Gene sequencing In the mid-1970s, a scientist called Fred Sanger pioneered a DNA sequencing technique (‘Sanger sequencing’) that determined the precise order and nucleotide type (thymine, cytosine, adenine and guanine) in a molecule of DNA. Modern Sanger sequencing uses fluorescently labelled, chain-terminating nucleotides that are sequentially incorporated into the newly synthesised DNA, generating multiple DNA chains of differing lengths. These DNA chains are subject to capillary electrophoresis, which separates them by size, allowing the fragments to be ‘read’ by a laser and producing a sequence chromatogram that corresponds to the target sequence (Fig. 3.12). Although transformative, Sanger sequencing was difficult and costly to scale, as exemplified by the Human Genome Project, which took 12 years to sequence the entire human genome at a cost approaching 3 billion dollars. Recently, DNA sequencing has been transformed again by a group of technologies collectively known as ‘next-generation sequencing’ (NGS; Fig. 3.13). This refers to a family of postSanger sequencing technologies that utilise the same five basic principles:

- Library preparation: DNA samples are fragmented (by enzyme cleavage or ultrasound) and then modified with a custom adapter sequence.
- Amplification: the library fragment is amplified to produce DNA clusters, each originating from a single DNA fragment. Each cluster will act as a single sequencing reaction.
- Capture: if an entire genome is being sequenced, this step will not be included. The capture step is required if targeted resequencing is necessary, such as for a panel gene test or an exome (Box 3.9).
- Sequencing: each DNA cluster is simultaneously sequenced and the data from each captured; this is known as a ‘read’ and is usually between 50 and 300 bases long sequenced (see Box 3.10 for a detailed description of the three most commonly used sequencing methods: synthesis, ligation and ion semiconductor sequencing).
- Alignment and variant identification: specialised software analyses read sequences and compares

the data to a reference template. This is known as 'alignment' or 'mapping' and, although there are 3 billion bases in the human genome, array-based approaches will not detect balanced chromosome rearrangements where there is no loss or gain of genes/chromosome material, such as balanced reciprocal translocations, or a global increase in copy number, such as triploidy. The widespread use of array-based approaches has brought a number of challenges for clinical interpretation, including the identification of copy number variants (CNVs) of uncertain clinical significance, CNVs of variable penetrance and incidental findings. A CNV of uncertain clinical significance describes a loss or gain of chromosome material where there are insufficient data to conclude whether or not it is associated with a learning disability and/or medical problems. While this uncertainty can be difficult to prepare families for and can be associated with considerable anxiety, it is likely that there will be greater clarity in the future as we generate larger CNV datasets. A CNV of variable penetrance, also known as a neurosusceptibility locus, describes a chromosome deletion or duplication associated with a lower threshold for manifesting a learning disability or autistic spectrum disorder. CNVs of variable penetrance are therefore identified at greater frequencies among individuals with a learning disability and/or autistic spectrum disorder than in the general population. The current understanding is that additional modifying factors (genetic, environmental or stochastic) must influence the phenotypic expression of these neurosusceptibility loci. Finally, an incidental CNV finding describes a deletion or duplication encompassing a gene or genes that are causative of a phenotype or risk unrelated to the presenting complaint. For instance, if, through the array CGH investigation for an intellectual disability, a deletion encompassing the BRCA1 gene were identified, this would be considered an incidental finding.

Looking at genes

Gene amplification: polymerase chain reaction

The polymerase chain reaction (PCR) is a fundamental laboratory technique that amplifies targeted sections of the human genome for further analyses – most commonly, DNA sequencing. The method utilises thermal cycling: repeated cycles of heating and cooling allow the initial separation of double-stranded DNA into two single strands (known as denaturation), each of which serves as a template during the subsequent replication step, guided by primers designed to anneal to a specific genomic region.

This Fig. 3.10 Detection of chromosome abnormalities by comparative genomic hybridisation (CGH). Deletions and duplications are detected by looking for deviation from the 1 : 1 ratio of patient and control DNA in a microarray. Ratios in excess of 1 indicate duplications, whereas ratios below 1 indicate deletions.

CGH

Patient DNA Label DNA with different fluorescent dyes Mix equimolar amounts of labelled DNA Apply DNA mix to glass slide with high-density array of different DNA probes with known location in the human genome

Patient/control ratio = 0.5:1 → deletion of patient DNA
 Patient/control ratio = 1.5:1 → duplication of patient DNA
 Patient/control ratio = 1:1 → normal
 Normal control DNA

Interrogating the genome: the changing landscape of genomic technologies • 53

human genome, allows the remarkably accurate determination of the genomic origin where a read consists of 25 nucleotides or more. Variants are identified as differences between the read and the reference genome. For instance, if there is a different nucleotide in half the reads at a given position compared to the reference genome, this is likely to represent a heterozygous base substitution. The number of reads that align at a given point is called the 'depth' or 'coverage'. The higher the read depth, the more accurate the variant call. However, in general, a depth of 30 or more reads is generally accepted as producing diagnostic-grade results. Rather than sequencing only one small section of DNA at a time, NGS allows the analysis of many hundreds of thousands of DNA strands in a single experiment and so is also commonly referred to as multiple parallel

sequencing technology. Today's NGS machines can sequence the entire human genome in a single day at a cost approaching 1000 US dollars. NGS capture Although we now have the capability to sequence the entire genome in a single experiment, whole-genome sequencing is not always the optimal use of NGS. NGS capture refers to the 'pull-down' of a targeted region of the genome and may constitute several to several hundred genes associated with a given phenotype (a gene panel), the exons of all known coding genes (an exome), or the exons of all coding genes known to be associated with disease (a clinical exome). Each of these targeted resequencing approaches is associated with a number of advantages and disadvantages (see Box 3.9). In order for NGS to be used for optimal patient benefit, it is essential for the clinician to have a good understanding of which test is the best one to request in any given clinical presentation.

Challenges of NGS technologies Genomic technologies have the potential to transform the way that we practise medicine, and ever faster and cheaper DNA sequencing offers increasing opportunities to prevent, diagnose and treat disease. However, genomic technologies are not without their challenges: for instance, storing the enormous quantities of data generated by NGS. While the A, C, T and G of our genomic code could be stored on the memory of a smartphone, huge computers, able to store several petabytes of data (where 1 petabyte is 1 million gigabytes of data), are required to store the information needed to generate each individual's genome. Even if we can store and handle these huge datasets successfully, we then need to be able to sift through the millions of Fig. 3.11

The polymerase chain reaction (PCR). PCR involves adding a tiny amount of the patient's DNA to a reaction containing primers (short oligonucleotides 18–21 bp in length, which bind to the DNA flanking the region of interest) and deoxynucleotide phosphates (dATP, dCTP, dGTP, dTTP), which are used to synthesise new DNA and a heat-stable polymerase. The reaction mix is first heated to 95°C, which causes the double-stranded DNA molecules to separate. The reaction is then cooled to 50–60°C, which allows the primers to bind to the target DNA. The reaction is then heated to 72°C, at which point the polymerase starts making new DNA strands. These cycles are repeated 20–30 times, resulting in exponential amplification of the DNA fragment between the primer sites. The resulting PCR products can then be used for further analysis – most commonly, DNA sequencing (see Fig. 3.12).

Cool ~60°C DNA sample DNA strands separate Primers bind to DNA DNA replicated Heat 95°C Heat 95°C Cool ~60°C Repeat cycles 20–30 times

PCR cycles Exponential amplification of DNA between primer sites DNA molecules

DNA strands separate Primers bind to DNA Cycle no. 1 DNA replicated Heat ~72°C Heat ~72°C Polymerase

- dNTPs Primers Cycle no. 2

54 • CLINICAL GENETICS their interpretation will require input from a genetics expert in the context of the clinical presentation, where an 'innocent until proven guilty' approach is often adopted. Finally, if we are to interrogate the entire genome or even the exome, it is foreseeable that we will routinely identify 'incidental' or secondary findings – in other words, findings not related to the initial diagnostic question. The UK has so far advocated a conservative approach to incidental findings. Uses of NGS NGS is now frequently used, within diagnostic laboratories, to identify base substitutions and indels (although the latter were Fig. 3.12 Sanger sequencing of DNA, which is very widely used in DNA diagnostics. This is performed using PCR-amplified fragments of DNA corresponding to the gene of interest. The sequencing reaction is carried out with a combination

dNTP and fluorescently labelled di-deoxy-dNTP (ddATP, ddTTP, ddCTP and ddGTP), which become incorporated into the newly synthesised DNA, causing termination of the chain at that point. The reaction products are then subject to capillary electrophoresis and the different-sized fragments are detected by a laser, producing a sequence chromatogram that corresponds to the target DNA sequence. Capillary electrophoresis Largest fragments migrate slowest Smallest fastest Laser fluorescence detector ddTTP ddCTP ddATP ddGTP DNA sample PCR DNA sequence chromatogram Key Fragments detected by laser fluorescence New DNA molecules terminated by incorporation of ddNTP Polymerase

- ddNTPs Primers 3.9 The advantages and disadvantages of whole-genome sequencing, whole-exome sequencing and gene panels Test Advantages Disadvantages Whole-genome sequencing (WGS) The most comprehensive analysis of the genome available More even coverage of genes, allowing better identification of dosage abnormalities Will potentially detect all gene mutations, including intronic mutations More expensive to generate and store Will detect millions of variants in non-coding DNA, which can be very difficult to interpret Associated with a greater risk of identifying incidental findings Shallow sequencing (few reads per gene) and so less sensitive and less able to detect mosaicism Whole-exome sequencing (WES) Cheaper than whole-genome sequencing Analysis is not restricted to only those genes known to cause a given condition Fewer variants detected than in WGS and so easier interpretation Deeper sequencing than WGS increases sensitivity and detection of mosaicism Less even coverage of the genome and so dosage abnormalities are more difficult to detect Less comprehensive analysis (1–2% of the genome) than WGS Increased risk of identifying incidental findings over targeted gene sequencing Gene panels Cost-effective Very deep sequencing, increasing the chances of mosaicism being detected Fewer variants detected and so data easier to interpret As analysis is restricted to known genes, the likelihood of a variant being pathogenic is greatly increased Will only detect variation in genes known to cause a given condition Difficult to add new genes to the panel as they are discovered normal variants to identify the single (or, rarely, several) pathogenic, disease-causing mutation. While this can, to an extent, be achieved through the application of complex algorithms, these take time and considerable expertise to develop and are not infallible. Furthermore, even after these data have been sifted by bioinformaticians, it is highly likely that clinicians will be left with some variants for which there are insufficient data to enable their definitive categorisation as either pathogenic or non-pathogenic. This may be because we simply do not know enough about the gene, because the particular variant has not previously been reported and/or it is identified in an unaffected parent. These variants must be interpreted with caution and, more usually,

Interrogating the genome: the changing landscape of genomic technologies • 55

Fig. 3.13 Sequencing by synthesis as used in the Illumina system. (1) Library preparation: DNA is fragmented and specialist adapters are ligated to the fragmented ends. (2) Cluster amplification: the library is loaded to a flow cell and the adapters hybridise to the flow-cell surface. Each bound fragment is hybridised. (3) Sequencing. (4) Alignment and variant interpretation: reads are aligned to a reference sequence using complex software and differences between reference and case genomes are identified. CCGATATCTAGCTTA ATATCTAGC CG TAGC TATCTAGC CCG TAGCTAGCTTA

1 Library preparation 2 Cluster amplification Genomic DNA Fragmentation Adapter ligation Flow cell Amplification 3 Sequencing Reads Reference genome 4 Alignment and variant interpretation G T A C A A 3.10 Next-generation sequencing methods Sequencing by synthesis (Fig. 3.13) • The most frequently used NGS method • Used in Illumina systems (commonly used in diagnostic laboratories) • Uses fluorescently labelled, terminator nucleotides that are sequentially incorporated into a growing DNA chain • Library DNA samples (fragmented DNA flanked by DNA adapter sequences) are anchored to a flow cell by hybridisation of the DNA adapter sequence to probes on the flow-cell surface • Amplification occurs by washing the flow cell in a mixture containing all four fluorescently labelled terminator nucleotides: A, C, T and G • Once the nucleotide, complementary to the first base of the DNA template, is incorporated, no further nucleotides can be added until the mixture is washed away • The nucleotide terminator is shed and the newly incorporated nucleotide reverts to a regular, non-fluorescent nucleotide that can be extended • The process is then repeated with the incorporation of a second base etc. • Sequencing by synthesis is therefore space- and time-dependent: a sensor will detect the order of fluorescent emissions for each spot on the plate (representing the cluster) and determine the sequence for that read Sequencing by ligation • Used in SOLiD systems • Uses DNA ligase rather than DNA polymerase (as is used in sequencing by synthesis) and short oligonucleotides (as opposed to single nucleotides) • Library DNA samples are washed in a mixture containing oligonucleotide probes representing 4–16 dinucleotide sequences. Only one nucleotide in the probe is fluorescently labelled • The complementary oligo probes will hybridise, using DNA ligase, to the target sequence, initially at a primer annealed to the anchor site and then progressively along the DNA strand • After incorporation of each probe, fluorescence is measured and the dye is cleaved off • Eventually, a new strand is synthesised (composed of a series of the oligo probes) • A new strand is then synthesised but is offset by one nucleotide • The process is repeated a number of times (5 rounds in the SOLiD system), providing overlapping templates that are analysed and a composite of the target sequence determined Ion semiconductor sequencing • When a nucleotide is incorporated into a growing DNA strand, a hydrogen ion is released that can be detected by an alteration in the pH of the solution. This hydrogen ion release forms the basis of ion semiconductor sequencing • Each amplified DNA cluster is located above a semiconductor transistor, capable of detecting differences in the pH of the solution • The DNA cluster is washed in a mixture containing only one type of nucleotide • If the correct nucleotide, complementary to the next base on the DNA template, is in the mixture and incorporated, a hydrogen ion is released and detected • If a homopolymer (sequence of two or more identical nucleotides) is present, this will be detected as a decrease in pH proportionate to the number of identical nucleotides in the sequence initially problematic). The current NGS challenge is to detect large deletions or duplications spanning several hundreds or thousands of bases and therefore exceeding any single read. Increasingly, however, this dosage analysis is being achieved using sophisticated computational methods, negating the need for more traditional technologies such as array CGH. Additional potential uses of NGS include detection of balanced and unbalanced translocations and mosaicism: NGS has proved remarkably sensitive at detecting the latter when there is high read coverage for a given region. Of note, however, NGS is still not able to interrogate the epigenome (and so will not identify conditions caused by a disruption of imprinting, such as Beckwith–Wiedemann, Silver–Russell, Angelman’s and Prader–Willi syndromes) and will not detect triplet repeat expansions such as those that cause Huntington’s disease,

56 • CLINICAL GENETICS regions of the genome, and therefore genes, more strongly associated with a given SNP profile and therefore more likely to contribute to the disease under study.

Genomics and obstetrics Prenatal genetic testing may be performed where a pregnancy is considered at increased risk of being affected with a genetic condition, either because of the ultrasound/biochemical screening results or because of the family history. While invasive tests, such as amniocentesis and chorionic villus sampling, have been the mainstay of prenatal diagnosis for many years, they are increasingly being superseded by non-invasive testing of cell-free fetal DNA (cffDNA), originating from placental trophoblasts and detectable in the maternal circulation from 4–5 weeks' gestation; it is present in sufficient quantities for testing by 9 weeks.

- Non-invasive prenatal testing (NIPT): the sequencing and quantification, using NGS, of cffDNA chromosome-specific DNA sequences to identify trisomy 13, 18 or 21. The accuracy of NIPT in detecting pregnancy-specific aneuploidy approaches 98%. A false-negative result can occur when there is too little cffDNA (possibly due to early gestation or high maternal body mass index) or when aneuploidy has arisen later in development and is confined to the embryo and not represented in the placenta. False positives can occur with confined placental mosaicism (describing aneuploidy in the placenta, not the fetus) or with an alternative cause of aneuploidy in the maternal circulation, such as cell-free tumour DNA.
- Non-invasive prenatal diagnosis (NIPD): the identification of a fetal single-gene defect that either has been paternally inherited or has arisen de novo and so is not identifiable in the maternal genome. Examples of conditions that are currently amenable to NIPD include achondroplasia and the craniosynostoses. Increasingly, however, NIPD is being used for autosomal recessive conditions such as cystic fibrosis, where parents are carriers for different mutations. The free fetal DNA is tested to see whether the paternal mutation is identified and, if absent, the fetus is not affected. If the paternal mutation is identified, however, a definitive invasive test is required to determine whether the maternal mutation has also been inherited and the fetus is affected. Where a genetic diagnosis is known in a family, a couple may opt to undertake pre-implantation genetic diagnosis (PGD). PGD is used as an adjunct to in vitro fertilisation and involves the genetic testing of a single cell from a developing embryo, prior to implantation.

Genomics and oncology Until recently, individuals were stratified to genetic testing if they presented with a personal and/or family history suggestive of an inherited cancer predisposition syndrome (Box 3.11). Relevant clinical information included the age of cancer diagnosis and number/type of tumours. For example, the diagnosis of bilateral breast cancer in a woman in her thirties with a mother who had ovarian cancer in her forties is suggestive of BRCA1/2-associated familial breast/ovarian cancer. In many familial cancer syndromes, somatic mutations act together with an inherited mutation to cause specific cancers (p. 50). Familial cancer syndromes may be due to germ-line loss-of-function mutations in tumour suppressor genes encoding DNA repair enzymes or proto-oncogenes. At the cellular level, loss of one copy of a tumour suppressor myotonic dystrophy and fragile X syndrome (see Boxes 3.8 and 3.2).

Third-generation sequencing Increasingly, third-generation or single-molecule sequencing is entering the diagnostic arena. As with next- or second-generation sequencing, a number of different platforms are commercially available. One of the most successful is SMRT technology (single-molecule sequencing in real time), developed by Pacific Biosciences. This system utilises a single-stranded DNA molecule (as compared to the amplified clusters used in NGS), which acts as a template for the sequential incorporation, using a polymerase, of fluorescently labelled nucleotides. As each complementary nucleotide is added, the fluorescence (and therefore the identity of the nucleotide) is recorded before it is removed and another nucleotide is added. A key advantage of third-generation sequencing is the long length of the read it generates: in the region of 10–15 kilobases.

It is also cheaper than NGS, as fewer reagents are required. Given these inherent advantages, third-generation sequencing is likely to supersede NGS in the near future. Given the confusion surrounding the terminology of NGS and third-generation sequencing, these technologies are increasingly referred to as 'massively parallel sequencing'. Genomics and clinical practice

Genomics and health care Genomics in rare neurodevelopmental disorders Although, by definition, the diagnosis of a rare disorder is made infrequently, rare diseases, when considered together, affect about 3 million people in the UK, the majority of whom are children. NGS has transformed the ability to diagnose individuals affected by a rare disease. Whereas previously, when we were restricted to the sequential analysis of single genes, a clinician would need to make a clinical diagnosis in order to target testing, NGS allows the interrogation of multiple genes in a single experiment. This might be done through a gene panel, a clinical exome or an exome (see Box 3.9 and p. 53), and has increased the diagnostic yield in neurodevelopmental disorders to approximately 30%. Not only does the identification of the genetic cause of a rare disorder potentially provide families with answers, prognostic information and the opportunity to meet and derive support from other affected families but also it can provide valuable information for those couples planning further children and wishing to consider prenatal testing in the future. Genomics and common disease Most common disorders are determined by interactions between a number of genes and the environment. In this situation, the genetic contribution to disease is termed polygenic. Until recently, very little progress had been made in identifying the genetic variants that predispose to common disorders, but this has been changed by the advent of genome-wide association studies. A GWAS typically involves genotyping many (> 500 000) genetic markers (SNPs) spread across the genome in a large group of individuals with the disease and in controls. By comparing the SNP genotypes in cases and controls, it is possible to identify

Genomics and clinical practice • 57

ever) in some members of these cancer-prone families. In DNA repair diseases, the inherited mutations increase the somatic mutation rate. Autosomal dominant mutations in genes encoding components of specific DNA repair systems are relatively common causes of familial colon cancer and breast cancer (e.g. BRCA1). Increasingly, genetics is moving into the mainstream, becoming integrated into routine oncological care as new gene-specific treatments are introduced. Testing for a genetic predisposition gene does not have any functional consequences, as the cell is protected by the remaining normal copy. However, a somatic mutation affecting the normal allele is likely to occur in one cell at some point during life, resulting in complete loss of tumour suppressor activity and a tumour developing by clonal expansion of that cell. This two-hit mechanism (one inherited, one somatic) for cancer development is known as the Knudson hypothesis. It explains why tumours may not develop for many years (or 3.11 Inherited cancer predisposition syndromes Syndrome name Gene Associated cancers Additional clinical features Birt-Hogg-Dubé syndrome FLCN Renal tumour (oncocytoma, chromophobe (and mixed), renal cell carcinoma) Fibrofolliculoma Trichodiscoma Pulmonary cysts Breast/ovarian hereditary susceptibility BRCA1 BRCA2 Breast carcinoma Ovarian carcinoma Pancreatic carcinoma Prostate carcinoma Cowden's syndrome PTEN Breast carcinoma Thyroid carcinoma Endometrial carcinoma Macrocephaly Intellectual disability/autistic spectrum disorder Trichilemmoma Acral keratosis Papillomatous papule Thyroid cyst Lipoma Haemangioma Intestinal hamartoma Gorlin's syndrome/basal cell naevus syndrome PTCH1 Basal cell carcinoma Medulloblastoma Odontogenic keratocyst Palmar or plantar pits Falx calcification Rib abnormalities (e.g. bifid, fused or missing ribs) Macrocephaly Cleft lip/palate

Li-Fraumeni syndrome TP53 Sarcoma (e.g. osteosarcoma, chondrosarcoma, rhabdomyosarcoma)
Breast carcinoma Brain cancer (esp. glioblastoma) Adrenocortical carcinoma Brain Lynch's
syndrome/ hereditary non-polyposis colon cancer MLH1 MSH2 MSH6 PMS2 Colorectal carcinoma
(majority right-sided) Endometrial carcinoma Gastric carcinoma Cholangiocarcinoma Ovarian
carcinoma (esp. mucinous) Multiple endocrine neoplasia 1 MEN1 Parathyroid tumour Endocrine
pancreatic tumour Anterior pituitary tumour Lipoma Facial angiofibroma Multiple endocrine
neoplasia 2 and 3 (also known as 2a and 2b, respectively) RET Medullary thyroid tumour
Pheochromocytoma Parathyroid tumour Polyposis, familial adenomatous (FAP) APC Colorectal
adenocarcinoma (FAP is characterised by thousands of polyps from the second decade; without
colectomy, malignant transformation of at least one of these polyps is inevitable) Duodenal
carcinoma Hepatoblastoma Desmoid tumour Congenital hypertrophy of the retinal pigment
epithelium (CHRPE) Polyposis, MYH-associated MYH (MUTYH) Colorectal adenocarcinoma Duodenal
adenocarcinoma Retinoblastoma, familial RB1 Retinoblastoma Osteosarcoma

58 • CLINICAL GENETICS azathioprine, a drug that is used in the treatment of autoimmune diseases and in cancer chemotherapy. Genetic screening for polymorphic variants of TPMT can be useful in identifying patients who have increased sensitivity to the effects of azathioprine and who can be treated with lower doses than normal. Gene therapy and genome editing Replacing or repairing mutated genes (gene therapy) is challenging in humans. Retroviral-mediated ex vivo replacement of the defective gene in bone marrow cells for the treatment of severe combined immune deficiency syndrome (p. 79) has been successful. The major problems with clinical use of virally delivered gene therapy have been oncogenic integration of the exogenous DNA into the genome and severe immune response to the virus. Other therapies for genetic disease include PTC124, a compound that can 'force' cells to read through a mutation that results in a premature termination codon in an ORF with the aim of producing a near-normal protein product. This therapeutic approach could be applied to any genetic disease caused by nonsense mutations. The most exciting development in genetics for a generation has been the discovery of accurate, efficient and specific techniques to enable editing of the genome in cells and organisms. This technology is known as CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated) genome editing. It is likely that ex vivo correction of genetic disease will become commonplace over the next few years. In vivo correction is not yet possible and will take much longer to become part of clinical practice. Induced pluripotent stem cells and regenerative medicine Adult stem-cell therapy has been in wide use for decades in the form of bone marrow transplantation. The identification of adult stem cells for other tissues, coupled with the ability to purify and maintain such cells in vitro, now offers exciting therapeutic potential for other diseases. It was recently discovered that many different adult cell types can be trans-differentiated to form cells (induced pluripotent stem cells or iPS cells) with almost all the characteristics of embryonal stem cells derived from the early blastocyst. In mammalian model species, such cells can be taken and used to regenerate differentiated tissue cells, such as in heart and brain. They have great potential both for the development of tissue models of human disease and for regenerative medicine. Pathway medicine The ability to manipulate pathways that have been altered in genetic disease has tremendous therapeutic potential for Mendelian disease, but a firm understanding of both disease pathogenesis and drug action at a biochemical level is required. An exciting example has been the discovery that the vascular pathology associated with Marfan's syndrome is due to the defective fibrillin molecules causing up-regulation of transforming growth factor (TGF)- β signalling in the vessel wall. Losartan is an antihypertensive drug that is marketed as an

angiotensin II receptor antagonist. However, it also acts as a partial antagonist of TGF- β signalling and is effective in preventing aortic dilatation in a mouse model of Marfan's syndrome, showing promising effects in early human clinical trials. to cancer is therefore moving from the domain of clinical genetics, where it has informed diagnosis, cascade treatment and screening/prophylactic management, to oncology, where it is informing the immediate management of the patient following cancer diagnosis. This is exemplified by BRCA1 and BRCA2 (BRCA1/2)-related breast cancer. Previously, women with a mutation in either the BRCA1 or BRCA2 gene would have received similar first-line chemotherapy to women with a sporadic breast cancer without a known genetic association. More recently, it has been shown that BRCA1/2 mutation-positive tumours are sensitive to poly ADP ribose polymerase (PARP) inhibitors. PARP inhibitors block the single-strand break-repair pathway. In a BRCA1/2 mutation-positive tumour - with compromised double-strand break repair - the additional loss of the singlestrand break-repair pathway will drive the cell towards apoptosis. Indeed, PARP inhibitors have been shown to be so effective at destroying BRCA1/2 mutation-positive tumour cells, and with such minimal side-effects, that BRCA1/2 gene testing is increasingly determining patient management. It is likely, with a growing understanding of the genomic architecture of tumours, increasing accessibility of NGS and an expanding portfolio of gene-directed therapies, that testing for many of the other inherited cancer susceptibility genes will, in time, move into the mainstream. Genomics in infectious disease NGS technologies are also transforming infectious disease. Given that a microbial genome can be sequenced within a single day at a current cost of less than 100 US dollars, microbiologists are able to identify a causative microorganism and target effective treatment rapidly and accurately. Moreover, microbial genome sequencing enables the effective surveillance of infections to reduce and prevent transmission. Finally, an understanding of the microbial genome will drive the development of vaccines and antibiotics, essential in an era characterised by increasing microbial resistance to established antibiotic agents. Treatment of genetic disease Pharmacogenomics Pharmacogenomics is the science of dissecting the genetic determinants of drug kinetics and effects using information from the human genome. For more than 50 years, it has been appreciated that polymorphic mutations within genes can affect individual responses to some drugs, such as loss-of-function mutations in CYP2D6 that cause hypersensitivity to debrisoquine, an adrenergic-blocking medication formerly used for the treatment of hypertension, in 3% of the population. This gene is part of a large family of highly polymorphic genes encoding cytochrome P450 proteins, mostly expressed in the liver, which determine the metabolism of a host of specific drugs. Polymorphisms in the CYP2D6 gene also determine codeine activation, while those in the CYP2C9 gene affect warfarin inactivation. Polymorphisms in these and other drug metabolic genes determine the persistence of drugs and, therefore, should provide information about dosages and toxicity. With the increasing use of NGS, genetic testing for assessment of drug response is seldom employed routinely, but in the future it may be possible to predict the best specific drugs and dosages for individual patients based on genetic profiling: so-called 'personalised medicine'. An example is the enzyme thiopurine methyltransferase (TPMT), which catabolises

Further information • 59

BRCA1/2 mutations), DCT is undertaken in isolation with no direct access to professional support. Furthermore, in addition to some (common) single-gene mutations, such as the founder BRCA1/2 mutations frequently identified in the Ashkenazi Jewish population and discussed in this example, current DCT packages utilise a series of SNPs to determine an overall risk profile; they evaluate the

number of detrimental and protective SNPs for a given disease. However, given that only a minority of the risk SNPs have so far been characterised, this is often inaccurate. Individuals may be falsely reassured that they are not at increased risk of a genetic condition despite a family history suggesting otherwise, resulting in inadequate surveillance and/or management. The ethical considerations listed in this clinical scenario give just a flavour of some of the issues frequently encountered in clinical genetics. They are not meant to be an exhaustive summary and whole textbooks and meetings are devoted to the discussion of hugely complex ethical issues in genetics. However, a guiding principle is that, although each counselling situation will be unique with specific communication and ethical challenges, a genetic result is permanent and has implications for the whole family, not just the individual. Where possible, therefore, an informed decision regarding genetic testing should be taken by a competent adult following counselling by an experienced and appropriately trained clinician. Further information Books and journal articles Alberts B, Bray D, Hopkin K et al. Essential cell biology, 4th edn. New York: Garland Science; 2013. Firth H, Hurst JA. Oxford desk reference: clinical genetics. Oxford: Oxford University Press; 2005. Read A, Donnai D. New clinical genetics, 2nd edn. Banbury: Scion; 2010. Strachan T, Read A. Human molecular genetics, 4th edn. New York: Garland Science; 2010. Websites bsgm.org.uk British Society for Genetic Medicine; has a report on genetic testing of children. decipher.sanger.ac.uk Excellent, comprehensive genomic database. ensembl.org Annotated genome databases from multiple organisms. futurelearn.com/courses/the-genomics-era Has a Massive Open Online Course on genomics, for which one of the authors of the current chapter is the lead educator. genome.ucsc.edu Excellent source of genomic information. ncbi.nlm.nih.gov Online Mendelian Inheritance in Man (OMIM). ncbi.nlm.nih.gov/books/NBK1116/ Gene Reviews: excellent US-based source of information about many rare genetic diseases. orpha.net/consor/cgi-bin/index.php Orphanet: European-based database on rare disease. Ethics in a genomic age As genomic technology is increasingly moving into mainstream clinical practice, it is essential for clinicians from all specialties to appreciate the complexities of genetic testing and consider whether genetic testing is the right thing to do in a given clinical scenario. To exemplify the ethical considerations associated with genetic testing, it may be helpful to think about them in the context of a clinical scenario. As you read the scenario, try to think what counselling/ethical issues might arise. A 32-year-old woman is referred to discuss BRCA2 testing; she is currently pregnant with her second child (she already has a 2-year-old daughter) and has an identical twin sister. Her mother, a healthy 65-year-old with Ashkenazi Jewish ancestry, participated in direct-to-consumer testing (DCT) for 'a bit of fun' and a BRCA2 mutation - common in the Ashkenazi Jewish population - was identified. There is no significant cancer family history of note. Consider the following issues:

- Pre-symptomatic/predictive testing: this describes testing for a known familial gene mutation in an unaffected individual (compared with diagnostic testing, where genetic testing is undertaken in an affected individual). Although this could be considered for the unaffected patient, in the current scenario any testing would also have implications for her identical twin sister. This needs to be fully explored with the patient and her sister prior to testing. There is also the potential issue of predictive testing in the patient's first child. A fundamental tenet in clinical genetics is that predictive genetic testing should be avoided in childhood for adult-onset conditions. This is because, if no benefit to the patient is accrued through childhood testing, it is better to retain the child's right to decide for herself, when she is old enough, whether she wishes to participate in genetic testing or not.
- Prenatal testing: the principles behind predictive genetic testing in childhood can be extended to prenatal testing, i.e. if a pregnancy is being continued, a baby should not be tested for an adult-onset condition that cannot be prevented or treated in childhood.

However, prenatal testing itself is hugely controversial and there is much debate as to how severe a condition should be to justify prenatal diagnosis, which would determine ongoing pregnancy decisions. • DCT: while DCT can be interesting and empowering for individuals wishing to find out more about their genetic backgrounds, it also has several drawbacks. Perhaps the main one is that, unlike face-to-face genetic counselling (which usually precedes any genetic testing, certainly where there are serious health implications for the individual and their family, such as is associated with

7KLVSDJHLQWHQWLRQDOO\OHIWEODQN

Revision #1

Created 2026-01-08 16:24:27 UTC by Omar Ayman

Updated 2026-01-08 16:24:28 UTC by Omar Ayman