



Genetic and Epigenetic Basis of Development and Disease

3

Peter A. Kaub and Christopher P. Barnett

The publication of the first draft map of the entire human genome in 2001 launched a new era in the field of genetic testing. Then, as now, the abundance of new data about an individual's genetic makeup has likely asked just as many questions as it has answered.

Historical context is useful to fully appreciate the pace of change in genetic testing. **Figure 3.1** shows a timeline of major scientific and technological breakthroughs in this field and highlights the pace of the rapidly evolving field of next-generation sequencing.

This chapter aims to give an overview of the basis, forms, and output of genetic testing. It is intended to be a quick introductory reference and primer to more detailed resources, such as the references (predominantly reviews) and many online sources cited. Divided into two parts, the first section aims to outline genetic structures and their modes of inheritance to explain the genetic basis of disease. The second section gives an overview of the main technologies currently available for genetic testing, outlining the basic concepts underpinning each test, simple laboratory considerations, plus some commentary on result interpretation and limitations. Genetic testing now pervades all fields of pathology and medicine, with education in this area becoming a core component in all sub-disciplines. Therefore, some of the concepts presented here may seem basic but the overall structure is deliberate to allow easy reference and jumping between sections if only after a definition, refresher of theory or specific details of a test, technology or link to an online resource or database. Many very useful online learning resources have emerged in this field, often with excellent

video and other visual aids, useful to clinicians and patients alike. A few highly recommended ones include the following: [1–6].

Clear definitions of nomenclature are necessary to navigate this complex and ever-expanding field. Keywords appear in bold type to enable easy identification where they are discussed or defined.

3.1 Genetic Structures

Human cells contain a nucleus consisting of highly condensed nucleic acids, mostly deoxyribonucleic acid (DNA) with some ribonucleic acid (RNA) and protein. Chromosomes connected by a **centromere**, contain **chromatin**, consisting of DNA tightly bound around discs of **histone** (an alkaline protein), to form a **nucleosome** [7]. Chromatin structure changes during the cell cycle to allow DNA replication and repair, as well as normal gene regulation and expression. In humans, chromosomes are classified into 22 pairs of **autosomes** (numbered chromosomes) and one pair of **allosomes** (**sex chromosomes**; XX female, XY male). **Chromosome number** is based on approximate size, with chromosome 1 much larger than chromosome 22. Ploidy refers to the chromosome state; e.g., diploid for pairs of chromosomes, haploid for single chromosomes, aneuploid for an incorrect number of chromosomes (e.g., triploid $n=3$ sets of chromosomes, tetraploid $n=4$ sets of chromosomes, trisomy = 3 of a specific chromosome).

Produced through the process of meiosis in the gonads, gametes retain only one member of each pair of chromosomes (haploid; $n=1$). When gametes fuse in the process of conception to form a zygote, a paired complement (diploid; $n=2$) of chromosomes is formed. Regions of chromosomes encoding functional products (DNA or RNA) are called **genes**. When there are differences in genes between a chromosome pair each corresponding gene or region (**locus**) is referred to as an **allele**.

P. A. Kaub (✉)
Adelaide Medical School, Faculty of Health and Medical Sciences,
University of Adelaide, Adelaide, Australia
e-mail: peter.kaub@adelaide.edu.au

C. P. Barnett
Women's and Children's Hospital, North Adelaide, Paediatric and
Reproductive Genetics Unit, North Adelaide, Australia
e-mail: christopher.barnett@health.sa.gov.au

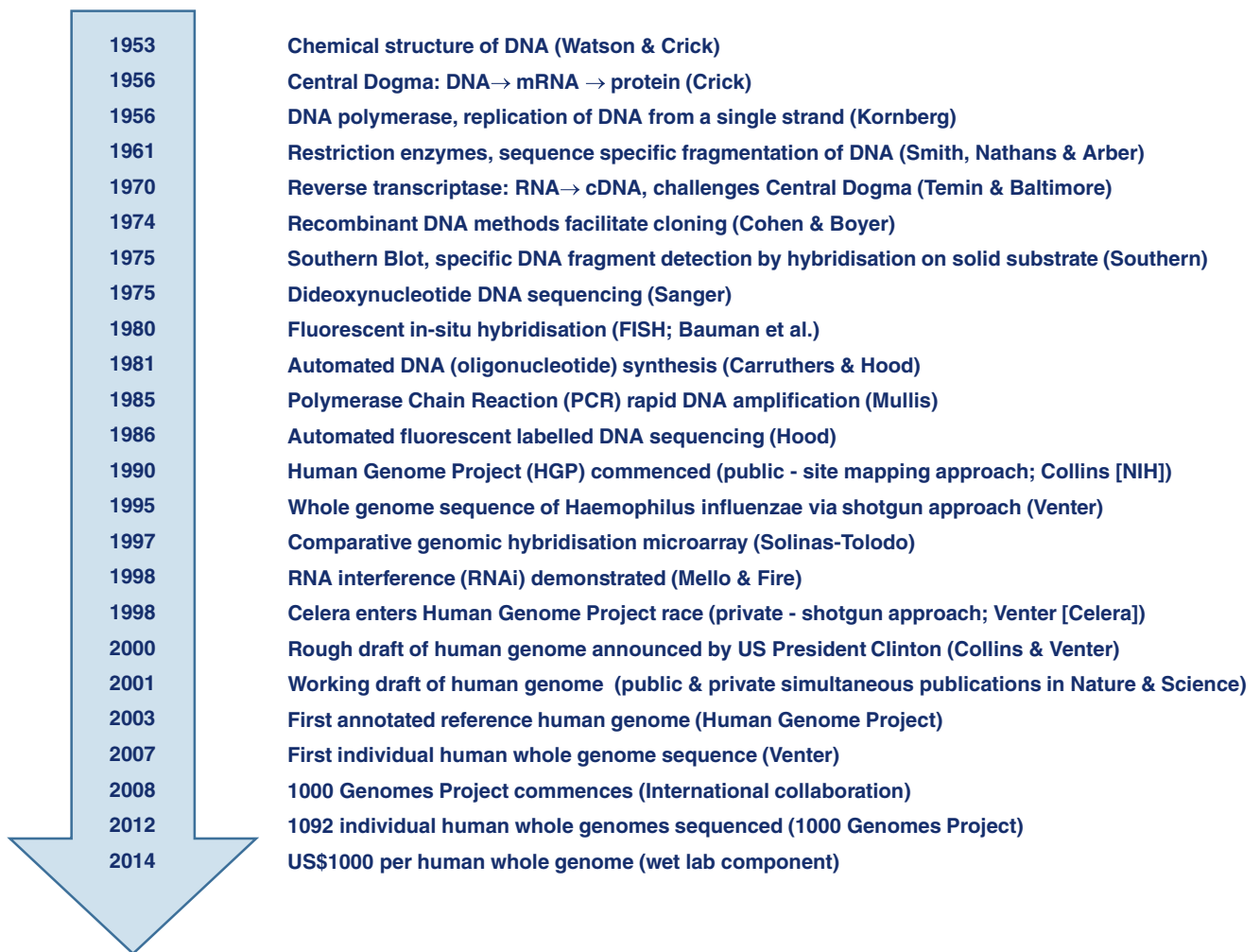


Fig. 3.1 Timeline of major scientific and technological milestones in genetic testing

Mitosis is the process of production of two daughter cells from a single cell. Each daughter cell contains identical copies of the full complement of chromosome pairs, tightly packed into the nucleus.

Cytogenetics classifies chromosomes according to well-characterised banding patterns, following special staining, to produce a **karyotype** (see Sect. 3.6.3).

Recombination is a process whereby DNA is swapped across chromosomes. It happens during meiosis across **homologous chromosomes** (containing the same alleles), to produce new variations of haploid chromosomes in the gametes—a normal function of sexual reproduction that generates diversity in offspring. Recombination can also occur during mitosis as part of normal mechanisms of **homologous recombinational repair**, usually after damage is sustained to one allele. **Non-allelic homologous recombination (NAHR)**, during meiosis or mitosis, occurs between regions with high sequence similarity, that are not alleles, resulting in deletion or insertion of whole regions, a frequent mechanism underlying **copy number variation (CNV)**.

Although there are many inbuilt checking and repair mechanisms, the processes above and other components of replication and repair have much potential for introduction of changes into DNA. Humans are about 99.5% identical at a genomic DNA level, with variations in the remaining small percentage responsible for the differences in specific traits or disease between individuals. Historically, these changes were called mutations if detrimental, or **variations** if not known to be detrimental. In response to the negative connotations historically associated with the term “mutation” it has become more acceptable to refer to a detrimental change as a **pathogenic variant** [8], or if a larger region, pathogenic structural variation. Due to long historical association, some compound terms may still be combined with “mutation” (e.g. sense and frameshift, below).

Variants and structural variation may be inherited from parents (**germline** or **constitutional**), generated during meiosis in sperm or ova (**de novo** or as **gonadal mosaicism** if more than one sperm or ova carry the same change), newly produced during the process of development of an embryo

(*de novo*), accumulated (**somatic**) from environmental exposure to chemicals, radiation, or toxins, or from normal accumulation of errors during the many cycles of replication and repair throughout life.

When unravelled, the chromosomes are found to consist mostly of a double-stranded helical structure of DNA (**Fig. 3.2**).

The chemical structure of DNA consists of a 5 carbon (pentose) sugar (deoxyribose) with base organic **nucleotides** (**cytosine, adenine, thymine, guanine**; abbreviated CATG) pairing to their complementary nucleotide (C will only bind with G, and A will only bind with T).

Nobel Laureate, Francis Crick proposed the concept of the “**Central Dogma**” to explain how DNA impacts on cell and organism-level functioning. **Translation** produces proteins via the intermediary of **messenger RNA (mRNA)** from the DNA blueprint (**transcription**) (**Fig. 3.3**). From this model also came the concept of the **gene**; i.e., a sequence of



Fig. 3.2 DNA double helix. Attached to a sugar/phosphate backbone (grey), complementary nucleotides A and T (green & red) or G and C (violet & blue) bind to each other, like rungs on a rope ladder, in the tightly wound double helix structure of DNA. The specificity of this complementary binding gives DNA its information coding and high fidelity replication abilities, plus underpins the fundamental basis for the vast majority of DNA test technologies used today

DNA responsible for producing a protein. Now too simplistic to encompass all of the increased knowledge surrounding genetic mechanisms and regulation, the current definition for a **gene** by the HGNC (HUGO Gene Nomenclature Committee) is “a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterized by sequence, transcription or homology” [9, 10]. This allows inclusion of genetic information that may play a role in modifying physiological function or regulation without explicit protein production or even without an immediately apparent functional process.

The entire sequence of nucleotides (**genome**) in humans consists of approximately 3.2 billion complementary nucleotide pairs (often called **base pairs “bp”**) bound together in double helical strands. The two strands contain an anti-parallel mirror of the sequence of each other, each nucleotide bound to its complementary pair on the opposing strand (**Fig. 3.2**). Replication of DNA requires a tightly orchestrated process involving several enzymes (described in detail in **Fig. 3.4**). The ends of a DNA strand are denoted as **5'** (five prime) or **3'** (three prime) and DNA replication always proceeds in a 5' to 3' direction. Very good animation of this process is abundant in free online video sharing sites e.g. [11, 12].

The process of DNA replication is performed with very high fidelity, but errors still occur at a rate of approximately one in every 100,000 bp. In a genome of 3 billion bp this can equate to up to 300,000 errors every time a cell divides. A key enzyme in this process, DNA polymerase, has a proof-reading mechanism that fixes about 99% of these errors. Mismatch repair is a mechanism that monitors for kinks in DNA secondary structure caused by incorrectly incorporated non-complementary nucleotides, replacing them with a complementary nucleotide. Whilst these processes are very robust, they can also cause introduction of errors in DNA sequence, which become permanent for all subsequent daughter cells.

A three nucleotide sequence (codon) and its relative alignment determines which amino acid will be **translated** into a growing protein chain; e.g., CAG for glutamine (a full codon usage table is available from the Human Genome Variation Society [13]). Variants are classified according to the impact a nucleotide change has on translation to an amino acid. Translation to the same amino acid is called a **synonymous mutation** or **variant**, to a different amino acid a **missense mutation** or **variant (non-synonymous)** and if translation is stopped by introduction of a **stop codon** it is called a nonsense mutation or **variant (Table 3.1)**.

Detrimental changes frequently occur for single nucleotides, but can involve **insertions** or **deletions** (concatenated to **indel**) of varying lengths. A change that occurs in greater than 1% of the population is, by virtue of its prevalence, likely to be a normal variant and not pathogenic. **SNPs (single nucleotide polymorphisms)** and **SNVs (single**

Fig. 3.3 Central Dogma of Genetics. In a one way, linear fashion, information coded in double-stranded DNA is transcribed into messenger RNA (mRNA), which is then translated into protein (with the assistance of ribosomal and transfer RNAs: rRNA & tRNA, respectively). Although not part of Crick's original Central Dogma description, it was subsequently determined that a precursor mRNA (pre-mRNA) step is where introns are removed and exons spliced together to form the mature mRNA transcript

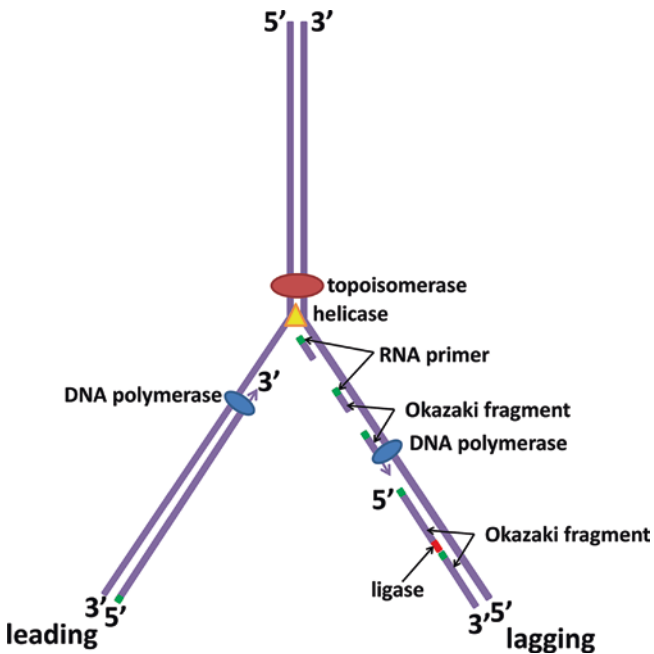
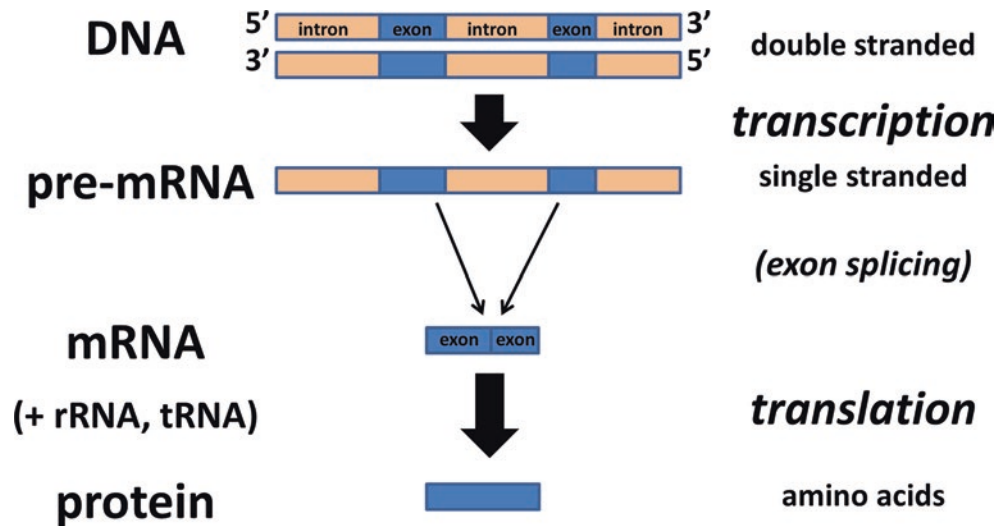


Fig. 3.4 DNA replication fork. By breaking the hydrogen bonds between complementary nucleotides and unwinding the DNA, enzymes topoisomerase and helicase combine to temporarily separate double-stranded DNA into single strands. This allows RNA primers produced by primase to anneal to complementary regions on target DNA. DNA polymerase binds next to these primers and makes a complementary copy of the single strand of DNA it is bound to, producing two molecules of double-stranded DNA. Replication can only occur in the 5' to 3' prime direction, a simple process on the "leading strand." However, the "lagging strand" requires a different approach for 5' to 3' replication, involving multiple RNA primers and piecewise production of "Okazaki fragments." The gaps between Okazaki fragments are then filled in by the enzyme DNA ligase

nucleotide variants) refer to single nucleotide changes that occur at a population and individual level, respectively. It is the latter that may be unique to an individual and worth investigating for its role in disease. A frameshift mutation or

variant involves insertion or deletion of one or more nucleotides that shift the reading frame of the following nucleotides so that the triplet codons now code for different amino acids. An **in-frame mutation** or **variant** is when the number of nucleotides changed is an exact multiple of three (**Table 3.1**). The amino acids before and after the change remain the same, but if it is at an important structural position for protein folding or subcellular localisation then it is more likely to be detrimental. In-frame expansions are also important mechanisms in triplet repeat disorders, such as Fragile X syndrome.

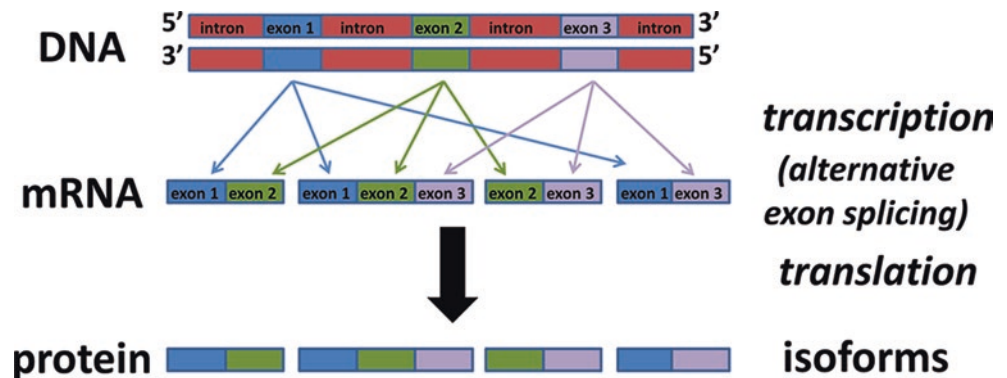
Approximately 99% of the genome consists of regions that do not code for proteins. Much of this was previously thought to be "junk DNA," but evidence continues to emerge of regulatory and other roles of untranslated regions related to tissue-specific expression, e.g., non-coding RNA (see epigenetics, Sect. 3.4). A protein-coding gene "edits" a large amount of information out in the process of transcription from DNA to mRNA. Introns are spliced out of the pre-mRNA and **exons** only are included in the mRNA transcript used for translation into protein (**Fig. 3.3**). Alternative splicing refers to a process whereby incorporation or exclusion of different exons results in alternative sizes (**isoforms**) of a protein produced from the same gene (**Fig. 3.5**). The mechanisms involved in this process are too complex to detail in this brief chapter, but suffice to say they are another potential source of detrimental changes (for review see [14]). Intronic changes are an increasingly recognised cause of human disease and much effort is currently being employed to standardise classification of their impact through tools utilising informatic analysis and functional models [15].

Genome refers to the entire genetic complement of a species or individual. The entire complement of exons is referred to as the **exome** and the entire complement of mRNA transcripts is referred to as the **transcriptome**. Similarly, the entire complement of proteins produced is called the

Table 3.1 Classification of variant (mutation) types

Variant type	DNA sequence	Amino acid (protein) change
Synonymous	CA <u>G</u> → CA <u>A</u>	glutamine → glutamine
Missense	CC <u>C</u> → CA <u>C</u>	proline → histidine
Nonsense	<u>A</u> AA → <u>I</u> AA	lysine → stop codon (*)
Frameshift (+1bp)	TGT <u>-</u> CAC → TGT <u>G</u> CAC	cysteine, histidine → cysteine, alanine & following amino acids likely changed
Deletion (in- frame)	CAG <u>TGT</u> CAC → CAG <u>-</u> CAC	loss of cysteine (often important site for disulfide bonds) between glutamine & histidine

Fig. 3.5 Alternative splicing of exons. Differential splicing of intron/exon junctions can produce different combinations of exons in the mature mRNA transcript. This results in different isoforms of protein from the same gene. Failure to remove introns or incorrect splicing of number and/or order of exons can also lead to disease



proteome, the entire methylation map, the **methylome** and all genes most commonly associated with Mendelian inheritance, the **mendeliome**.

Mitochondrial DNA (mtDNA) is a separate entity to DNA in the nucleus (**nuclear DNA**). It is a circular, small, double-stranded entity of only 16.6 kbp, coding 37 known genes, associated with oxidative phosphorylation and translation regulation. Immensely important for energy (ATP) production, mitochondria contain more than 1500 proteins, with most coded by nuclear DNA, and subsequently transported into the mitochondria. Importantly, the relatively small size and maternal inheritance pattern of mtDNA allows it to be used effectively in forensic identification on poorly preserved post-mortem material (e.g., from bones), where normal nuclear DNA may have long past degraded.

3.2 Nomenclature

It is very easy to get lost in the sea of nomenclature conventions, so only some general principles and a few examples will be given here, with links to the main nomenclature bod-

ies for detailed descriptions. A reference summary of genetic nomenclature and database sources, illustrated using pathogenic variants from two well-characterised genes and diseases is provided in **Table 3.2**. The Atlas for Genetics and Cytogenetics in Oncology and Haematology [16] has a useful, short summary of nomenclature conventions for describing genetic variation, with more detailed explanations available from the Human Genome Variation Society (HGVS) [17, 18].

HGNC [10] is responsible for overseeing gene nomenclature. Their overarching principles for gene nomenclature are:

- Try to maintain consistency of names across species
- Full gene names should be brief, specific and convey character or function (not italicized); e.g., Spinal Motor Neuron Protein 1
- Gene name abbreviations should be italicized, a combination of uppercase letters and numerals; e.g., *SMN1*
- Protein names should be the same as the gene name but not italicized; e.g., SMN1

The difference between gene (italics only) and protein name is a subtle but important one that, if adhered to, helps reduce confusion. Note that sometimes gene names change;

Table 3.2 Genetic variation nomenclature and database references

Gene name	Spinal motor neuron protein 1	Cystic fibrosis transmembrane conductance regulator
HGNC gene abbreviation (<i>italics</i>)	SMN1 (HGNC ID:11117)	CFTR (HGNC ID:1884)
Protein abbreviation (no italics)	SMN1	CFTR
Gene size (bp)	27,929	188,702
Variant description (cDNA level)	c.836G>T	c.1521_1523delCTT
Variant description (protein level)	p.G279V or p.Gly279Val	p.F508del or p.Phe508del
Cytogenetic gene co-ordinate	5q13.2	7q31.2
Gene OMIM entry	600354	602421
Disease OMIM entry	253300	219700
Gene genomic co-ordinates (GRCh38/hg20) UCSC Genome Browser	chr5: 70,925,087-70,953,015	chr7:117,479,963-117,668,664
Variant genomic co-ordinate/s (GRCh38/hg20) ENSEMBL	chr5:70,951,942	chr7:117,559,092-117,560,094
Variant OMIM entry	600354.0005	602421.0001
Variant RefSeq (rs)	rs76163360	rs113993960
Variant dbSNP	76163360	113993960
Variant Clinical significance (ClinVar)	NM_000344.3(SMN1):c.836G>T	NM_000492.3(CFTR):c.1521_1523delCTT
Genetics Home Reference	<i>SMNI</i>	<i>CFTR</i>

Two examples have been chosen to illustrate nomenclature for genetic variation. The *SMNI* and *CFTR* gene variants have a relatively high carrier frequency in many human populations. If both alleles of these genes have pathogenic change, they can result in the conditions spinal muscular atrophy (SMA) and cystic fibrosis (CF), respectively

Entries in blue have specific online entries for that gene or variant listed as hyperlinks in the references. Tools and information resources associated with genetic analysis can be explored by following the hyperlinks for these genes

URLs:

SMN1—www.genenames.org/cgi-bin/gene_symbol_report?match=SMN1

600354—www.omim.org/entry/600354

253300—www.omim.org/entry/253300

chr5: 70,925,087-70,953,015—<http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr5:70925087-70953015&hgside=1117343519>

600354.0005—www.omim.org/entry/600354#0005

rs76163360—www.ensembl.org/Homo_sapiens/Variation/Summary?v=rs76163360;toggle_HGVS_names=open

76163360—www.ncbi.nlm.nih.gov/snp/76163360

NM_000344.3(SMN1):c.836G>T—www.ncbi.nlm.nih.gov/clinvar/RCV000009738/

SMN1—<http://ghr.nlm.nih.gov/gene/SMN1>

CFTR—www.genenames.org/cgi-bin/gene_symbol_report?match=CFTR

7q31.2—www.omim.org/geneMap/7/504?start=-3&limit=10&highlight=504

602421—www.omim.org/entry/602421?search=CFTR&highlight=cftr

219700—www.omim.org/entry/219700

chr7:117,479,963-117,668,664—http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&position=chr7%3A117479963-117668664&hgside=389964619_QjqACetwCd6XdgESezNnq7o3bzEJ

chr7:117,559,092-117,560,094—http://asia.ensembl.org/Homo_sapiens/Location/View?db=core;r=7:117559092-117560094;v=rs113993960;vf=b=variation;vf=26649996

602421.0001—www.omim.org/entry/602421#0001

rs113993960—www.ensembl.org/Homo_sapiens/Variation/Summary?v=rs113993960;toggle_HGVS_names=open

113993960—www.ncbi.nlm.nih.gov/snp/113993960

NM_000492.3(CFTR):c.1521_1523delCTT—www.ncbi.nlm.nih.gov/clinvar/?term=RCV000058929%20OR%20RCV000007523%20OR%20RCV000007524%20OR%20RCV000119038

SMNI—<https://medlineplus.gov/genetics/gene/smn1/>

CFTR—<https://medlineplus.gov/genetics/gene/cftr/>

however, HGNC, OMIM [19] and other databases frequently list alternative names or symbols, alongside the currently approved gene symbol.

Units of quantity for nucleotide base pair are abbreviated following standard SI units, as in computing and other science fields (kilo, mega, giga, tera), dropping the “p” from “bp” when given a quantity prefix (i.e., Mbp becomes just Mb). For example, 32.1 kb = 32,100 bp; 3.12 Gb = 3.12 billion (3,120,000,000) bp; chr7:117.48-117.69 Mb

Accepted convention for describing genetic variation is to use the following prefixes denoting the nucleic acid type or protein the coordinates relate to:

- c. coding DNA (cDNA)
- g. genomic DNA (gDNA)
- m. mitochondrial DNA (mtDNA)
- r. RNA
- p. protein (using amino acid single- or three-letter abbreviation; e.g., G or Gly for glycine)

There are several potential ways to describe a variant. At a minimum, the description must show the genome reference consortium (**GRCh**) assembly used, the gene or locus reference sequence (NCBI/EBI RefSeq [20] or LRG [21]), and the nucleic acid level change, using the conventions described by HGVS [18]. The chromosome number, HGNC gene name and the predicted protein change (with either 3-letter or single letter amino acid abbreviations) are also often included for clarity. All the descriptions below describe the same well known “deltaF508” 3bp deletion commonly detected in the CFTR gene associated with cystic fibrosis (GRCh listed first).

[GRCh38/hg38] NM_000492.4(*CFTR*):c.1521_1523delCTT p.(F508del)

LRG_663t1:c.1521_1523delCTT p.(Phe508del)

NC_000007.14:g.117559592_117559594del

NG_016465.4:g.98809_98811del

It is acceptable to describe a variation by referencing only nucleic acid-level coordinates, but protein-level description on its own is not acceptable and must be accompanied by the nucleic acid-level coordinates to ensure precision.

Examples of some common types of genetic variations are listed below, without including gene assembly or reference sequence for brevity (see also variant descriptions in **Table 3.2**):

- *Single nucleotide substitution*: **c.456G>T** (resulting in non-synonymous amino acid change **p.G152C** or **p.Gly152Cys**)
- *Deletion (-3bp)*: **c.1521_1523delCTT** (resulting in single amino acid deletion **p.F508del**)
- *Insertion (+6bp between 343 and 344)*: **c.343_344insCAGTGG** (resulting in two amino acid insertion between arginine at 113 and the amino acid at 114; **p.R113_114insQW** or **p.Arg113_114insGluTrp**)
- *Inversion*: **c.342_1856inv** (of 1514bp fragment)
- *Frameshift (downstream stop codon)*: **p.L125QfsX20** or **p.Lys125Glufsstop20** (lysine at amino acid position 125 is changed to glutamine, with the frameshift extending for 20 amino acids, until a last stop codon).
- *Frameshift—from combined deletion (-2bp) & insertion (+1bp)*: **c.2051_2052delAAinsG**

It should be noted that shorthand to denote presence (+) or absence (-) of certain alleles has been historically used; i.e., homozygous (+/+ or -/-) and heterozygous (+/-), but is currently discouraged, unless the allele is obvious by context in the report.

The new field of epigenetics (discussed in Sect. 3.4) has spawned its own nomenclature challenges, see [22].

In cytogenetics, the definitive nomenclature reference is *An International System for Human Cytogenetic Nomenclature* [23]. Basics of cytogenetic nomenclature, with a few simple examples, are given in the section on cytogenetics (Sect. 3.6.3).

3.3 Inheritance

Knowledge of modes of inheritance is essential to understanding genetic disease processes. Most of our attributes are directly linked to inheritance from our ancestors. Austrian monk, Gregor Mendel, is credited as the first to describe the process of genetic inheritance, from experiments conducted in the mid-1800s, many years prior to elucidation of DNA as the carrier of the genetic blueprint. Although not coining the term himself, he was the first to outline the concept of an “**allele**” to describe alternative forms of the same gene or genetic element (**genotype**). As humans normally have two copies of the same gene (one inherited from each parent), it is the expression and interplay of these two alleles that determine expression of **traits**; i.e., characteristics. Phenotype refers to the trait/s actually expressed physiologically and may diverge from that expected for a certain genotype.

Mendel’s experiments with breeding garden peas and assessing mainly binary traits (e.g., color) led to three laws:

- *Law of Segregation*: when gametes form, they only retain one copy of a gene for a given location (one **allele**).
- *Law of Independent Assortment*: genes can segregate independently when gametes are formed (**recombination**).
- *Law of Dominance*: some alleles are **dominant** (express even if another allele is present) and some are **recessive** (only express if both alleles are recessive). The Law of Dominance underpins what is referred to today as “**Mendelian inheritance**” or a “**Mendelian trait**”; i.e., inheritance follows an autosomal **dominant** or autosomal **recessive** pattern in a single gene.

A compiled summary of listings on OMIM indicated 94% autosomal, 6% X-linked, 0.3% Y-linked and 0.3% mtDNA diseases [24].

Conventionally, a dominant Mendelian allele is represented by a capitalised letter (M) and recessive allele by a lowercase letter (m). There are then three possibilities of segregation depending on what alleles the parents have: *M/m* (**heterozygous**); *M/M* or *m/m* (**homozygous**) (see “Punnet square” box **Fig. 3.6**). A dominant allele will express if pres-

		Father (M/m)	
		M	m
Mother (M/m)	M	M/M	M/m
	m	M/m	m/m

Fig. 3.6 Inheritance pattern from heterozygous parents. “Punnet square” indicating inheritance of autosomal recessive (m) or dominant (M) allele from heterozygous carrier parents

ent, whether a recessive allele is present or not (M/M or M/m). A recessive allele will only be expressed in the phenotype if both alleles are recessive (m/m).

Autosomal recessive traits are inherited in a horizontal manner (see **Fig. 3.7a**). In the offspring of heterozygous (carrier) parents there is a 25% chance of autosomal recessive allele being expressed and 50% chance of being a carrier (not expressed).

Cystic fibrosis (CF) is an example of autosomal recessively inherited disease, most frequently homozygous for the most common pathogenic variant, NM_000492.4(*CFTR*):[c.1521_1523delCTT];[c.1521_1523delCTT] p.[(F508del)];[(F508del)]. However, CF also demonstrates

the concept of a **compound heterozygote**, when two different disease-associated recessive alleles in the same gene are expressed e.g., NM_000492.4(*CFTR*):[c.1521_1523delCTT];[c.1624G>T] p.[(F508del)];[(G542X)], also resulting in a disease phenotype (N.B. both descriptions use the [GRCh38/hg38] assembly).

Closely related individuals have a higher chance of carrying similar DNA, as they have closer common ancestors. Therefore, mating between genetically related individuals (**consanguinity**) increases the chance of autosomal recessive traits being expressed; i.e., the chance of alleles from parents being the same is increased the more closely related they are. The **co-efficient of inbreeding (f)** measures the theoretical

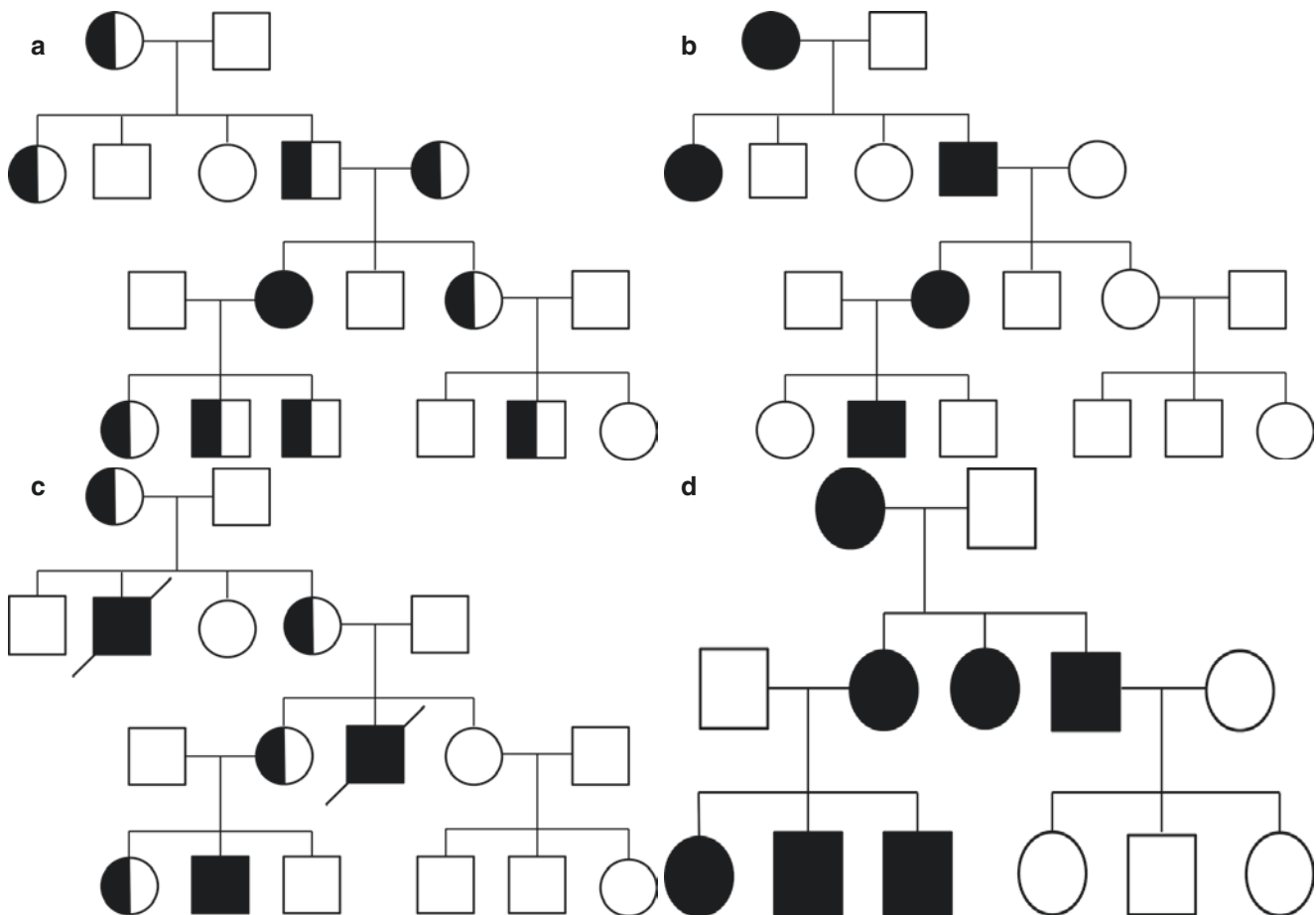


Fig. 3.7 Inheritance pattern genograms (pedigree). (a) Autosomal recessive (AR): If both parents are a carriers of an AR pathogenic change, there is a 25% chance of their child being homozygous for the change and 50% chance of them being a carrier. All children from one homozygous and one non-carrier parent will be carriers of an AR change (bottom left). (b) Autosomal dominant (AD): If either parent is affected, there is a 50% chance that their child will be affected. Age of onset and severity of disease will be dependent on penetrance and expressivity, respectively. (c) X-linked (XL): a pathogenic change is passed on through an X-chromosome. As females have two X-chromosomes, a healthy allele on one X-chromosome most often compensates for a pathogenic change on the other X-chromosome. X-linked conditions most often affect males, as they only have one copy

of the X-chromosome, with no other allele to compensate, leading to disease if their only X-chromosome contains a pathogenic change. Sons of carrier mothers have a 50% chance of being affected. Daughters of carrier mothers or affected fathers have a 50% chance of being a carrier of an X-linked condition. (d) Mitochondrial (mtDNA): all children of an affected mother will carry a mitochondrial DNA pathogenic change, as this sub-organelle DNA is only inherited from the mother. Affected males do not pass on mitochondrial DNA changes to their children, as mtDNA is normally only inherited from the mothers (female gamete). **Key** Square: male, Circle: female, Full-shading: affected, Half-shading: carrier, No shading: unaffected, Diagonal line: deceased

level of homozygosity based on **pedigree**, with first cousins expected to share 1/8 of their DNA, therefore have approximately 12.5% homozygosity ($f = 12.5\%$). The prevalence of certain alleles also differs between ethnic groups, again due to effects of closer common ancestors.

Autosomal dominant traits are inherited in a vertical manner, with a 50% chance of being passed on to offspring (**Fig. 3.7b**). There may, however, be a range (from minor to severe) of disease traits expressed in different individuals with the same dominant allele (**variable expressivity**). Some alleles may be present, but not express themselves in all individuals; **penetrance** refers to the percentage of individuals expressing the phenotype associated with a specific allele by a certain age (e.g., evidence of autosomal dominant hypertrophic cardiomyopathy is dependent on age and differs even within families). For a specific allele, penetrance refers to the chance of a phenotype being present (or not). In contrast, expressivity refers to the severity of traits expressed, implying that there is a level of phenotypic expression present, however minor it may be.

Pleiotropy (literally “affecting many”) describes where a single allele manifests phenotypically in multiple, apparently unrelated traits. Modulation of these traits may be impacted by environmental and other factors. Monozygotic twins demonstrate this concept well. Despite identical genotypes (i.e., an identical complement of alleles), monozygotic twins can express traits differently—i.e., have **discordant phenotypes**. This indicates that there are factors other than genotype that can affect phenotype (see epigenetics, below).

Haplotype refers to a subset of the genotype, usually of alleles that tend to be inherited together and frequently from the one parent. The concept of haplotype is important in historical methods used to isolate candidate disease genes through **linkage analysis** of affected individuals and families (e.g., *CFTR* gene in cystic fibrosis). This method relies on non-disease marker genes, in close proximity to a disease gene, frequently being inherited together, acting like a flag to the disease gene. Sometimes, genes in close proximity (**contiguous**) may all be affected together by relatively large DNA changes, leading to complex phenotypes that are a combination of the multiple allele changes (e.g., 11p14 deletion causing aniridia and increased risk of Wilm tumor). Hemizygous describes when there is only one of a pair of chromosomes in an individual, e.g. males are hemizygous for the X-chromosome. Haploinsufficiency refers to a reduction in relative gene expression from loss of one allele resulting in insufficient gene product to preserve normal function (e.g., 7q11.23 deletion of 26 genes in Williams syndrome).

Sex-linked inheritance follows an oblique inheritance pattern associated with segregation of the X and, very rarely, the Y chromosome (**Fig. 3.7c**). Fabry disease and hemophilia (A & B) are **X-linked disorders**, expressed in males in a hemizygous manner. Fabry disease may also present in the pheno-

type of heterozygous females to varying degrees, through the process of **X-inactivation (lyonization)**. This is the process whereby one X chromosome in each cell is randomly made transcriptionally inactive through chromatin structure changes at the time of embryo development (see epigenetics, below). Sex-determining region Y (SRY) protein on the Y chromosome is responsible for initiation of male sex determination, and faults in its expression can be responsible for differences in sex development (DSD).

Most traits are thought to be under more complex control than Mendelian inheritance, via **incomplete dominance** (both alleles expressed to some degree, with the phenotype a combination of their expression; e.g., sickle cell trait that is milder than the homozygous [*Hb^s/Hb^s*] sickle cell anemia), **co-dominance** (both alleles expressed in the phenotype; e.g., ABO blood grouping) or **digenic/polygenic** (influenced by two or more genes; e.g., autosomal recessive retinitis pigmentosa, autosomal recessive hearing loss). Mitochondrial disease follows a pattern of **maternal inheritance** only (**Fig. 3.7d**), as mitochondrial DNA (mtDNA) in a zygote is derived exclusively from the maternal oocyte. Therefore, all children from the same mother can have the same mitochondrial-derived trait, however, only daughters can pass it on to their offspring.

A **genogram** (family tree or **pedigree**) is a useful method for visualising inheritance and is often used to elicit the likely segregation pattern (**Fig. 3.7**). This can be a useful aid in refining differential diagnoses and genetic tests to be performed.

3.4 Epigenetics

Epigenetics is a relatively new field that has generated a wealth of interest, especially in its implications for genetic disease and testing. The prefix *epi* (Greek for “over” or “above”) infers a meaning of genotypic effect over and above that performed by the genome; however, its definition continues to be debated, particularly with regard to mechanisms that are not heritable. It is generally agreed that **epigenetics** refers to modulation of gene activity or expression without modification to gene sequence. The term epigenome is used to describe the complement of all epigenetic effects. The NIH Roadmap Epigenome Project [25] includes both heritable and non-heritable mechanisms in its definition, agreed to here for the purposes of discussion.

The starkest demonstration of epigenetic mechanisms is when monozygotic twins with identical genotypes express differences in phenotype, by the presence or absence of disease [26]. The depth of knowledge of this mechanism of genetic modulation and its impact on all manner of disease is still relatively new, but is increasingly finding its way into genetic diagnostics. Epigenetics may well turn out to be the

previously hidden mechanism behind a range of phenotypes not explained using classical genetic models. The hope is that it will become an important aid in determining why one person gets a disease and another of similar genotype remains unscathed.

3.4.1 Genomic Imprinting

Genomic imprinting, where an allele is completely silenced based on its parental origin, is an epigenetic phenomenon responsible for diseases such as Beckwith-Wiedemann syndrome, Prader-Willi syndrome (paternal inheritance), and Angelman syndrome (maternal inheritance) [27]. Epigenetic phenomena also underlie the process of X-inactivation (for review [28]).

Whilst further types of epigenetic regulation are likely to be discovered, the following mechanisms (all post-translational) are already known to be the basis of several epigenetic phenomena, with relevance in disease. This whole field remains one of the most active areas in biomedical research.

3.4.2 Nucleosome Position

DNA is packaged into the nucleus wrapped around **histone** proteins to form **nucleosomes**, making up the majority of the **chromatin** complex. Changes in the position of nucleosomes in the chromatin structure can affect gene transcription mechanisms by altering proximity and/or access to transcription start sites.

3.4.3 Histone Modification

Modification of histone N-terminal tails by methylation, phosphorylation, acetylation, ubiquitination, sumoylation, ribosylation or citrullination can alter the initiation of transcription of a gene. Like nucleosome positioning, it can act by altering the chromatin structure, modifying, either positively or negatively, the ability for transcription to initiate at specific sites. Histone modification has also demonstrated wider reach, able to affect DNA repair and replication, plus alternative splicing mechanisms.

3.4.4 CpG Methylation

Probably the most widely known and tested form of epigenetic modification, **methylation** of specific cytosine nucleotides can repress gene expression by inhibiting transcription factor binding and enhancing recruitment of chromatin co-

repressors. Cytosine nucleotides adjacent to a guanine (commonly referred to as **CpG** for **C**ytosine joined by a **p**hosphodiester bond to adjacent **G**uanine) are the targets for this methylation via DNA methyltransferase (DNMT) enzymes. This tends to happen in CpG-rich regions (called **CpG islands**), which frequently occur near to 5' gene promoter regions. Their effect is to repress transcription, effectively silencing a gene. The equivalent of single nucleotide polymorphisms (SNPs) for the genome, **methylation variable positions (MVPs)** are sites that show common variability in their effect on epigenetic regulation. Epigenomic maps of such information are continuing to evolve and the term **methylome** is now used to describe the entire complement of methylated CpG sequences.

3.4.5 Non-Coding RNA

Surprisingly, only 20% of RNA (mRNA) is translated into protein, posing the question of what might be the function of the remaining 80% of RNA transcripts (termed **non-coding RNA; ncRNA**). At least some ncRNAs are involved in epigenetic forms of regulation, through what is termed **RNA interference (RNAi)**. The short (20–25 bp) double-stranded molecules of **microRNA (miRNA)** not to be confused with messenger RNA [mRNA]) and **silencing RNA (siRNA)** have different but overlapping roles. Both act by directly binding to mRNA molecules, miRNA less specifically than siRNA. siRNA actively degrades already transcribed mRNA through the actions of the enzyme Dicer and protein complex RISC (see [29] for excellent animation of the process). miRNA acts to indirectly prevent translation to protein just by virtue of it binding to the 3' untranslated region of an mRNA molecule, but it can also utilise the same degradation pathway of Dicer and RISC as siRNA.

Although an arbitrary value to distinguish them from the shorter ncRNAs, **long non-coding RNAs (lncRNAs)** are at least 200 bp, but frequently much larger [30]. They work in a variety of ways, but an example is the very well characterised X-active specific transcript (XIST). XIST is a 17 kb lncRNA responsible for mediating **X-inactivation** by effectively coating the X-chromosome it is transcribed from, rendering it inactive. LNCipedia is a compendium focusing on human lncRNAs [31].

There are many other ncRNAs involved in epigenetic processes, including ribozymes (“gene shears”), Piwi-interacting (Pi-RNA), small nuclear (snRNA), small nucleolar (snoRNA) and transcription initiation (tiRNA) RNA, but they are beyond the scope of this chapter.

Exogenous manipulation and monitoring of ncRNAs, especially miRNA and siRNA, have spawned a whole new range of potential diagnostic and therapeutic possibilities.

It should be noted that the epigenetic mechanisms are often interactive, not necessarily acting in isolation, each able to up- and down-regulate the likelihood of one of the others coming into play and acting in concert to modify chromatin structure and/or gene expression. X-inactivation is an example of several of these mechanisms working in tandem for epigenetic regulation.

The International Human Epigenome Consortium (IHEC) launched the Epigenome Project [32, 33] in 2010, aiming to determine epigenomic impact on “... key cellular status relevant to health and disease”. Genome RNAi is a database compiling phenotypes resulting from RNA interference [34].

3.5 Somatic Variation

The genotype of subsets of cells and tissues may change throughout life from normal wear and tear, accumulation of errors through normal regulation and repair, or exogenous factors, such as adverse environmental exposures (e.g., radiation, toxins). Cancers, on the whole, develop in this manner, first localising abnormalities to cell subtypes, tissues and regions, then spreading through metastasis. Genetics in this area would require another whole chapter to discuss but it is just highlighted here in order to flag the rare occasions where tumors can develop *in utero* and be the obvious cause of pathology. The genetic tests for somatic cancer are obviously indicated at these times.

3.6 Genetic Testing

3.6.1 Sampling

Genetic testing requires isolation of nucleic acid (DNA or RNA) (for a quick reference summary see **Table 3.3**). RNA degrades much more rapidly than DNA, and therefore requires more careful handling and extraction. In general, the

most reliable and most frequently used sample type for genetic testing is blood transported at room temperature in an EDTA tube. Cord blood can be a useful source for testing in the early neonatal period. If blood is not available (e.g., postmortem cases), then heart, lung, and other tissues may be used directly to isolate DNA (preferably not liver as its protein and enzyme-rich composition tends to hamper good nucleic acid isolation). Skin is very robust, but lung and other tissue may also be used to culture cells from which DNA can be isolated. For cell culture, this tissue is best provided fresh on its own in a sterile sample container or in culture media (e.g., RPMI) or normal saline, stored at room temperature for short periods or 4–8 °C (not frozen) for up to a few days.

Amniocentesis (amniocytes) and chorionic villous (placenta) samples can be used directly to extract DNA, but also frequently rely on cell culture to obtain sufficient DNA or for karyotyping. Given the relatively small amount of starting material, processing is best performed immediately, therefore forewarning the laboratory about sample availability is essential. Cytogenetics uses blood in lithium heparin or sodium heparin tubes for isolation of peripheral blood lymphocytes (PBLs) to culture for isolation of chromosomes.

It is possible to isolate DNA from formalin-fixed, paraffin-embedded (FFPE) tissue but the process of fixation causes significant degradation to nucleic acids. DNA extraction can be attempted on these samples, but quality and quantity isolated is inconsistent, with a high failure rate, making this not a preferred option for germline genetic testing. For somatic genetic testing, where the majority of the FFPE sample is tumor DNA, extraction can be more useful and consistent. If used, FFPE samples should be provided dewaxed on original slides, with tumor-rich regions marked in some way.

Maternal blood in EDTA tubes is used to isolate circulating free DNA from plasma (see NIPS, Sect. 3.6.15).

It should be noted that while theoretically all of our cells should have the same genotype, mosaicism (genotypes divergent between cells in the same individual) can occur. Any

Table 3.3 General guidelines for obtaining DNA samples

Test	Tissue	Target	Collection vessel	Transport/Storage
Karyotype/FISH	Blood	Culture PBLs	Li Hep/Na Hep	Room temp. <72 h
	CVS	Placenta	Sample jar with sterile culture media	Room temp. if immediate processing, 4 °C if >48 h until processing
	Amniotic fluid	Amniocytes	Plain sterile tubes	Room temp. <48 h
All other nucleic acid-based testing	Blood	PBL DNA	K-EDTA	Room temp. <72 h
	Heart, lung, other tissue	Direct DNA isolation	In sterile sample jar	4 °C or –20 °C (do not freeze if also used for culture)
	Skin, lung, heart, liver, other tissue	Cultured cells used for DNA isolation	In sterile sample jar in RPMI, normal saline or tissue on its own	Room temp. or 4 °C <48 h
	FFPE tissue (tumor)	DNA isolation	Dewaxed on slide	Room temp. in slide box

PBL Peripheral blood lymphocyte, *Li Hep* Lithium heparin blood collection tube, *Na Hep* Sodium heparin blood collection tube, *Room temp.* Room temperature, *CVS* Chorionic villus sample, *RPMI* A type of cell culture media, *FFPE* Formalin-fixed, paraffin-embedded, *h* hours

isolated DNA will be representative of the cell or tissue type it is derived from, which may not always be representative of the genotype of all cells in the body (e.g., **placental mosaicism**).

These are general guidelines only and laboratory resources or staff should be consulted to determine what tests are available, plus the most suitable sampling, storage and transport methods.

3.6.2 Complementarity: The Basis of Genetic Testing

The machinery of DNA replication underpins the mechanism behind almost all genetic testing, other than karyotyping. Binding of a nucleotide to its complementary nucleotide in an anti-parallel, mirror-like fashion gives the structure of DNA many advantages in terms of fidelity for replication and repair. Genetic testing relies on the fact that a nucleotide sequence AGCTGGCT will only bind to its complementary sequence TCGACCGA (UGCTGGCT if RNA) and is the basis of the incredible precision possible with genetic testing. Harnessing the power of enzymes involved in the fundamental processes of DNA replication, also allows very small amounts of starting material to be amplified into sufficient quantities for a range of different genetic tests. Cytogenetics is the exception.

3.6.3 Cytogenetics

Cytogenetics is the study of chromosomes, with their number and characteristics assessed to produce a karyotype (*karyon* from Greek for nucleus). By visualising banding patterns on stained chromosomes DNA can be analysed at a gross level, with changes detectable in the 5–10 Mb range (~400 bands per haploid set (bhp) resolution).

In cytogenetics, it is important to be aware of two of the phases of mitosis. Approximately 90% of a cell's lifecycle happens in interphase, where chromosomes are highly condensed in the nucleus. As most cells are already likely to be in interphase before cell culture begins, the lead time to being able to harvest interphase cells can be as short as 24–72 h. In metaphase, chromosomes align along the equator of the cell guided by microtubules. It is at this time that chromosomes are most easily visualised, which is therefore the preferred state for karyotyping. The disadvantage of examining metaphase cells though, is the process can take considerably longer than preparation of interphase cells (usually one week, but often longer for slow growing cells or other problems requiring repeat culture).

Each chromosome has a consistent and well-characterised banding pattern, centromere location, and length allowing it

to be identified and classified. On G-banding, heterochromatin refers to the dark bands from densely packed DNA and euchromatin refers to the lighter regions, gene rich and more accessible for active transcription. Scoring individual chromosomes, from a number of cells on a slide, allows determination of gross changes that may indicate aneuploidy (anomalies in total number or character of chromosomes).

After replication, chromosomes are arranged in pairs of sister chromatids connected by a **centromere**. When a cell divides, one chromatid from each pair goes to each daughter cell. The centromere creates a division into two arms for each chromatid, with the shorter arm labeled **p** (from the French “petit”) and the longer arm labeled **q** (as it follows p in the alphabet). Location is classified by sequential numbering starting from the centromere and moving outwards (i.e., proximal to distal) on both arms. The first two numbers are region and band, respectively, (e.g., q23 is region 2, band 3). The region and band should always be stated as single numbers (i.e., for the previous example two-three, not twenty three) unless you want to raise the ire of a cytogeneticist. The centromere is the start of region 1 and sub-bands follow a decimal point after the region and band number; e.g., 13q23.1 is sub-band 1 of band 3, region 2 distal from the centromere on the q (long) arm of chromosome 13.

A karyotype is reported by a numerical value of the number of chromosomes (normal in humans is 23 pairs = 46), then sex chromosomes, then, if present, any aneuploidy. Parentheses identify the type of rearrangement, a semicolon separates alterations in two or more chromosomes and tilde (~) is used to show uncertainty in the location. Total number of cells counted is indicated in square parentheses at the end. Strict nomenclature guidelines are provided by the International Standing Committee on Cytogenetic Nomenclature [23].

A **normal karyotype** is **46,XX** (female) and **46,XY** (male). Examples of a female trisomy 13 (47,XX,+13) and a female triploid karyotype (69,XXX) are given in **Figs. 3.8** and **3.9**, respectively.

The main types of aneuploidy are **duplication, deletion, translocation, inversion, isochromosome, ring chromosome** and **uniparental disomy (UPD)**. **Terminal** and **interstitial** changes (usually deletions or duplications) refer to those near the ends or within the internal part of a chromosome, respectively. **Table 3.4** gives examples of several types of aneuploidy with an example karyotype and common disease name.

Mosaicism refers to cases where there are cells with more than one karyotype in the same individual. There are many causes, especially ageing, but all mosaic karyotypes are generated from only one zygote (**Table 3.4**). Placental mosaicism can be a cause for apparent trisomy (in cells from the placenta only) that is not present in the fetus.

Although very rare, chimerism, is where more than one karyotype exists in the same individual, originating from

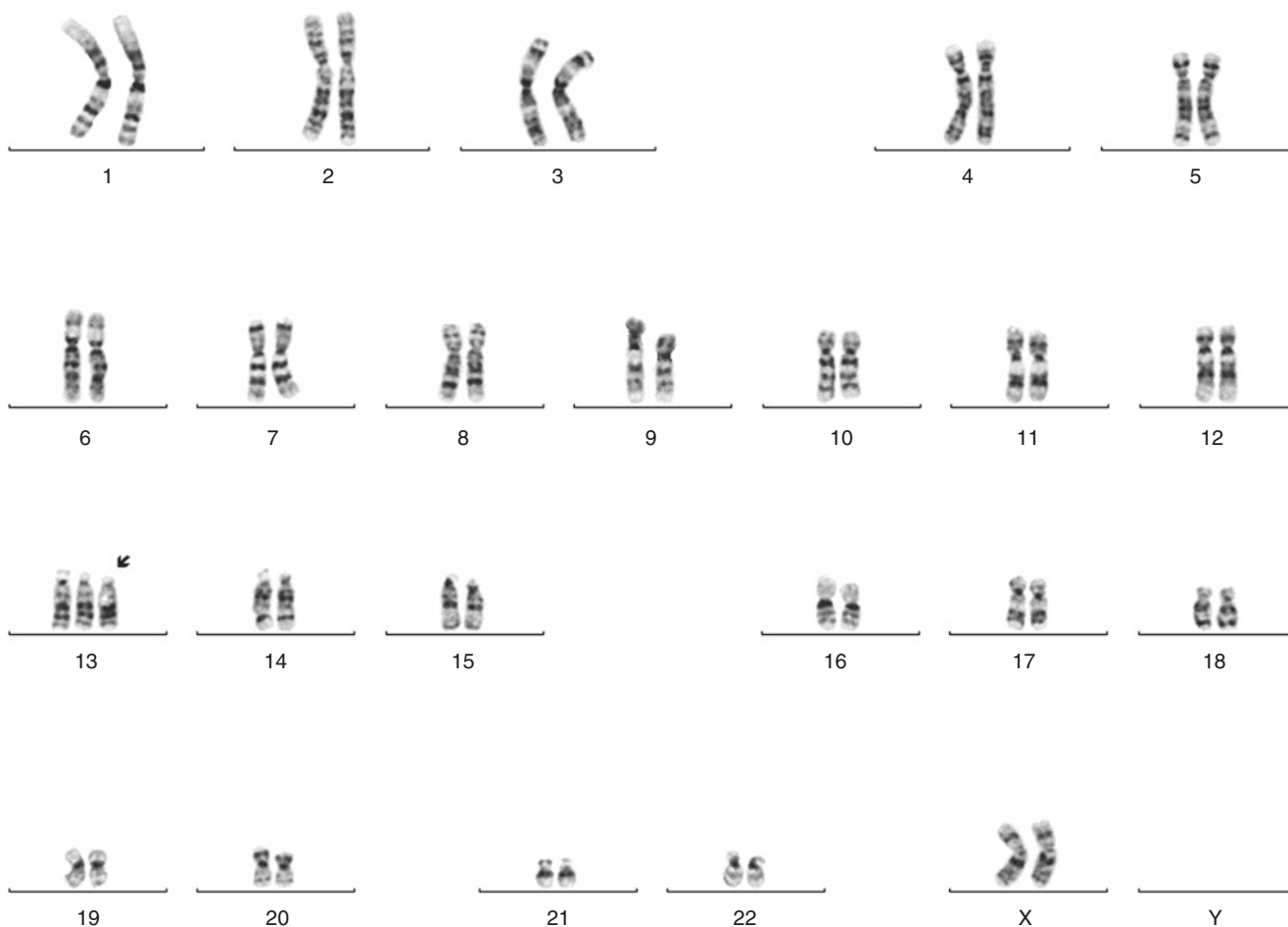


Fig. 3.8 Trisomy 13 (Patau syndrome) karyotype. One extra copy of chromosome 13 (arrow) indicating a female with Trisomy 13 (karyotype notation = 47,XX,+13). Figure courtesy of Ms. R. Hutchinson, SA Pathology, Australia

separate individual zygotes (**Table 3.4**). This occurs after successful bone marrow or other tissue transplants, but prenatally is usually the result of early embryonic twin-twin fusions resulting in a dual karyotype singleton.

Non-invasive prenatal screening (NIPS, Sect. 3.6.15) is continuing to reduce the amount of karyotyping performed for prenatal screening. However, karyotyping of amniocytes or other fetal tissue remains the gold standard, and is still used to confirm potentially pathogenic NIPS results. Also, microarray is unable to detect balanced translocations, therefore there is still likely to be a role for “classical” karyotyping for some time yet.

- Traditional karyotyping is a good test for detecting or confirming aneuploidy.

3.6.4 Fluorescent *In Situ* Hybridisation

Fluorescent *In Situ* Hybridisation (FISH) is used in cytogenetics as an alternative, as well as adjunct to karyotyp-

ing. As it can be used on interphase cells, it allows for more rapid detection of suspected aneuploidy. It can also be used to confirm or further characterise karyotype results. It relies on fluorescently labeled DNA probes (10–100 kb) that hybridise to complementary regions of DNA on chromosomes. Tens of thousands of commercial and in-house probes exist, many generated from the sequencing techniques employed in early parts of the Human Genome Project. Usually, only a small subset is used for rapid assessment or confirmation according to the suspected aneuploidy. FISH has the advantage of a relatively quick turnaround time (approximately 48–72 h from sample receipt).

The technique is similar to most nucleic acid hybridisation techniques; i.e., heat to denature DNA into single strands, followed by addition of a labeled single strand DNA probe that will bind to its complementary sequence. For FISH, this occurs in the fixed tissue (cells) on a slide, hence the “*in situ*” component of its name.

A range of different FISH probes exist, allowing different lengths, parts and characteristics of chromosomes to be visu-

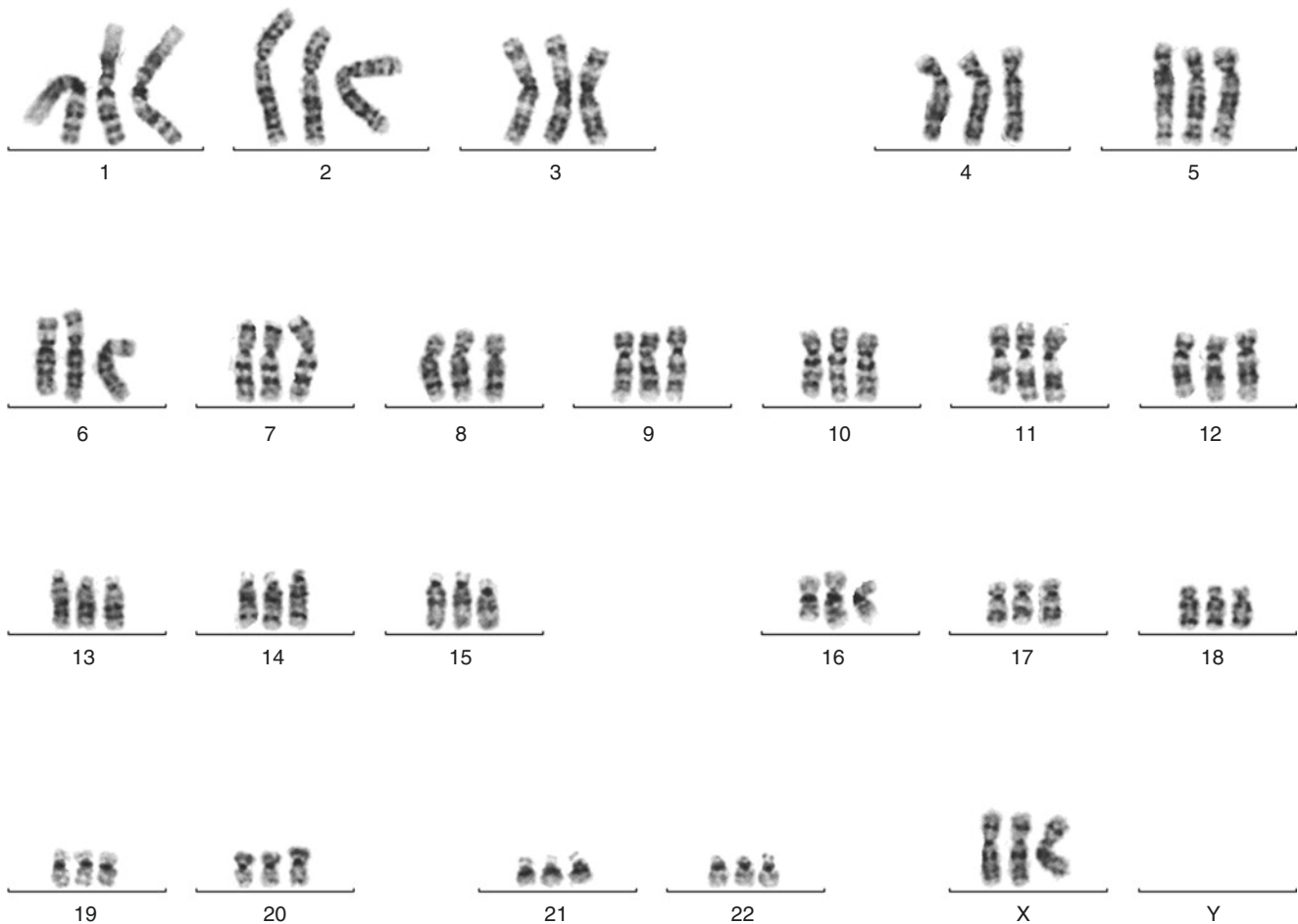


Fig. 3.9 Triploid karyotype. Three copies (3n) of each chromosome in a female (karyotype notation = 69, XXX). Figure courtesy of Ms. R. Hutchinson, SA Pathology, Australia

alised (e.g., translocation, centromere, subtelomere, fusion, breakpoint, and painting probes). The latter use multiple probes to color code all chromosome pairs different colors in the one reaction. Simple examples of trisomy 21 and sex determination by FISH are shown in **Figs. 3.10** and **3.11**, respectively.

A FISH result is denoted by “nuc ish” (for nuclear *in situ* hybridisation) for the karyotype, with probe name in parentheses and cell number counted in square parentheses following; e.g., nuc ish(D21S259/D21S341/D21S342)x3 [200/200]. Often the FISH result is reported first verbally, but usually karyotyping is also commenced in parallel, and reported later with a metaphase FISH result for confirmation. A standard karyotype is listed first, followed by the FISH result (see **Table 3.4**).

The principles behind FISH also form the basis of microarray hybridisation techniques (see Sect. 3.6.13).

- **FISH** is a good test for rapid assessment of trisomies, frequently used for fetuses rapidly approaching the cut-off age for termination. FISH is useful when targeting particular areas of the genetic code.

3.6.5 Automated DNA Sequencing

Named after its inventor, dual Nobel Laureate in Chemistry, Frederick Sanger, **dideoxynucleotide (Sanger) sequencing** was one of many systems he trialled, outlasting them and other competitors. It was the mainstay of DNA sequencing until the rise of affordable NGS platforms, but is still a cost-effective method when sequencing of only a known short region is required (e.g. cascade testing of a known familial pathogenic variant).

Sanger sequencing replicates the targeted region into many individual fragment chains differing by single nucleotides in their size and then separating them by size gives a ladder- or barcode-like pattern indicating the DNA sequence (**Fig. 3.12a**):

Output from automated sequencing is in the form of electrophoretic spectra (**Fig. 3.12b**). It should be noted that this technology essentially produces a sequence that is an average (mean) of all the DNA molecules in the sample, and therefore changes that are only a small percentage of the whole (e.g., low-level mosaicism or somatic change) are difficult to detect by this method.

Table 3.4 Karyotype—examples of cytogenetic abnormalities & nomenclature

Duplication: replication of all or part of a chromosome		
47,XY,+21	Trisomy 21 (Down syndrome)	one extra chr21 ⇒ total 47
47,XX,+18	Trisomy 18 (Edwards syndrome)	one extra chr18 ⇒ total 47
47,XXY	Klinefelter syndrome	one extra sex chromosome ⇒ XXY (total 47)
46,XX,dup(8)(p22p21.1)	partial trisomy 8	duplication & inversion of part of chr8 between region 2, band 1, sub-band 1 & band 2 (N.B. inversion in duplications is indicated by reversal of band number order i.e. 22 before 21)
Deletion: loss of all or part of a chromosome		
45,X	Turner syndrome	one missing X chromosome
46,XX,del(5)(p13)	Cri du chat syndrome	deletion of short arm (p) of one chr5 from region 1, band 3 to the subtelomere of the short arm
46,XX,del(1)(p36.3)	1p36 deletion syndrome	deletion of short arm (p) of one chr1 from region 3, band 6, sub-band 3 to the subtelomere of the short arm
Translocation: relocation of all or part of a chromosome so that it is incorporated into another chromosome		
46,XY,t(9;22)(q34;q11.2)	Philadelphia chromosome	translocation between long arms (q) of chr9 & 22
47,XX,+der(22),t(11;22)(q23;q11)	miscarriage	translocation between long arms (q) of chr11 & 22, with derivative of chr22 producing trisomy
46,X,der(X),t(X;Y)(q28;p11.31)	ambiguous genitalia	translocation (Robertsonian) between the long (q) & short (p) arms of chrX & chrY, respectively
Inversion: part of a chromosome that has reversed its direction 180 degrees, so that it is oriented in the opposite direction on the chromosome to normal		
46,XY,inv(7)(p22;q22)	some cases of fetal demise	inversion between region 2 band 2 on short arm (p) & region 2 band 2 of long arm (q) of chr7
Isochromosome: a chromosome that has lost one of its arms & replaced it with a copy of the same arm (i.e., p-p or q-q)		
46,XX,i(18)(q10)	isochromosome 18q syndrome	one chr18 has two long arms (q); the breakpoint is assigned the centromere location q10 (region 1 band 0)

- DNA (Sanger) sequencing is a good test for confirming changes in small known regions of the genome, e.g. in a

Table 3.4 (continued)

Ring chromosome: arms of a chromosome have fused together in a ring shape		
46,XX,r(15)	ring chromosome 15 syndrome	one chr15 has fused short & long arms into a ring
Uniparental disomy: both chromosomes of a pair, or parts of them, are derived from the same parent		
46,XY,upd(16)mat	associated with some cases of IUGR	both of the chr16 pair are derived from the mother
Mosaic: more than one karyotype in the same individual (derived from one zygote)		
mos 45,X/46,XX	Turner syndrome mosaicism	two subsets of cells from same zygote with different karyotypes
mos 46,XX ^{SRY+} /45,X ^{SRY+}	ovotesticular disorder of sexual development	two subsets of cells, both containing male sex determining region (SRY), on the X chromosome, one subset mosaic for Turner syndrome
Chimera: more than one karyotype in the same individual (derived from more than one zygote)		
chi 46,XX/46,XY	female/male chimera	two subsets of cells from two zygotes with different karyotypes

FISH: for rapid analysis (interphase) or confirmation (metaphase) of aneuploidy

Interphase (not usually reported, other than verbally):

nuc ish(D21S259/D21S341/D21S342)	Trisomy 21	three fluorescent signals are detected for chr21 in the same cell
----------------------------------	------------	---

Metaphase (clinical FISH, reported with karyotype):

46,XX,ish del(22)(q11.2q11.2)(TUPLE1-)	22q11.2 deletion (Di George) syndrome	absence of signal on the long arm (q) for region 1, band 1, sub-band 1 of chr22, confirmed by FISH
--	---------------------------------------	--

N.B. Cytogeneticists will not normally refer to a location according to region, band, or sub-band; it is used here only to illustrate the systematic approach of defining chromosomal coordinates in a vertical, branch-like manner. In normal communication (written and verbal) a chromosomal coordinate is likely to just be referred to as a band or simply the arm and number; e.g., “band p22.3” or just “p22.3.” (chr = chromosome)

familial disorder where the change has already been detected in another family member.

3.6.6 Restriction Fragment Analysis

This technique relies on cutting enzymes (“restriction enzymes”) that cleave double-stranded DNA molecules at specific sequences. Recognition sites are usually short (4–8 bp; e.g., the enzyme EcoRI only cuts DNA at sites with the sequence GAATTC) and their frequency—i.e., number of times they cut—is often characteristic in a particular gene. If sequence changes occur in these recognition sites, it changes

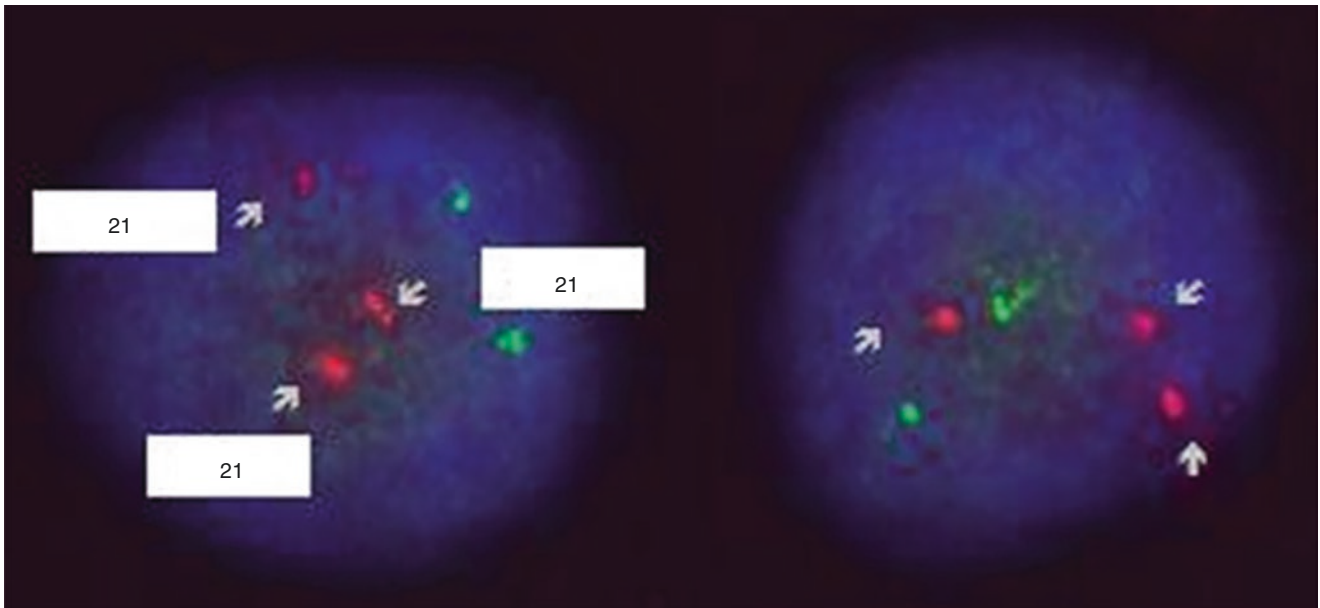


Fig. 3.10 Fluorescent *In Situ* Hybridisation (FISH) of autosomes. Trisomy 21 indicated by three red fluorescently labeled copies of chromosome 21 (arrows) in two adjacent cells. Figure courtesy of Ms. R. Hutchinson, SA Pathology, Australia



Fig. 3.11 Fluorescent *In Situ* Hybridisation (FISH) of sex chromosomes. Male sex indicated by one copy each of the X (green fluorescence) and Y (red fluorescence) chromosomes (arrows) in two adjacent cells. Figure courtesy of Ms. R. Hutchinson, SA Pathology, Australia

the number of times the restriction enzyme cuts. Ultimately this leads to a difference in the number and size of fragments of DNA when separated by electrophoresis, giving a different banding pattern, called restriction fragment length polymorphism (RFLP). Amplification fragment length polymorphism (AFLP) relies on generation of amplified polymerase chain reaction (PCR) products after restriction enzyme cutting of DNA, followed by ligation of specific PCR primers to the cut

fragments. This enables only cut fragments to be subsequently amplified in a PCR reaction. The principle of generating a range of different sized fragments that characterise presence or absence of a variant is however, overall the same as for RFLP.

The power of PCR (see later) in tandem with restriction fragment analysis, in a technique called cleaved amplified polymorphic sequence (CAPS), is more commonly utilised. Initially

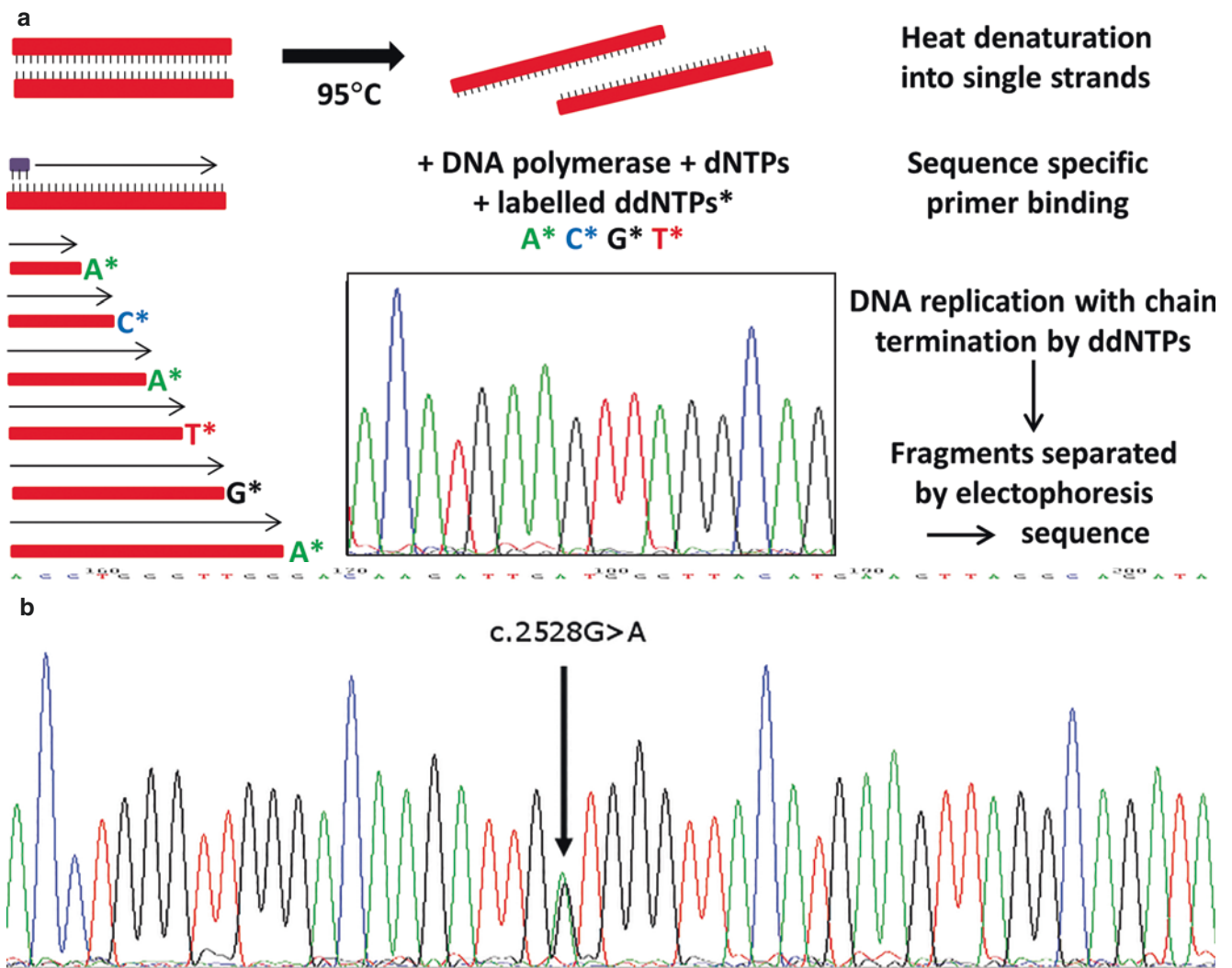


Fig. 3.12 DNA (Sanger) sequencing. **(a)** Method: Sanger sequencing utilises labeled dideoxynucleotides (ddNTPs) to terminate chains of replicating DNA, initiated by a sequence specific primer (purple). This generates many DNA fragments that differ in size by only 1bp, with their last incorporated nucleotide labeled. Separating these fragments according to size by electrophoresis, allows a profile of the last incorporated nucleotide to be determined alongside the fragment just 1bp shorter than it. A linear harvesting of DNA sequence data from a barcode-like readout of adjacent fragments is thus possible. This pro-

cess was made much easier by automation of DNA sequencing, using fluorescently labeled ddNTPs, capillary electrophoresis, and software-based sequence analysis. **(b)** Example of a Sanger sequencing readout: DNA sequencing spectra (capillary electrophoresis) indicating a heterozygous variant G>A (black arrow). At this position there are two peaks of similar height—green (A) and black (G)—indicating presence of both the normal sequence (GGT) on one allele and the pathogenic (mutated; GAT) sequence on the other allele (*PEX1* gene: c.2528G>A; p.G843D). **(b)** courtesy of Mr. T. Pyragius, SA Pathology, Australia

PCR is used to generate a shorter fragment from a well characterised region of interest using PCR. Restriction enzyme treatment then cuts the PCR product into separate smaller fragments according to presence or absence of a specific variant (**Fig. 3.13**).

- **RFLP/AFLP** is sometimes used to diagnose spinal muscular atrophy prenatally.

3.6.7 Linkage Analysis

The principle behind linkage analysis uses alleles that are commonly inherited together as markers for specific genes,

although they are unlikely to be the actual disease cause. These marker regions may be detected by DNA sequencing, RFLP, AFLP (see previous), PCR or Southern blotting (see later).

Historically, linkage analysis was responsible for discovery of many genes (e.g., *CFTR*); however, the increasing availability of SNP arrays, exome- and genome-wide association studies using newer technologies means use of this technique has continued to decrease other than in families where there already exists significant historical linkage data for known heritable disorders.

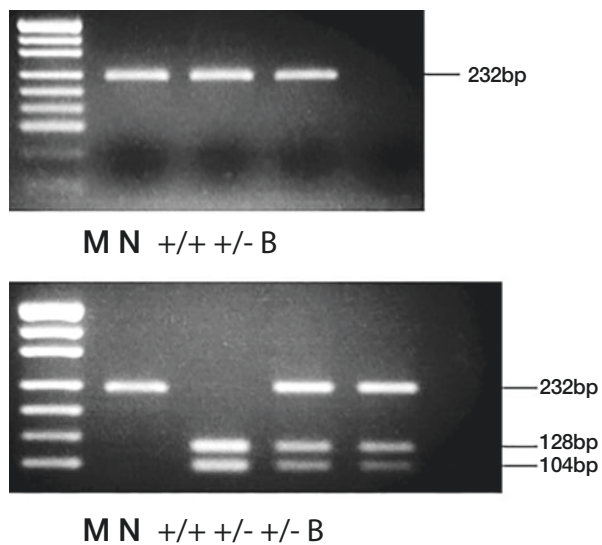


Fig. 3.13 Restriction fragment analysis of a PCR amplicon. An example of cleaved amplified polymorphic sequence (CAPS). PCR primers targeting a region of the *PMM2* gene amplify a 232bp product in all samples, visualized on agarose gel electrophoresis (upper panel). Differences in DNA sequence produce differences in the ability for restriction enzymes to cut at their specific sequence targets. Differences in the DNA fragment profile after restriction enzyme digestion are referred to as CAPS. Shown in the bottom panel is a restriction analysis-based method that detects a pathogenic variant in the *PMM2* gene, associated with the condition Congenital Disorder of Glycosylation Type 1a (CDG-1a). Restriction enzyme BtsC1 cuts only at a single nucleotide polymorphism in the amplified region of the *PMM2* gene. PCR products are cut into two smaller fragments only if this variant is present. Individuals that are heterozygous for this variant will have both the uncut (232bp) and cut fragments present (128bp and 104bp). M: molecular weight markers, N: normal, no pathogenic variant (–/–), +/-: homozygous pathogenic variant, +/-: heterozygous pathogenic variant, B: blank (no DNA) control. Figure courtesy of Mr. K. Brion, SA Pathology, Australia

3.6.8 Southern, Northern & Western Blots

As described in the historical timeline (Fig. 3.1) this technique was named after its developer, Edwin Southern, not a map direction, hence the capitalisation of “Southern.” It was the first time that the techniques of complementary hybridisation and fixation of DNA to a solid substrate after separation by electrophoresis were combined.

The same principle underlying this technique was then used for protein (Western blot) and RNA (Northern blot), a play on words from the map direction nuance. Like FISH (see previous), all of these techniques rely on labeled probe hybridising to a region of interest, but after electrophoretic separation of the sample, then immobilisation on a solid substrate (Fig. 3.14). Like FISH, Southern and Northern blotting use a complementary nucleic acid, whilst Western blotting uses an antibody to the epitope of interest as the probe. The size of a nucleic acid probe and therefore the region of its complementary binding may be small (oligonucleotide) or very large (cDNA).

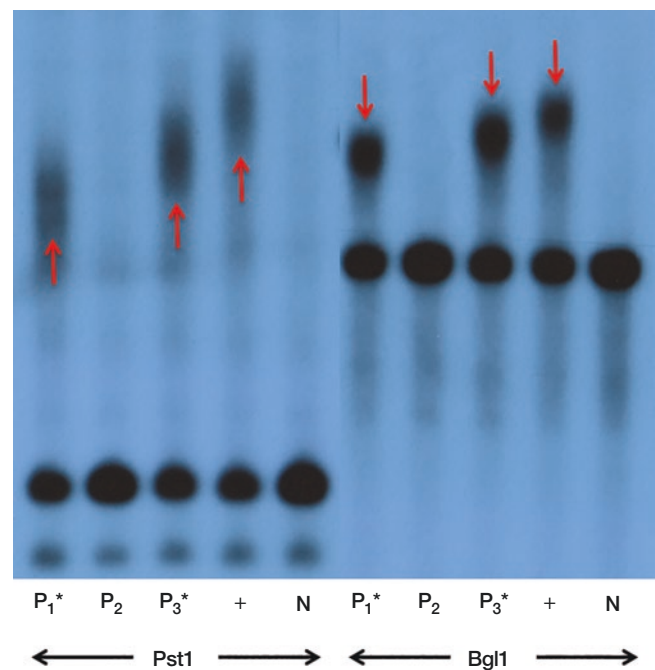


Fig. 3.14 Southern blot. DNA is cut into smaller fragments by restriction enzymes (here Pst1 and Bgl1) that only cut at specific recognition sequences, then electrophoresed on agarose gel and transferred (blotted) onto a nitrocellulose sheet. A radiolabeled piece of DNA specific to the gene or region being probed hybridises to regions containing complementary DNA (here M10M6 probe for the *DMPK* gene). Size of DNA fragments is estimated by how far they migrate from the origin during electrophoresis (larger fragments migrate more slowly, here closer to the top). Red arrows indicate restriction fragments from one allele that are greater in size than the normal range. Expansion of the number of CTG repeats in the non-coding region of the *DMPK* gene is associated with the autosomal dominant disorder myotonic dystrophy type 1 (DM1; Normal: 5–37, Pre-mutation 38–49, Mild: 50–150, Classical: 100–1500, Congenital: 1000–2000 CTG repeats). Number of CTG repeats can be determined from the size of labeled DNA fragments. P₁ = approx. 1.2–2.2 kb fragment (412–743 CTG repeats), P₂ = approx. 1.9–2.7kb fragment (629–904 CTG repeats), positive control = approx. 2.6–4.2 kb (867–1400 CTG repeats). P: patient sample, *: pathogenic CTG expansion present, +: positive control, N: normal control. Figure courtesy of Ms. R. Catford & Dr K. Friend, SA Pathology, Australia

- **Southern blot** is commonly used to determine the length of a repeat sequence in fragile X syndrome or congenital myotonic dystrophy.
- **Western blot** is a good test for HIV antibody test confirmation.

3.6.9 Polymerase Chain Reaction (PCR)

Most genetic testing technologies used today rely on amplification of identical copies of DNA region/s of interest from relatively small amounts of starting material.

Polymerase chain reaction (PCR) is the technology that underpins this amplification. Invented by Nobel Laureate

Kary Mullis in the mid-1980s, it essentially harnesses the inbuilt machinery of DNA replication, revolutionising molecular biology to this day. Relying on variability in the strength of DNA binding to its complementary nucleotide sequence at different temperatures, PCR utilises tightly controlled automated temperature cycling and a special heat-tolerant form of DNA polymerase (*Taq*—isolated from the thermophilic bacterium *Thermus aquaticus*) to rapidly and exponentially replicate specific sequences of DNA.

PCR consists of three phases repeated many times to exponentially amplify the target (Fig. 3.15a). The amplified product is referred to as an amplicon. Amplification of nucleic acids by PCR has many variations. Three of the most important variations (GAP-PCR, long-range PCR, and MLPA) are discussed below. However, direct differences in the size of PCR amplicons alone can be used to detect well-characterised genetic variations (Fig. 3.15b, c). Sequencing and/or MLPA (below) are often used as subsequent confirmatory methods following PCR positive results. **Triplet repeat primed PCR (TP-PCR)** is another common PCR method, frequently used to follow up or replace the primary PCR method for Fragile X syndrome testing detailed in Fig. 3.15c. A useful explanatory video about TP-PCR is available here [35].

- **PCR-based** amplification is used, at some stage, in most genetic tests. It is often confused as “the” genetic test itself but invariably its primary use is to amplify enough DNA to do “the” test.

3.6.9.1 Gap PCR

This form of PCR relies on well-characterised deletions, resulting in normally distant sequences being found very close together, so that primers to those sequences are close enough to now be successfully amplified by PCR.

- Gap PCR is a good test for detecting hemoglobinopathies, such as Hb Barts in hydrops fetalis (Fig. 3.16a).

3.6.9.2 Long Range PCR (L-PCR)

In standard PCR, there is an underlying error rate for misincorporation of nucleotides (of the order of once per 10,000–100,000 nucleotides). *Taq* polymerase stalls to correct these errors. The longer a DNA strand, the more likely there will be errors and the efficiency of replication compromised by *Taq* stalling to repair them. This sets a practical limit on the length of DNA able to be amplified using standard PCR to a few thousand base pairs.

Incorporation of a proofreading enzyme into a PCR mix helps to iron out these errors earlier, allowing *Taq* and/or other DNA polymerases to produce longer amplified products of the order of tens of kilobases. This is called **long range PCR (L-PCR)**. It is used in applications where ampli-

fication with good fidelity over larger stretches of DNA is required; e.g., complete mitochondrial DNA sequence.

- **L-PCR** is a good test for detecting Incontinentia Pigmenti (Fig. 3.16b).

3.6.9.3 Multiplex Ligation-Dependent Probe Amplification (MLPA)

A PCR technique, MLPA is used to detect copy number variations (deletions or duplications) in genes (see discussion in array, Sect. 3.6.13). It uses one primer pair to amplify PCR products from multiple regions in the one reaction. Each region produces a uniquely sized amplicon due to differences in the size of stuffer and gene specific regions of hybridisation probes, but flanked by the same common primer sequence that will amplify in PCR. If an exon or part of it is missing, then that region will not be amplified in an MLPA reaction. The specificity of this technique lies in the fact that amplification will only be successful if sequence is identified where the probes sit adjacent to each other, so that a small gap between them can be filled in by the enzyme ligase, that then allows PCR amplification to proceed (Fig. 3.17a). Examples of its application are given in references [36] and [37].

- **MLPA** is a good test for Duchenne’s Muscular Dystrophy (DMD) and microdeletion syndromes; e.g., 22q11.2 deletion (velocardiofacial or DiGeorge) syndrome, and Spinal Muscular Atrophy (SMA) (Fig. 3.17b–d).

3.6.10 Matrix-assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) Mass Spectrometry

This technology has been adapted for use in identifying many biomolecules. The most common use in genetic testing is looking for well-characterised single nucleotide variants in DNA. It begins with a PCR amplification step to generate starting material specific for the gene of interest. In a second separate reaction, there is extension of one single nucleotide onto the amplified product using nucleotides modified to have a specific mass (Fig. 3.18a). Resulting samples are purified then spotted by a robotic device in nanoliter quantities onto a silica chip, much like in microarrays. This gives the advantage of very high density throughput so that many samples can be assessed in tandem. Firing of a finely controlled laser precisely onto each individual spot, rapidly and sequentially converts it into ionized plasma for passing through a connected mass spectrometer, creating a mass particle profile for each sample. Variants and normal sequence have characteristic mass particle signatures, assessed and called automatically in software.

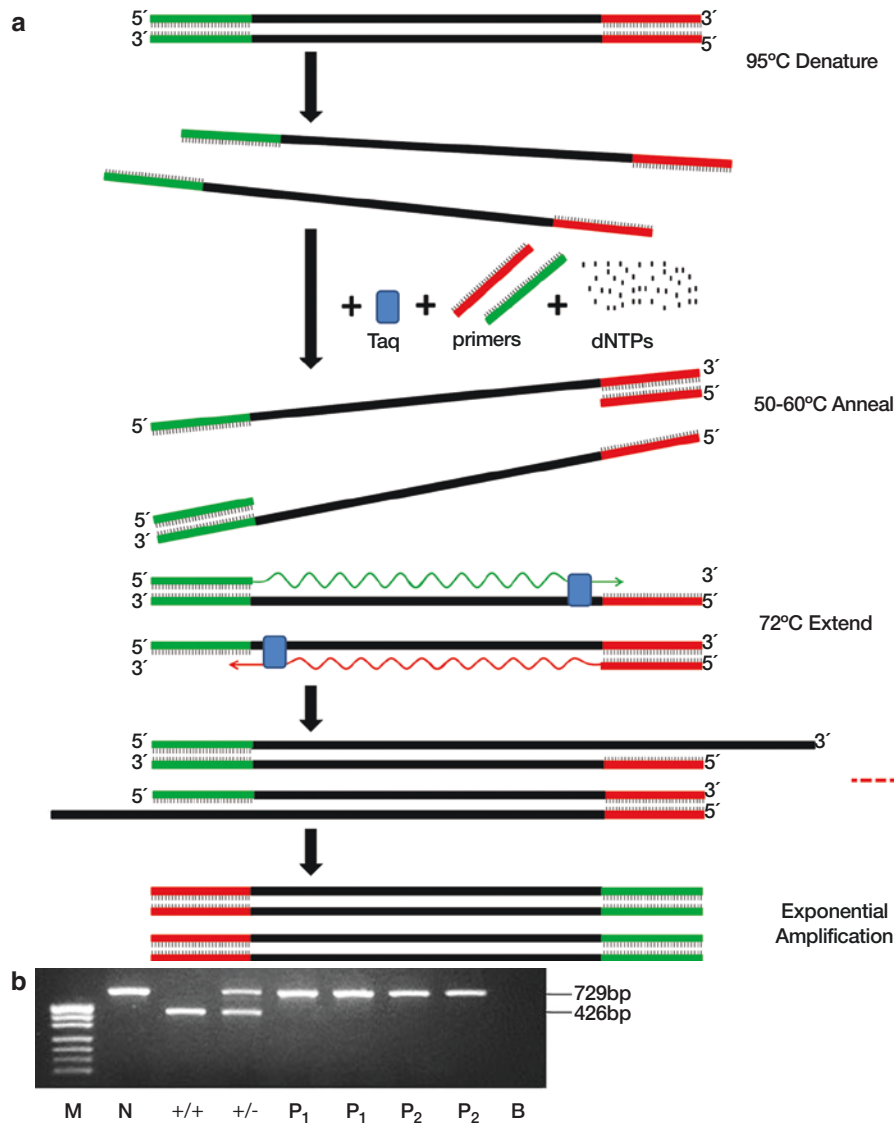


Fig. 3.15 Polymerase chain reaction (PCR). (a) PCR amplification of DNA. DNA replication requires a DNA polymerase, primers to initiate the region of replication, and nucleotides (dNTPs). In PCR, the steps of denaturation, annealing of primers, and extension of sequence from the primers happen at tightly controlled temperatures. *Thermus aquaticus* (*Taq*) polymerase and other DNA polymerases that can perform and survive at relatively high temperature allow rapid cycling of these steps to produce an exponential amplification of target DNA. (b) PCR analysis of a gene deletion (agarose gel electrophoresis). Differences in the length of DNA of a PCR amplified product can indicate deletions or duplications to that region. Differences in the profile of PCR amplified products are visualised by electrophoresis, separating amplicons according to size. Shown here on agarose gel electrophoresis, is a 203bp decrease in size of the PCR amplified product targeting a pathogenic deletion in the *CLN3* gene (associated with Ceroid Lipofuscinosis, Neuronal, type 3 [Batten Disease]). Individuals that are heterozygous for this variant will produce amplified PCR products both with (426bp) and without (729bp) the deletion. M: molecular weight markers; N:

normal control (this pathogenic variant absent), $-/-$; 729bp, $+/+$: homozygous pathogenic variant on both alleles; 426bp, $+/-$: heterozygous pathogenic variant on one allele (426bp and 729bp); $P_{1,2}$: two normal patient samples (this pathogenic variant absent); 729bp, B: blank (no DNA) control. Figure 3.15b courtesy of Mr. K. Brion, SA Pathology, Australia. (c) PCR analysis of CGG repeats in Fragile X syndrome (capillary electrophoresis). PCR primers target the CGG repeat region of the *FMR1* gene on the X-chromosome, associated with Fragile X syndrome (FXS). Capillary electrophoresis differentiates PCR products according to size, allowing the number of CGG repeats to be determined (normal 5–44, grey zone 45–54, pre-mutation 55–200, FXS >200). Shown here are 30 CGG repeats in a male (only one X-chromosome; upper panel), an unaffected female (22 & 29 repeats; middle panel), and a female normal on one allele and a pre-mutation on the other allele (29 & 54 repeats; lower panel). This method will not detect deletion or missense variant causes for FXS. Figure 3.15c courtesy of Dr. K. Friend, SA Pathology, Australia

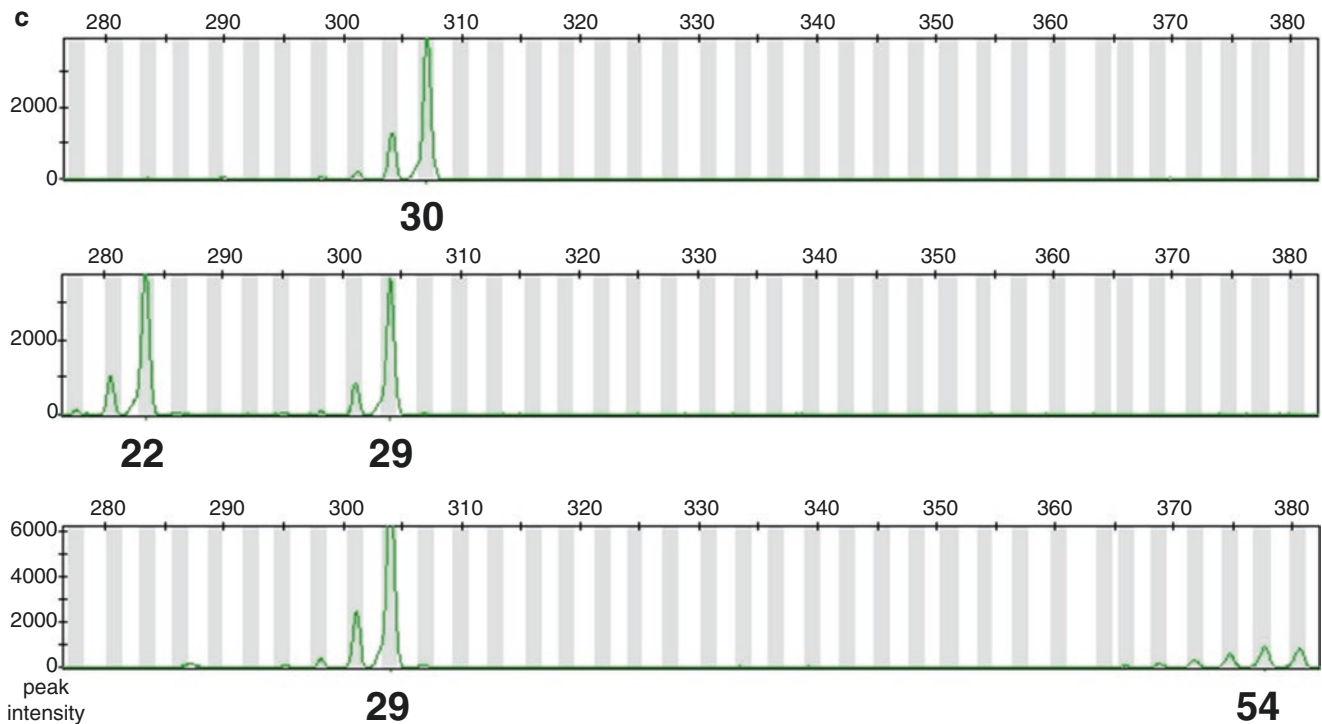


Fig. 3.15 (continued)

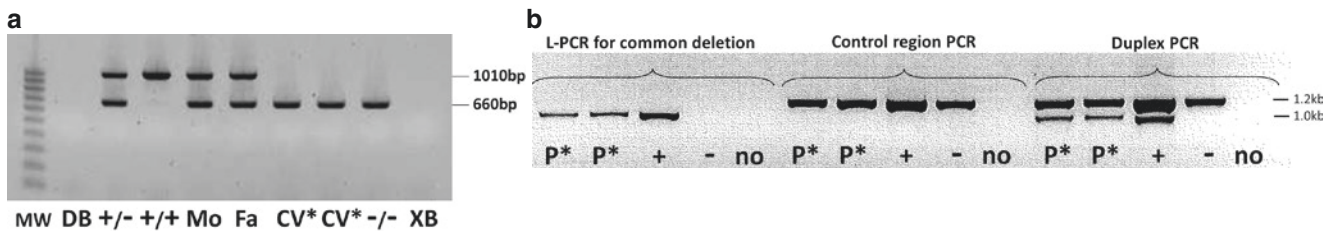


Fig. 3.16 (a) Gap PCR analysis of alpha thalassemia. Two closely located genes encode alpha globin (*HBA1* & *HBA2*; both on chr16p13.3). Common deletions in alpha globin can be detected by Gap PCR. Pathogenic deletions in these genes result in various forms of alpha thalassemia, depending on the number of functional alpha globin alleles (normal = 4, one from each gene on each allele). Homozygous deletions on both alleles for both genes result in no functional alpha globin protein (Hb Barts, causing fetal demise from hydrops fetalis). Shown here is one gene deletion found predominantly in those of South-east Asian ethnicity that spans both alpha globin encoding genes. This deletion involves the removal of about 19.4Kb of DNA—including the $\psi\alpha 2$, $\psi\alpha 1$, $\alpha 2$, $\alpha 1$ and $\theta 1$ globin genes. Gap PCR produces a smaller sized amplicon if a deletion is present—wild type (no deletion): 1010bp; heterozygous deletion: both 1010bp & 660bp; homozygous deletion: only smaller 660bp amplicon. Parents are seen to both be heterozygous for this deletion, with their fetus affected (homozygous; Hb Barts). MW: molecular weight markers, DB: DNA blank control for PCR; +/-: heterozygous deletion control; +/+ : wild type (no deletions) control; Mo: mother; Fa: father; CV*: fetal chorionic villous sample;

-/-: homozygous deletion control, XB: DNA extraction blank control. Figure courtesy of Dr. K. Simons & Dr C. Nicholls, SA Pathology, Australia. (b) Long-range PCR (L-PCR). Conventional PCR utilises thermostable *Taq* polymerase for amplification of DNA targets. *Taq* allows rapid amplification but has limitations on the maximum size of the amplified product. Long-range PCR utilises high-fidelity DNA polymerases with proofreading ability to allow amplification of very long DNA fragments (up to 40 kb). The most common deletion in the *IKBKKG* gene (associated with Incontinentia Pigmenti) is 11.7 kb (spanning exons 4–10). The markedly decreased size of the amplification product containing this deletion (1.0 kb) is shown here from an L-PCR reaction (left). Samples without the deletion will produce a much larger amplification product (~13 kb; not shown), and no smaller 1.0 kb amplification product. An unrelated, ubiquitous region of DNA acts as an amplification reaction control (middle). Duplex PCR combines both the *IKBKKG* gene and control primers in the same tube to control for any differences in amplification efficiency. P*: patient with deletion, +: positive control, -: negative control, no: no DNA control. Figure courtesy of Dr. K. Friend, SA Pathology, Australia

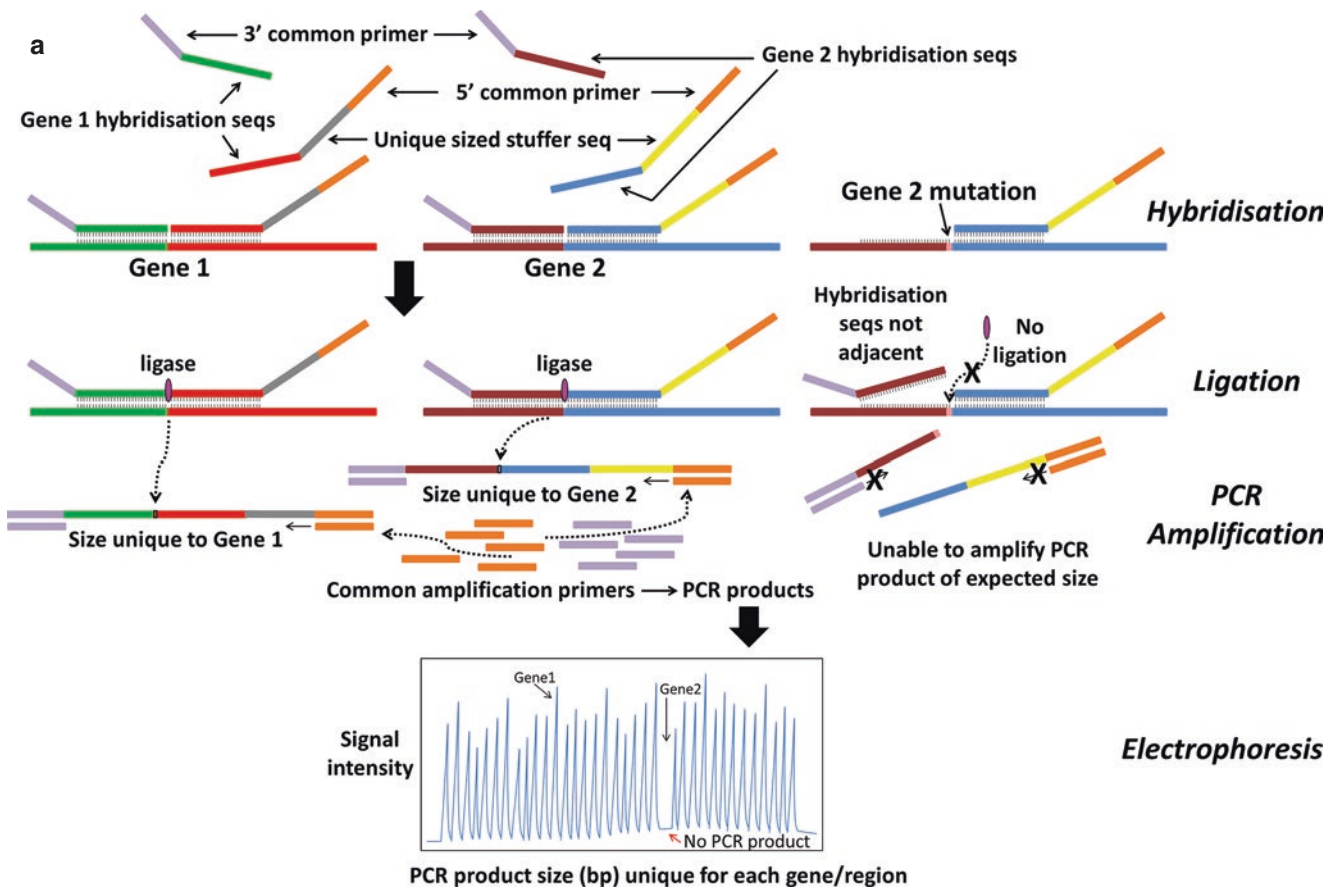


Fig. 3.17 (a) Multiplex ligation-dependent probe amplification (MLPA). Many changes in multiple genes (or even different regions of the same gene) can be tested in the same single reaction tube (multiplexed). Common PCR primer sequences (violet & orange) flank hybridisation sequences (seqs; Gene 1: green & red; Gene 2: brown & blue) specific for individual gene changes. A ligation step after hybridisation will only occur if both hybridisation sequences (probes) for that region completely hybridise, so that they lie adjacent to each other. DNA ligase (pink) is then able to fill in the gap between these adjacent hybridised probes (black rectangle). This allows the common primers to amplify a PCR product from any region that has completely hybridised to their hybridisation probes. Any change in sequence in sample DNA (Gene 2: light pink; far right) will not allow complete hybridisation of the hybridisation probes resulting in failure of the ligation step and subsequent failure to amplify a PCR product for that region. The combination of stuffer sequence (Gene 1: grey; Gene 2: yellow) and hybridisation sequence is designed so that each gene region produces a uniquely sized PCR amplification product, when analysed on capillary electrophoresis (bottom). In this way many genes (or regions from the same gene) can be analysed simultaneously in the one reaction. Similar to CGH array (Fig. 3.21), comparison of copy number variation (CNV) between a control and test sample analysed by MLPA indicates if there have been deletions or duplications of DNA, but over much smaller regions (50–70bp) than possible with CGH array. (b, c) MLPA analysis of microdeletion syndromes. Analysis of 20 microdeletion syndromes

simultaneously, using a commercial MLPA kit (P0245; MRC Holland). PCR amplified products are separated by size on capillary electrophoresis. Amplified product size and relative quantity from a test sample (blue trace) is compared to a normal control (red trace) to determine copy number variations (CNVs) in amplified regions. (b) Heterozygous deletions of two probes (256 & 335bp) within the region associated with autosomal dominant neurofibromatosis type 1 (NF1) microdeletion syndrome are indicated by a reduction of blue trace to approximately half of the red trace peak height for the size of amplified product expected for this region (arrows). (c) The same data can be presented as a peak ratio to more clearly delineate CNVs. Peak ratios of approximately 1 indicate no CNV (green boxes). Peak ratios greater than 1.25 or less than 0.75 (green horizontal lines) suggest marked CNV; i.e., duplications and deletions, respectively. Heterozygous deletion is indicated by a peak ratio of approximately 0.5, corresponding to an expected decrease in amplified product by one half if it has been deleted from one of a pair of alleles. The probes deleted are at chr17q11.2 within exons 12 and 20 of the *NF1* gene (red boxes, arrowed). (d) MLPA analysis for Spinal Muscular Atrophy. MLPA peak ratio analysis indicates homozygous deletion of regions of the *SMN1* gene (peak ratio = zero, indicating no amplified product detected in this region i.e. deletion from both alleles). Deletions of two probes to exons 7 & 8 (182 & 218bp, respectively; red boxes, arrowed) of the *SMN1* gene are associated with the autosomal recessive condition Spinal Muscular Atrophy (SMA). (b–d) courtesy of Dr. K. Friend, SA Pathology, Australia

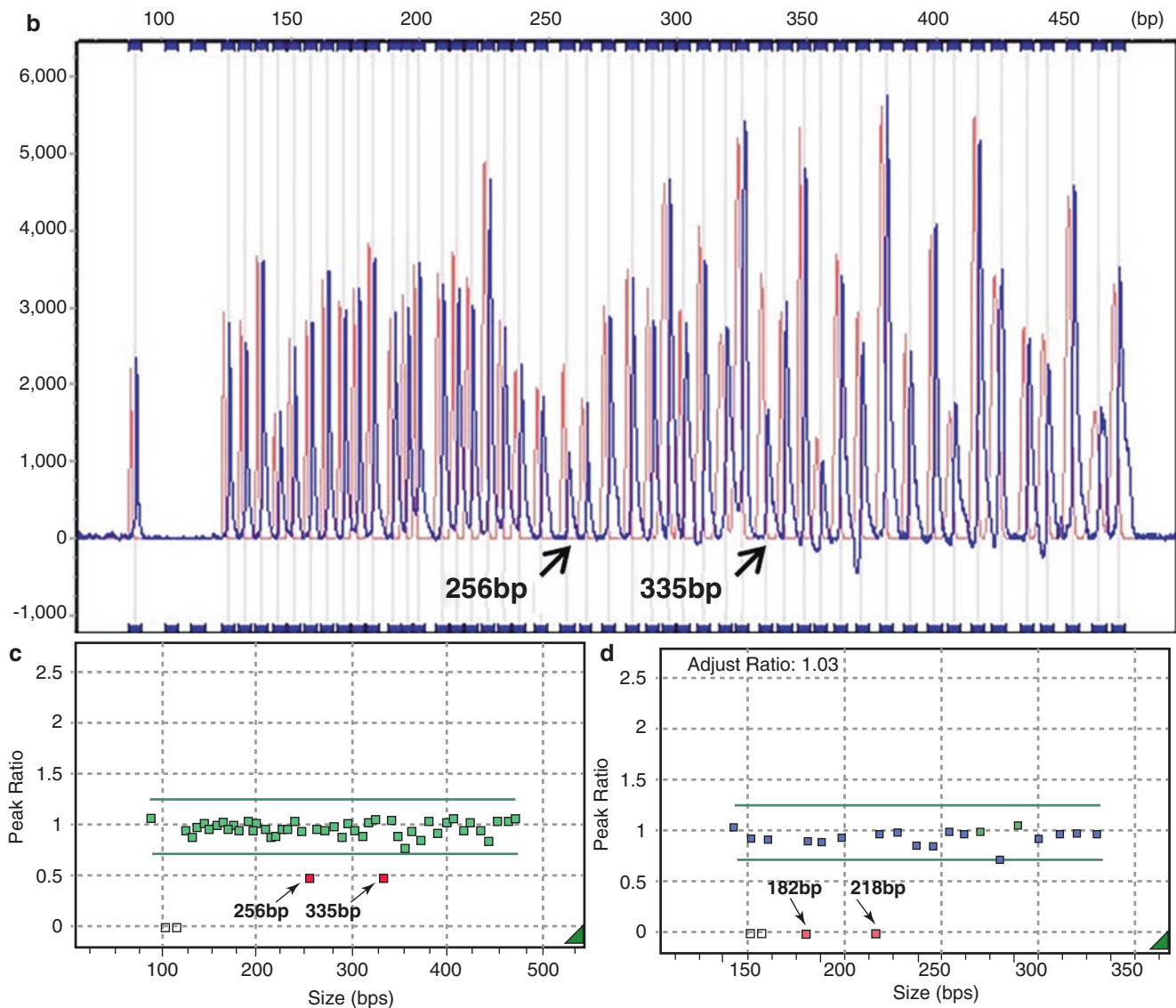


Fig. 3.17 (continued)

A great advantage of this technique is that several different variants can be assessed in the one tube, as long as each amplified, mass-labeled product has a unique mass compared to other products in the same tube. This multiplexing of both an increased number of individual samples in the one run, as well as the number of variants that can be assessed simultaneously, has markedly increased the power and decreased the cost of this technology for variant screening in conditions with high carrier prevalence (e.g., cystic fibrosis; **Fig. 3.18b**).

- **MALDI-TOF** is a good test for many of the common variants found in the cystic fibrosis gene (CFTR).

3.6.11 Mini/Micro Satellite Repeats

Satellite repeats are short sequences of DNA repeated next to each other (called **variable number tandem repeats, VNTRs**) at specific sites throughout the genome, often in non-coding regions. Microsatellites are repeats of 2–6 bp (**short tandem repeats; STR**), while minisatellites are longer VNTRs of 10–60 bp. The number of times the sequence is repeated in tandem is highly variable between individuals. PCR-based techniques can be used to amplify these repeat regions, and the number of times a VNTR is repeated can be determined from their size on electrophoresis. The number

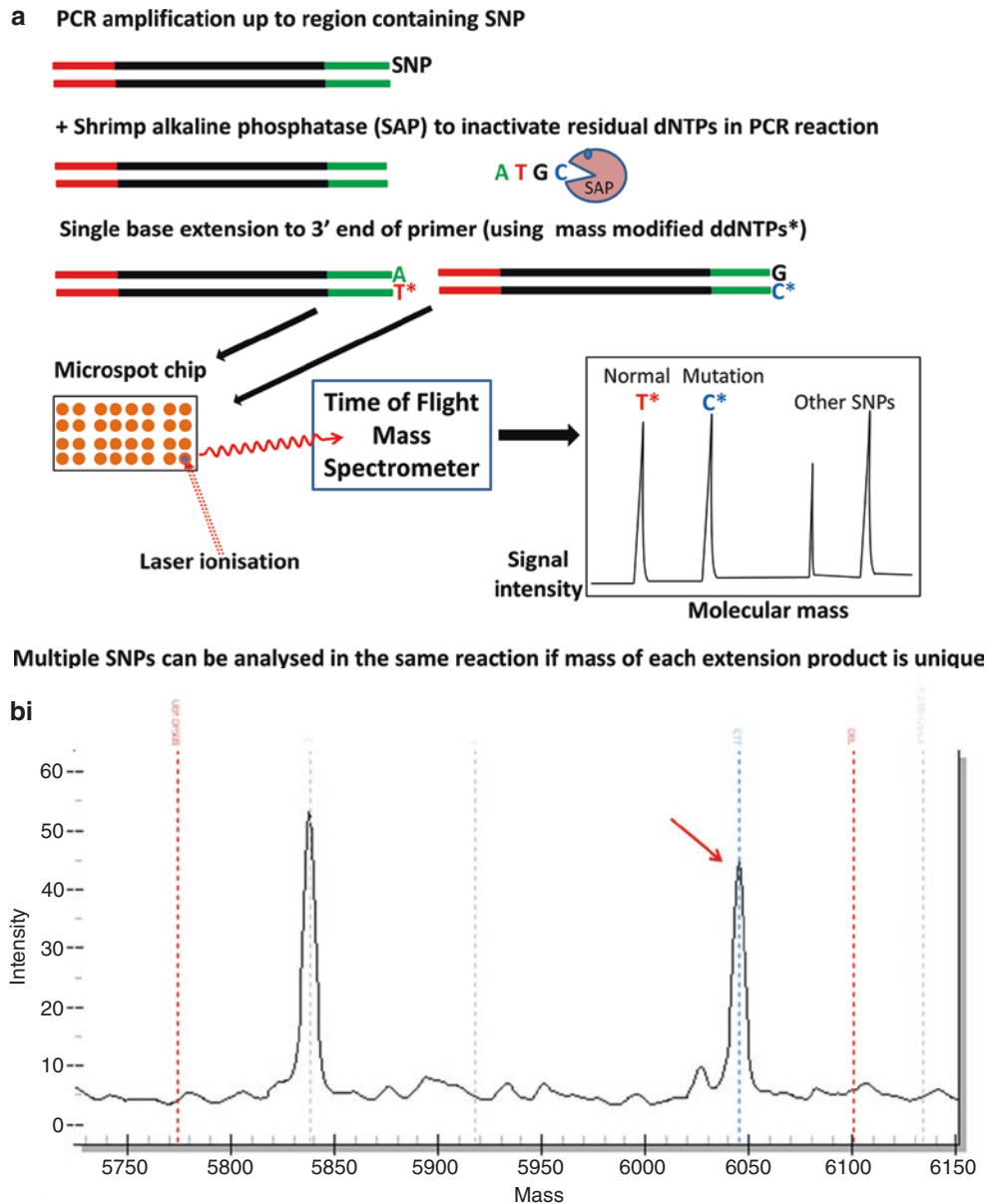


Fig. 3.18 Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. **(a)** Method: This test relies on incorporation of mass modified ddNTPs to produce a specific mass spectrometric signature. It allows multiple single nucleotide polymorphisms (SNPs) to be assessed in the same reaction. PCR generates an amplified product next to the SNP of interest. Clean up of the PCR reaction with shrimp alkaline phosphatase (SAP) removes any remaining dNTPs so that they will not interfere with the subsequent single base extension step. Use of chain terminating ddNTPs with a modified mass, ensures that only one single base extension will occur and that the extension product will have a unique mass based on the nucleotide

incorporated. Micro-spotting onto a silicon chip, followed by laser ionisation feeding directly into a mass spectrometer allow rapid, automated analysis both of many SNPs in the one reaction and multiple samples spotted at high density onto the same microchip. **(b)** *CFTR* gene c.1521_1523delCTT (p.F508del) pathogenic variant in cystic fibrosis: (i) Normal (CTT intact; red arrow). (ii) Heterozygous (both CTT and deletion (DEL) with similar peak heights; red & green arrows, respectively). (iii) Homozygous (deletion (DEL) peak only with no CTT peak; red & green arrows, respectively). **(b)** courtesy of Mr. T. Pyragius, SA Pathology, Australia

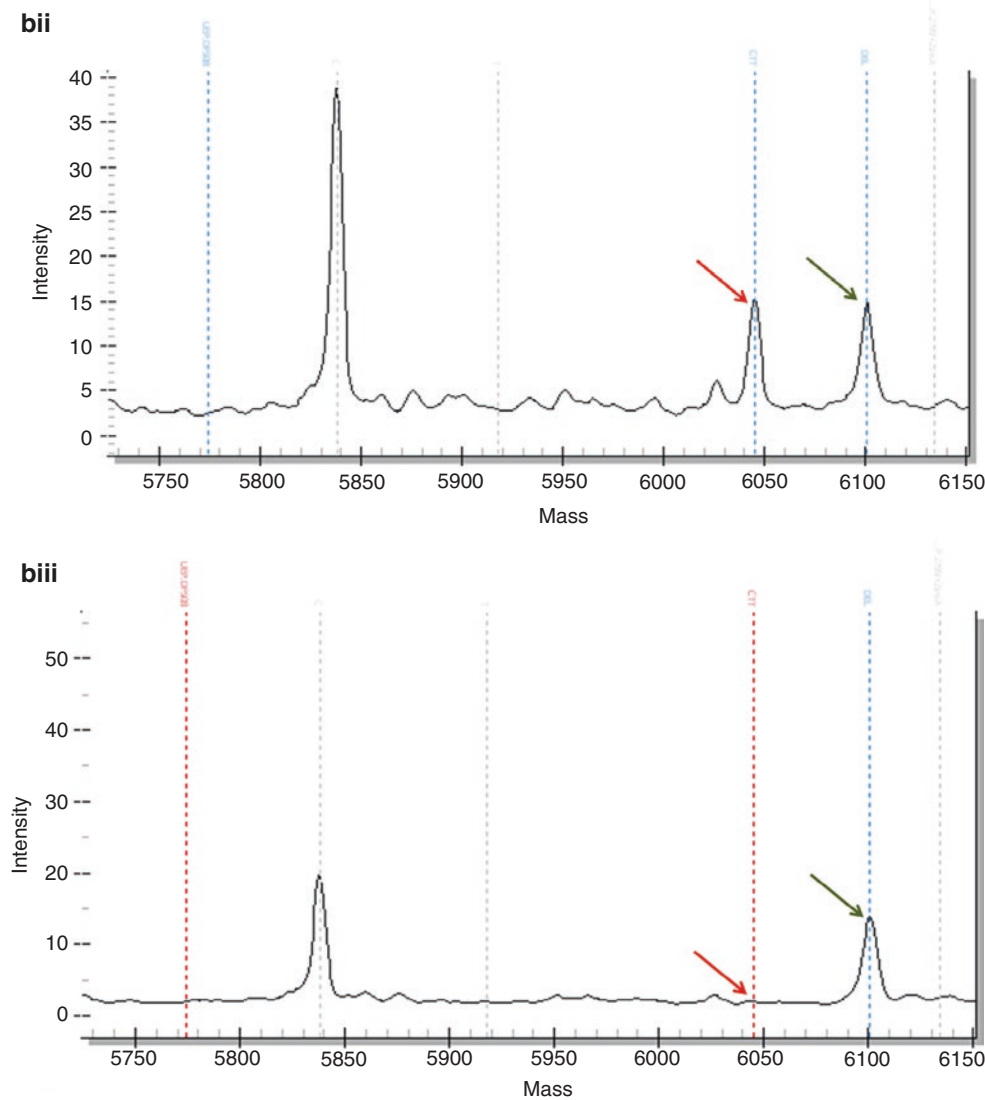


Fig. 3.18 (continued)

of repeats in several VNTRs will be characteristic for each individual and forms the basis of DNA fingerprinting.

As half of our genome is inherited from each parent, we also get half of our satellite repeat patterns from each parent. Therefore, this technique is useful for parentage analysis; e.g., in paternity cases or in determining the level of maternal cell contamination in a fetal sample (**Fig. 3.19**).

Satellite repeat results are generally presented as electrophoresis spectra indicating the number of repeats found in a range of different VNTRs in the same individual.

The technique is useful in molar pregnancy testing where mis-expression of imprinted genes leads to a complete or partial hydatiform mole. Determining parental origin of the imprinted genes is helpful for proper classification, determining likely pathology and most appropriate management [38].

- **Satellite repeat marker analysis** is a good test for maternal cell contamination, molar pregnancy, forensic identification and paternity testing.

3.6.12 CpG Methylation

Data on CpG sites where cytosine is methylated to produce 5-methyl cytosine is used to determine regions of epigenetic gene silencing. Conversely, CpG hypomethylation at known MVPs indicate increases in gene expression at these sites (see epigenetics, Sect. 3.4).

Many CpG methylation tests rely on treatment of genomic DNA with bisulphite (alkylation) (**Fig. 3.20a**).

High resolution melting analysis, methylation-specific PCR, standard PCR followed by MALDI-TOF, RFLP or

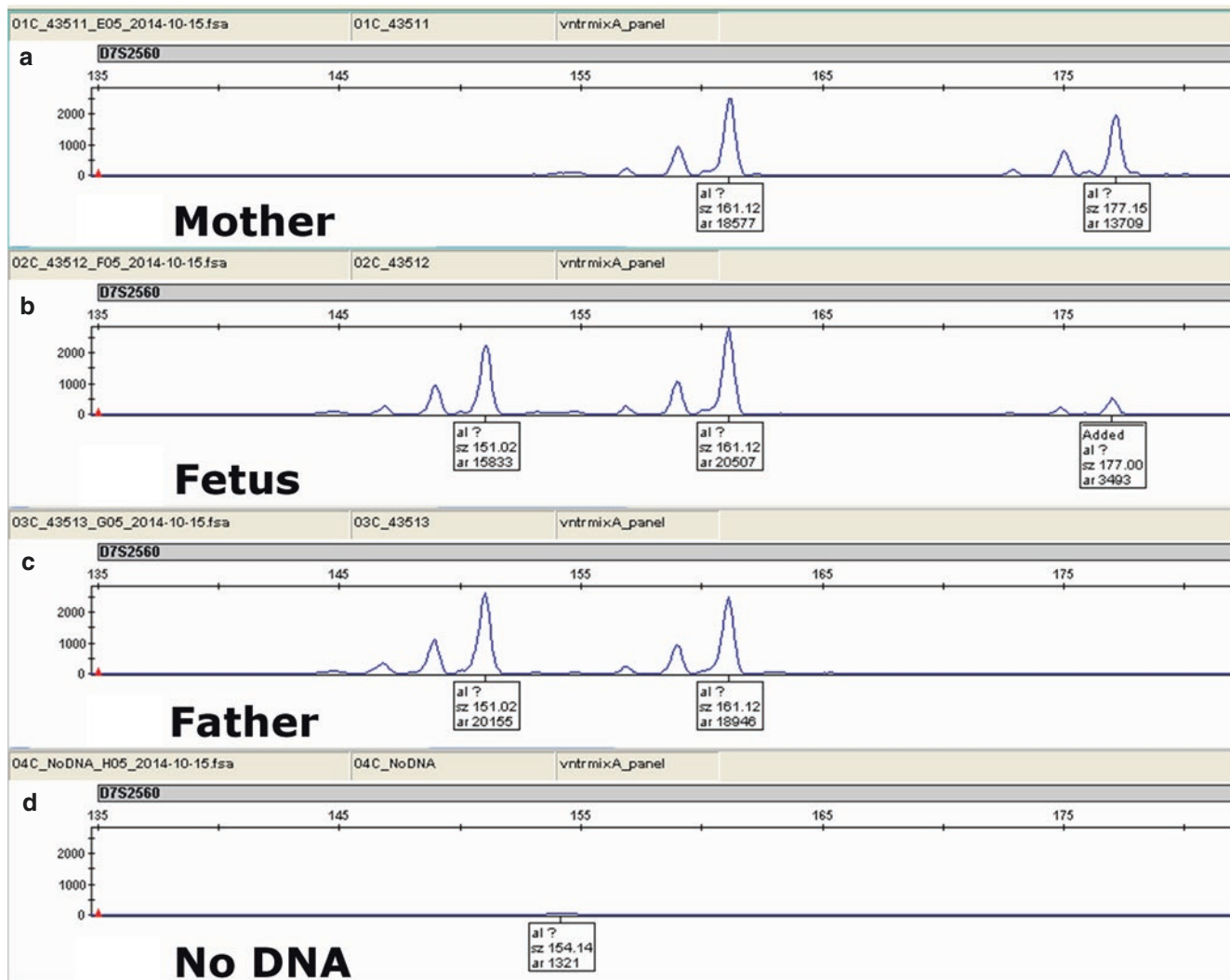


Fig. 3.19 Satellite repeat based DNA fingerprinting for detecting maternal cell contamination. Sets of DNA microsatellite markers are amplified for maternal, paternal, and prenatal (fetal) samples. The distance between microsatellite markers and therefore size of amplified products will differ between individuals, acting as a unique DNA fingerprint. Peaks are separated according to molecular size. Although for this set of markers, the mother and father share a common 161bp marker on one allele; the father (c) has a 151bp marker and the mother (a) a 177bp marker on the other allele. The fetus (b) should only inherit

one allele from each parent; however, there are three peaks present (151, 161 & 177bp) indicating that the sample is contaminated with some maternal tissue. A no DNA control (d) does not produce any amplified products. Unrelated individuals may share some common microsatellite markers on one allele; however, using many sets of markers ensures a unique profile for each individual. A similar strategy is used for forensic DNA fingerprinting and determining parentage. Figure courtesy of Ms. R. Catford & Dr K. Friend, SA Pathology, Australia

sequencing and methylation arrays are all techniques used to determine methylated CpG sites that are resistant to bisulphite treatment. Other methylation tests rely on utilisation of the differential cutting ability of methylation sensitive restriction enzymes e.g. **methylation-specific MLPA (MS-MLPA)** for imprinting disorders [39, 40]. The technique chosen depends on the type of lesion and total length of coverage required.

- **CpG methylation analysis** is useful in assessing diseases related to genomic imprinting such as Angelman & Prader-Willi syndromes (**Fig. 3.20b**) (see epigenetics, Sect. 3.4).

3.6.13 Cytogenetic Microarray (CGH & SNP Array)

Microarrays are an important part of the fetal diagnostic process. They can indicate differences in chromosome structure at a higher resolution than attainable by karyotyping (microarray can detect deletions as small as 10 kb). The American College of Medical Genetics (ACMG) review of clinical use of array-based technologies recommend them as a first-tier test for investigating developmental delay/intellectual disability, multiple congenital abnormalities, and autism spectrum disorders [41]. They cite evidence from

a C*GGCTGC*GG - PATERNAL
 C GGCTGC GG - MATERNAL
 U GGCTGU GG - SODIUM BISULPHITE modification
 A CCGACA CC - PATERNAL primer
 G CCGACG CC - MATERNAL primer

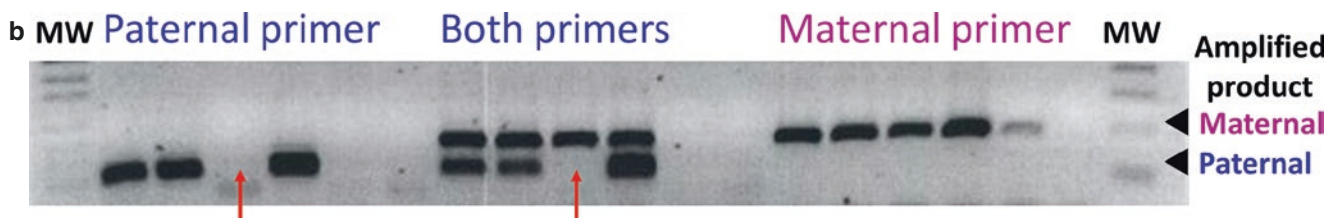


Fig. 3.20 Methylation PCR. (a) Paternal and maternal alleles have different methylation patterns (imprinting). Bisulfite alkylation of cytosine to uracil (U) does not occur at CpG methylated sites (C*). This enables design of primers that will only bind to non-methylated regions after bisulfite alkylation. (b) This technique can be employed to determine imprinting patterns important in conditions such as Angelman & Prader-Willi syndromes (PWS), as well as other epigenetic modifica-

tions. Following bisulfite alkylation, PCR is conducted with primers specific for maternal non-alkylation and paternal alkylation products (a). Agarose gel electrophoresis indicates the absence of the paternal amplification product (red arrows) and presence of the maternal amplification product, consistent with the paternal imprinting pattern found in PWS. MW: molecular weight markers. Figures courtesy of Dr. K. Friend, SA Pathology, Australia

large cohort studies estimating between 10 and 20% improved diagnostic yield compared to karyotyping.

This technology relies on robotic workstations to spot well-characterised DNA fragments at very high density in specific order onto silicon microchips (**microarrays**). The entire genome of an individual, fragmented into smaller pieces, can then be applied to the chip where it will hybridise to its complementary sequence at a specific location, already mapped on the chip.

Differences in hybridisation patterns between a test and reference genome indicate **copy number variation (CNV)**; i.e., differences in the number of times one of the smaller fragments of DNA is present within the genome.

The hybridisation component of both **CGH (comparative genomic hybridisation)** and **SNP (single nucleotide polymorphism)** microarray techniques is analogous to 100,000s of FISH hybridisation reactions being run in parallel next to each other on the one chip. Automated microscopic imaging and analysis is then used to determine fluorescent intensity at each spot, to assess differences in hybridisation compared to a reference (“normal”) genome.

CGH relies on a test genome being fluorescently labeled a different color (green) to the reference genome (red). The two samples are then combined and hybridised to the microarray chip together. Identical sequences will hybridise to the same locations on the chip. Differences in signal intensity between different spots on the chip are easily evident when imaged i.e. equal signal intensities will result in a yellow spot (combination of red & green). Spots that are more green

or red, indicate copy number variations between the test and reference genome (**Fig. 3.21a**).

SNP microarray, in contrast, utilises hundreds of thousands of specific oligonucleotides, generated to cover the entire genome, with inclusion of many containing single nucleotide polymorphisms (SNPs), located in regions known to be commonly associated with copy number disorders. This can increase the resolution and precision of the sequences found to have CNVs. SNP array does not use a reference genome in the same hybridisation reaction, rather it compares the test genome fluorescent hybridisation signal to an archived, well characterised reference genome through software (**Fig. 3.22a**). Current commercially available SNP arrays use at least 850,000 unique oligonucleotides on their microchips.

Virtual karyotypes can be constructed in software from both CGH and SNP arrays (**Figs. 3.21b, c and 3.22a**) as the chromosomal location of probes used is well characterised, with coverage across all chromosomes.

Identified CNV regions may contain multiple candidate genes that could be causative for disease. Online tools such as the UCSC Genome Browser (**Fig. 3.22b**) and OMIM [19], plus biomedical literature searches are used to help interrogate array results. Recent technical standards for CNV pathogenicity classification [42], introduced a quantitative, evidence-based scoring framework, similar to that widely used in sequence variant classification (five tiers). This, along with uncoupling classification from potential implications for an individual, has gone some way to increasing consistency and transparency in CNV classification. Online

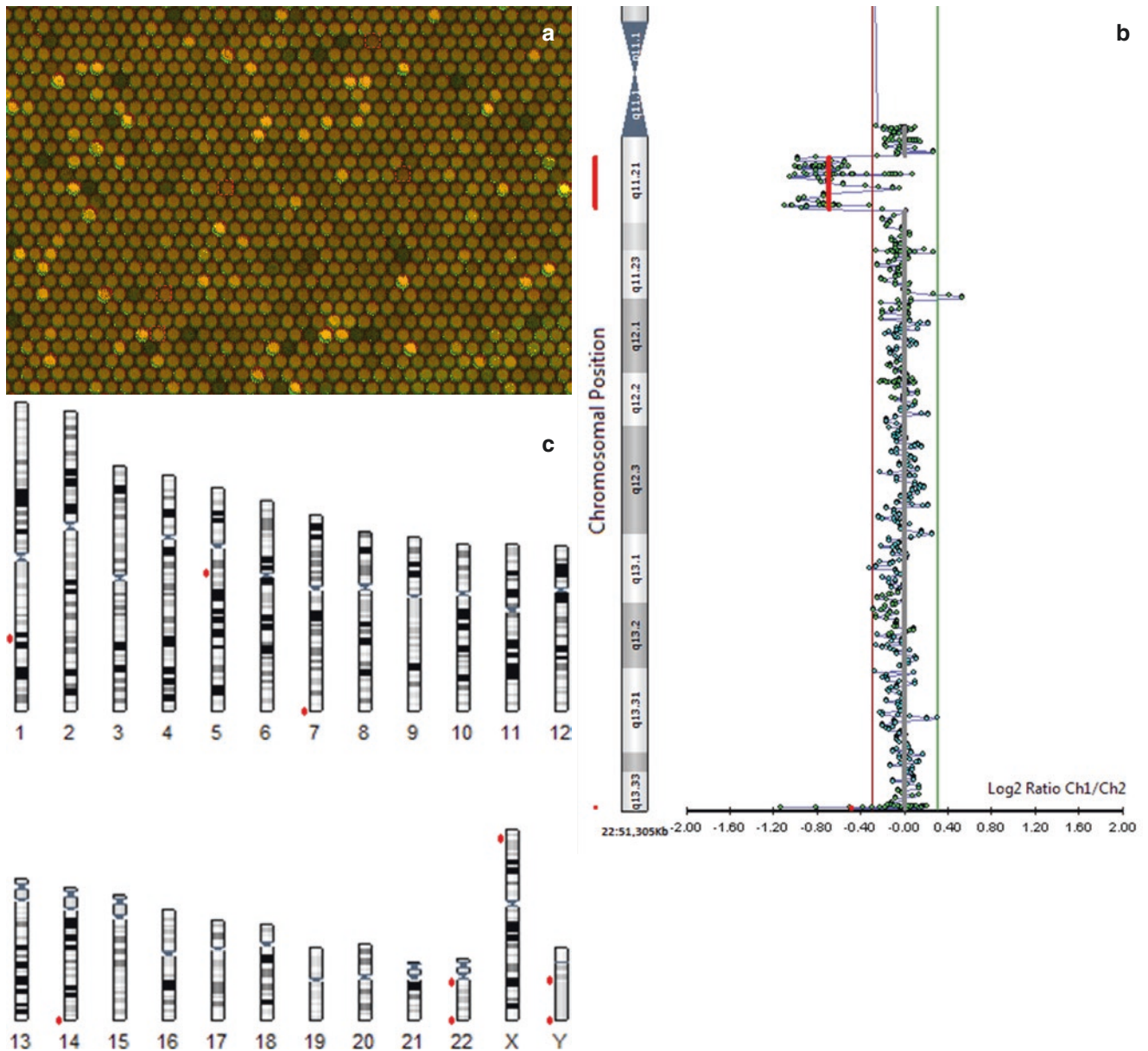


Fig. 3.21 Comparative genomic hybridisation (CGH) microarray. (a) Fluorescent imaging of CGH microarray chip. Yellow spots are the result of equal levels of hybridisation between control (red labeled) and test DNA (green labeled), indicating no copy number variation (CNV). Greater green intensity indicates a relatively greater level of test sample hybridisation; i.e., a CNV increase (e.g., from a duplication). Greater red intensity indicates a relatively higher level of control DNA hybridisation; i.e., a CNV decrease in the test sample compared to control (e.g., from a deletion). Although subtle and not very obvious to the human eye, sophisticated imaging technology is able to discriminate small differences between red and green intensity, with software indicating (by red broken line squares) spots that have a greater red intensity (deletion). (b) CGH array readout showing a heterozygous deletion in

chr22q11.21, indicated by a cluster with a marked decrease in the log₂ value (< -0.5 ; highlighted by the red line). Classical karyotyping for this child was normal, demonstrating the utility of the higher resolution genetic information obtained by CGH array. (c) A virtual karyotype generated from CGH array data in (b). Decreased CNVs are indicated by red dots, including chr22q11.21 (close to the centromere), associated with 22q11.2 deletion (velocardiofacial or DiGeorge) syndrome and consistent with the presenting phenotype. Note that not all CNVs are necessarily pathogenic. The region and nature of the CNV must be consistent with the presenting phenotype and currently available evidence of pathogenicity. Figures courtesy of Dr. J. Nicholl, SA Pathology, Australia

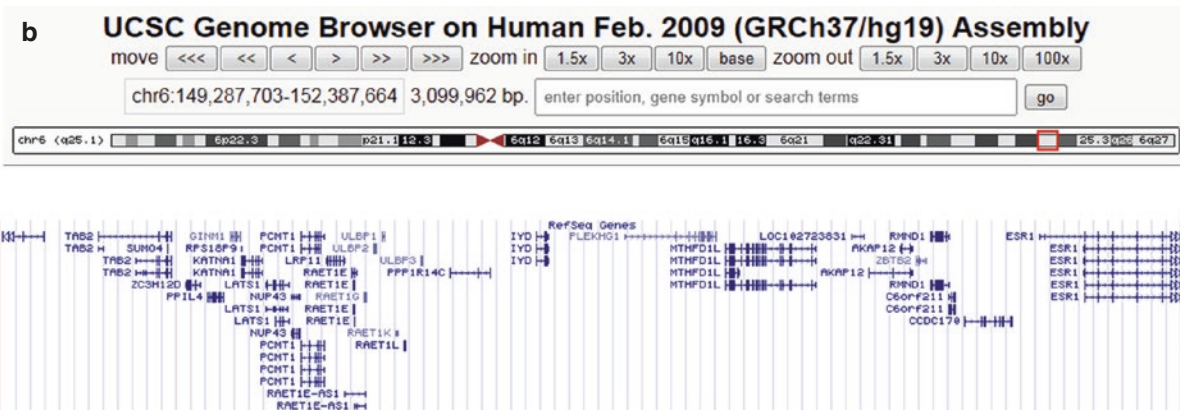
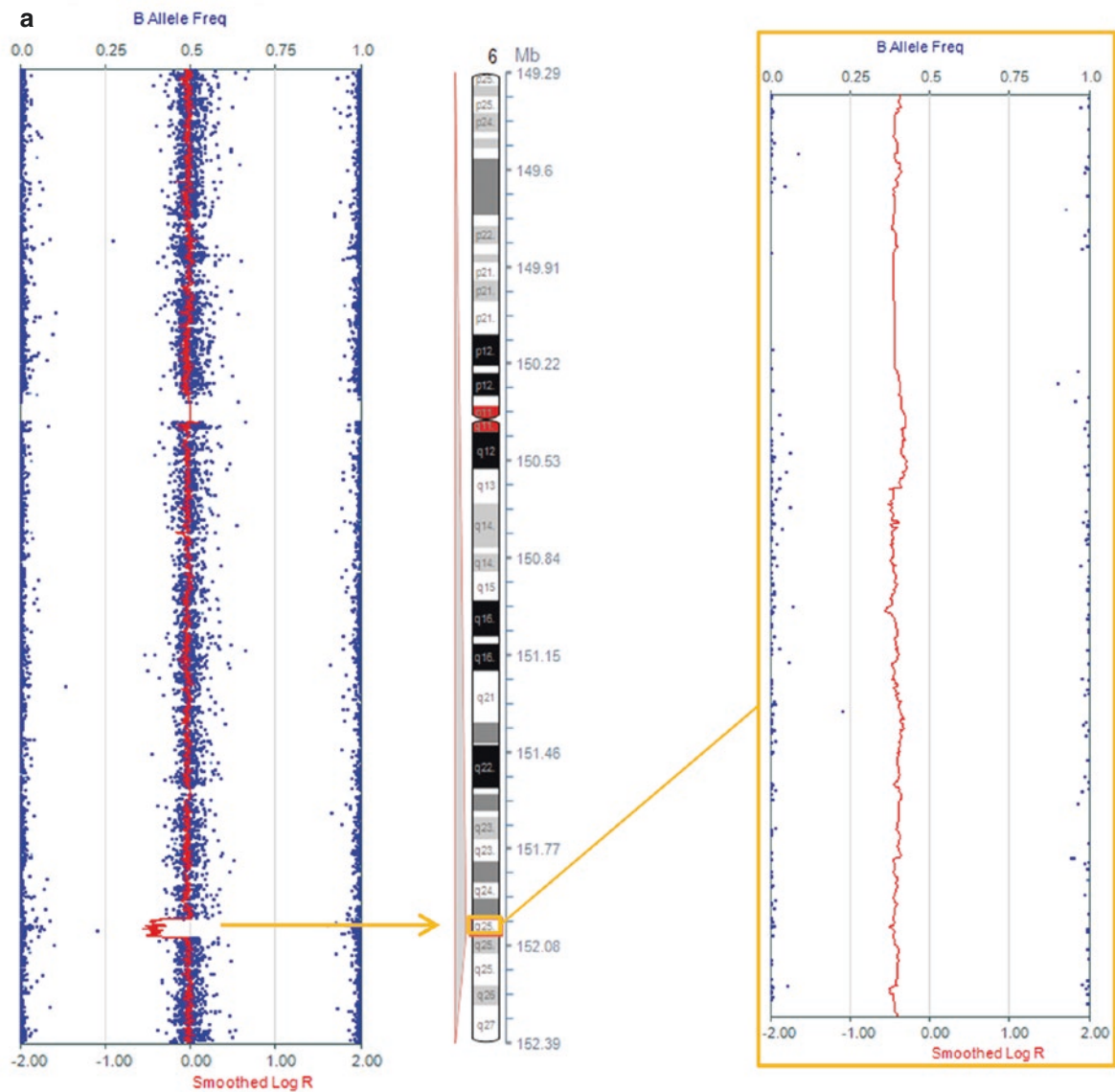


Fig. 3.22 Single nucleotide polymorphism (SNP) microarray. (a) Heterozygous deletion in chr6q25. Copy number variation (CNV) is indicated by a change in both the B allele frequency (<0.5) and Smoothed Log R (<0) values, shown as a dipping red line in both the gross (left) and fine (right) readouts. A virtual karyotype (centre) indicates the region the deletion is found in chromosome 6 (orange box).

(b) List of known RefSeq genes in the deleted region (red box) using UCSC Genome Browser. This includes the gene *TAB2*, associated with heterozygous cardiac development conditions, consistent with the phenotype. Figures courtesy of Ms. F. Norris, Victorian Clinical Genetics Services, Australia

gene dosage sensitivity databases can search by genomic region, for evidence of **haploinsufficient** or **triplosensitive** mechanisms of disease [43]. An online CNV pathogenicity calculator tool based on the scoring metric in the technical standard has also helped to streamline analysis [44].

Loss of heterozygosity (LOH) refers to deletion of an entire gene and/or surrounding chromosomal region, so that an allele from one parent is entirely lost. Regions with LOH are worth closer examination as potential hotspots for disease, often through gene dosage effects, i.e., reduction in relative expression of a gene product, indicated by CNV.

Microarrays are unable to determine balanced chromosomal anomalies (e.g., balanced translocations) or low levels of mosaicism.

The increased specificity of SNP compared to CGH arrays allows **copy-neutral LOH** to be detected. Also known as uniparental disomy (UPD), it refers to replication of the same chromosome from one parent, after loss of the chromosome from the other parent, during early development. It is important as a potential hotspot for recessive allele expression (given both alleles are copies of each other, and therefore automatically homozygous) as well as flags for chromosomes associated with well characterised imprinting disorders impacted by UPD (chromosomes 6, 7, 11, 14, 15, and 20) [45]. As cost differences have continued to decrease, these advantages have resulted in SNP array becoming the predominant microarray format offered by diagnostic laboratories.

- **Microarrays** are a good test to identify the cause of congenital abnormalities and intellectual disability (10–15% more chromosomal diagnoses made if the standard karyotype is normal).

3.6.14 Next-Generation Sequencing (NGS)

The power of **next-generation sequencing (NGS)**, also known as **massively parallel sequencing (MPS)**, comes from the ability to quickly and cheaply sequence billions of small fragments of DNA simultaneously (in parallel), combined with powerful, affordable computing for analysis of the large data sets produced (bioinformatics).

Improvement in sequencing technology output and reduction in cost have allowed it to be offered clinically to individual patients. Recently and remarkably, rapid NGS has been applied to the most acutely ill cohorts of patients, with turnaround times of 72 h for results [46], perhaps the benchmark for the future.

Most common clinically available NGS formats currently utilise short-read sequencing technologies (useful explanatory video here [47]). An example of the output of an NGS procedure is given in **Fig. 3.23**. Note, in this case

there are single nucleotide variations (**SNVs**) on both alleles, one a substitution and the other a deletion, indicating compound heterozygous variants. This demonstrates the power of NGS in that a very large number of individual fragments covering this region are sequenced individually, rather than the averaging approach of Sanger sequencing. It is also useful for demonstrating somatic differences present at very low percentage compared to germline tissue.

The number of times a region is individually sequenced is called **coverage depth** and obviously the larger this number with the same sequence result the greater the confidence it is a real variation in that individual's DNA.

Limitations of the different short-read platforms available for NGS include difficulties in sequencing GC rich regions and with length of reads limited to hundreds of base pairs range, making it difficult to detect **insertions** or **deletions (indels)** greater than approximately 50 bp. **Internal tandem repeats** or **homopolymer repeats** (of the same nucleotide; e.g., CCCCCC) can also cause sequencing problems or artefacts in short-read NGS.

There are a range of types of NGS based on how much of the genome is actually sequenced:

- **Panel:** uses a pre-amplification step, to select/enrich for regions of interest (e.g., only exons associated with cardiomyopathy)
- **Whole exome sequencing (WES):** exons only
- **Whole genome sequencing (WGS):** the entire genome

It should be noted that the NGS platform is suitable for application to any form of nucleic acid-based sequencing (e.g. genome, exome, transcriptome, methylome/epigenome, microbiome) as long as appropriate library preparation precedes the input onto a sequencing machine.

The post-wet lab component of data analysis to classification of findings and generation of a report is summarised in bioinformatics (Sect. 3.7). Sequencing in parallel with parents (**trio**) or other genetic relatives of a clinically affected individual (**proband**), can help to more rapidly identify **de novo** variants by ignoring sequence in common with unaffected relatives, especially where the phenotype is consistent with an autosomal dominant mechanism.

- **NGS panel testing** is a good test for congenital cardiomyopathies.
- **WES** is a good test for identifying new disease genes or non-classic presentation of a known syndromic condition, where insufficient clinical features have not raised suspicion regarding that syndromic diagnosis.
- **WGS** looms as a good test for almost every diagnostic genetic indication as cost and analysis times continue to decrease.

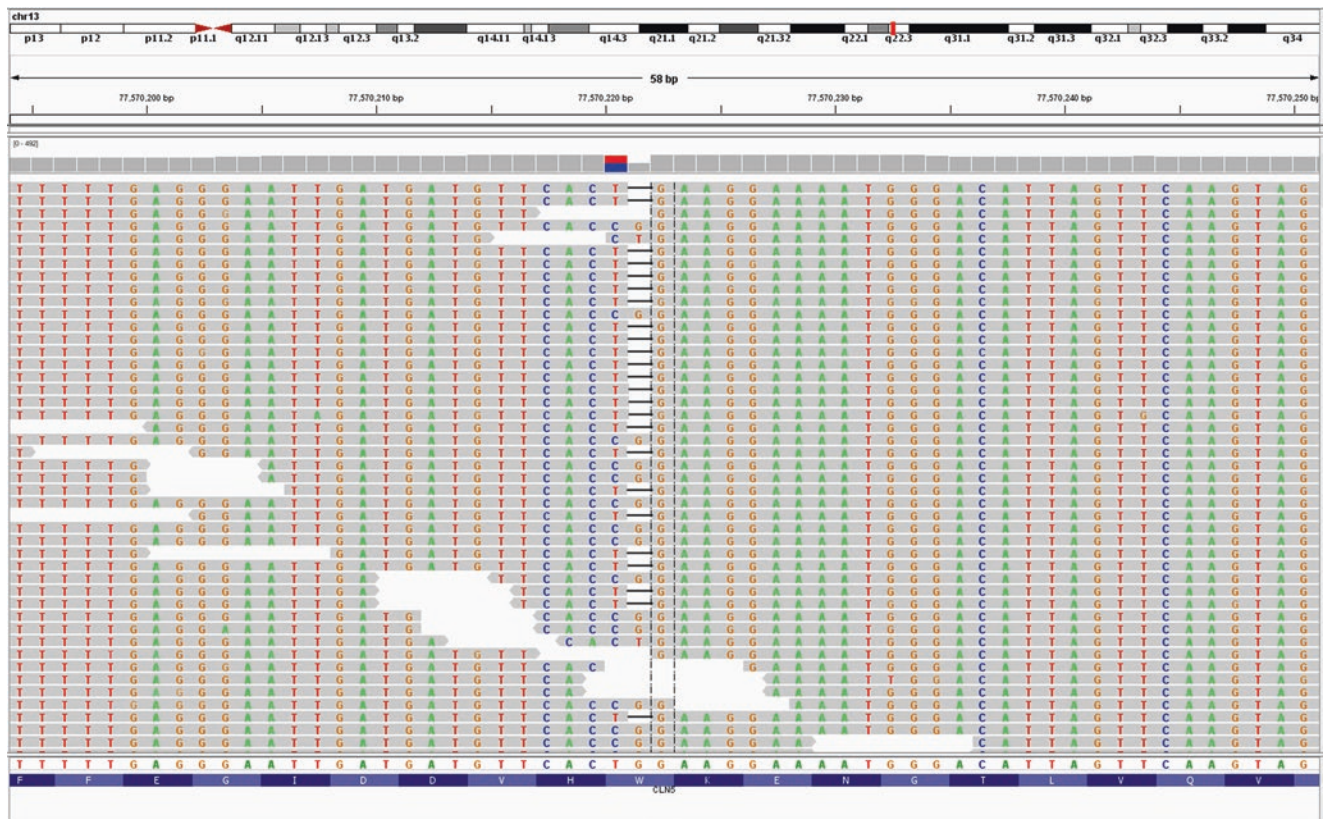


Fig. 3.23 Next-generation sequencing (NGS) of a compound heterozygous variant. In contrast to Sanger sequencing, NGS produces sequence readouts able to show differences down to the level of individual fragments of DNA. Sophisticated bioinformatic pipelines allow the data to be filtered according to many criteria, including quality of sequence, confidence of results, frequency, prevalence, clinical phenotype and known disease associations. Sequencing occurs in both the forward and reverse directions simultaneously, with even adjacent variants on separate alleles able to be clearly visualised. Shown is compound heterozygous pathogenic variants in the *CLN5* gene, associated

with Ceroid Lipofuscinosis, Neuronal, type 5. Gene location is indicated by a red vertical line through chr13q22.3 (top line) and numerically by genomic coordinates below that. The c.670T>C variant on one allele is indicated by a colour change from red to blue in the rectangles immediately above the sequence data as well as the individual letters of the sequence. Immediately adjacent, the c.671delG deletion variant on the second allele is denoted by a white rectangle above the data, with a black horizontal line in the individual sequences. Figure courtesy of Mr K. Brion, SA Pathology, Australia

3.6.15 Non-Invasive Prenatal Screening (NIPS)

Frequently called **Non-Invasive Prenatal Testing (NIPT)** it is important to underline that this is a screening test, resulting in a high or low risk profile. Although highly accurate as a screening test, with impressive sensitivity, specificity and positive predictive values for major trisomies [13, 18, 21], there are low but significant false positive and negative rates, meaning it should not be considered a diagnostic test. It utilises **cell-free DNA (cfDNA)**, small fragments of DNA (150–200 kb) freely circulating in plasma, no longer associated with the cell of origin, probably arising from a combination of cell death (i.e., apoptosis) plus extracellular “shedding” from intact cells. From 7 weeks gestation, in addition to their own maternal cfDNA, some fetal cells and fetal cfDNA (**cffDNA**) derived from placenta are present in the plasma of pregnant women.

Harnessing the power of massively parallel sequencing or microarray, minute amounts of fetal cfDNA can be detected even when only a small percentage of the total cfDNA (combination of cfDNA from both mother and fetus) in maternal plasma. By increasing the coverage depth and decreasing the numbers of regions assessed, massively parallel sequencing can theoretically sequence all molecules of cfDNA within a single sample. If there are even small amounts of change in the relative quantities of sequence associated with specific chromosomes in the cfDNA it can indicate aneuploidy. It may also be used for sex determination (detection of any Y chromosome cfDNA can indicate a male fetus, as the maternal cfDNA should have no Y chromosome material).

The main advantage for this technique is its non-invasive nature, compared to other prenatal cytogenetic techniques (CVS and amniocentesis). It can be performed with nothing more invasive than venepuncture for the mother and essen-

tially none of the risk of fetal loss associated with other invasive techniques (see Chap. 8).

Analysis relies on a statistical number crunching exercise. For example if 6% of total cfDNA is fetal and chromosome 21 (being one of the smaller chromosomes) represents 1.5% of the DNA in a genome, then a trisomy 21, will increase the amount of chromosome 21 fetal cfDNA in maternal plasma by 0.15% to give 1.625% of total fetal cfDNA, indicating Down syndrome.

The negative predictive value of the test, with a high-risk antenatal serum screen is of the order of 99%. Therefore, increasing adoption in antenatal screening has continued to decrease the number of cases that progress to invasive sampling for classical karyotyping. Fetal cfDNA less than 4% of the total cfDNA in maternal plasma is generally not sufficient for a reportable result and factors such as weight, age, ethnicity, twin/multiple and previous pregnancies, maternal disease or aneuploidy can impact cfDNA quality and quantity. NIPS using a WGS platform has also had reported utility in testing for rare autosomal trisomies, subchromosomal abnormalities and prenatal screening for parents with known balanced translocations [48, 49]. This technology has continued to decrease use of antenatal serum screening and invasive prenatal karyotyping but there still appears to be some inconsistencies in application and reporting of standardised measures across testing laboratories [50].

Currently it is recommended that positive NIPS tests are confirmed with confirmatory fetal karyotyping before any irreversible procedures are undertaken: the source of fetal cfDNA is placental and therefore a healthy fetus with placental mosaicism would be incorrectly classified using NIPS results alone.

- **NIPS** is a good screening test for Trisomies 13, 18, and 21, monosomy X and sex determination for X-linked disorders, but a positive result requires confirmation by an invasive test.

3.7 Bioinformatics

NGS has reached its current level of relatively wide availability due to both rapid advances and cost reduction in the core sequencing technology with parallel development of analytical tools on very powerful, yet affordable, computing platforms. The latter has pushed the field of bioinformatics to the very prominent position it enjoys today, as the engine behind NGS, deriving clinically significant meaning from the “big data” generated by this technology.

The main proprietary NGS technology platforms offer locked down software analysis tools but a very collaborative bioinformatics research community continues to produce very powerful and more customisable analysis “pipelines”.

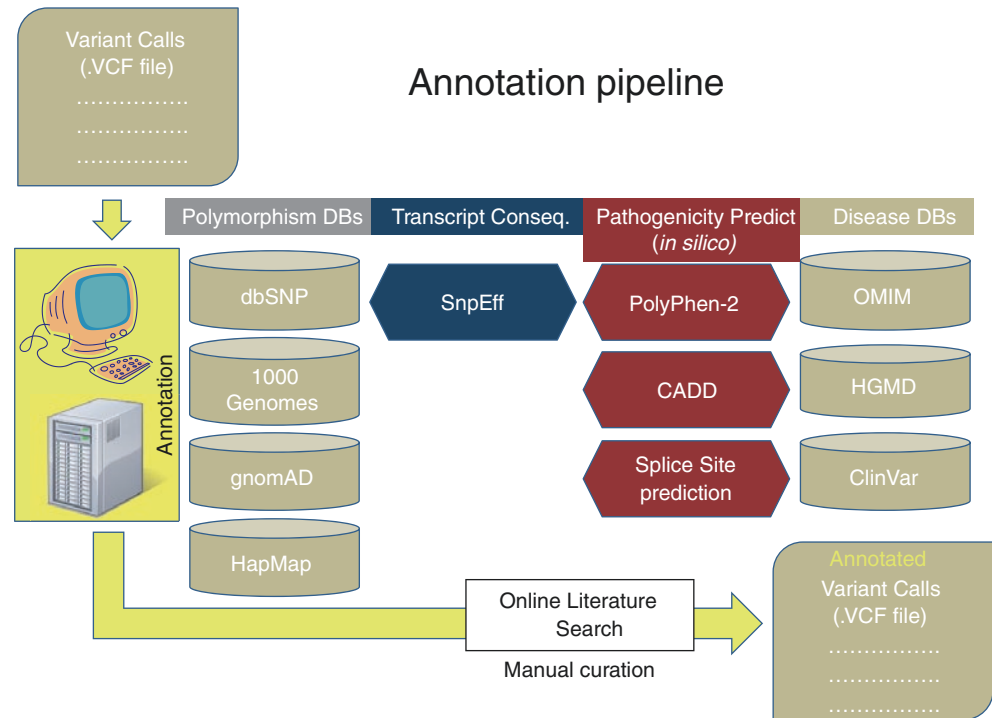
The basics of a bioinformatic analysis pipeline from the filtering stage are illustrated in flow diagram form in **Fig. 3.24**. The Broad Institute offer a useful set of imaging and analysis tools (e.g. Genome Analysis Tool Kit, **GATK**; Integrative Genome Viewer, **IGV**), that are a good starting point [51, 52]. Overall bioinformatic steps can be summarised as:

- Sequencing machine data generation and storage
- Convert instrument signal data into individual sequence fragment nucleotide calls with quality scores (**primary analysis** e.g. **FASTQ** file)
- Alignment of all sequences from the fragments, using a reference human genome **assembly** e.g. GRCh38/hg38 (**secondary analysis** e.g. **BAM** file)
- Variant calling, i.e. determining which changes are divergent from the reference genome (**secondary analysis** e.g. **VCF** file)
- Use filters (e.g. based on biological effect, population frequency, previous classification, gene/variant-phenotype association) to produce a list of candidate variants related to the testing indication (**tertiary analysis, annotation**)
- Manually interrogate candidate variants for integrity and suitability (e.g. using IGV [51]), proceeding to classification for suitable candidates
- Use databases and predictive algorithms to assess gene/variant-disease and/or phenotype correlation, mechanism and segregation of disease, population frequency, protein structure, function, evolutionary conservation, physiochemical change, allelic context, splicing impact and previous classifications or other evidence (**tertiary analysis, curation**)
- Use standardised guidelines with a Bayesian probabilistic risk framework to classify variants according to one of five tiers (**tertiary analysis, classification**) [8, 53–55]
 - (1) Benign (probability <0.1%)
 - (2) Likely benign (probability 0.1–5%)
 - (3) Variant of uncertain (or unknown) significance (**VUS**) (probability 5–95%)
 - (4) Likely pathogenic (probability 90–99%)
 - (5) Pathogenic (probability >99%)

Although commercial curation packages can offer significant time (and therefore cost) savings, useful free online tools for variant analysis abound. A starting list that is far from comprehensive is included here [20, 56–71], although the utility and availability is likely to vary over time as the next best tool comes along. Literature searches remain an essential part of variant curation, utilising well known search engines [72, 73].

While the data acquisition component is usually comprehensive for the entire set of changes detected compared to a reference genome, it is normally only practical to filter a subset of the genetic information sequenced, for intensive interrogation (**tertiary analysis**), based largely on the clinical indication (phenotype). For NGS, the term *in silico* is cur-

Fig. 3.24 Example of a bioinformatic annotation pipeline for next-generation sequencing (NGS). The annotation process combines polymorphism and disease databases (DBs) with transcription consequence and pathogenicity prediction tools, plus manual curation by traditional literature searches. Figure courtesy of Dr K. Kassahn, SA Pathology, Australia. Web pages for all of the resources in this Figure are included in references [19, 55, 59, 67, 68, 86-90]



rently used to describe computer-based analysis or simulation, particularly with regard to variant pathogenicity prediction algorithms (Fig. 3.24).

Quality of the bioinformatic pipeline in generating clinically meaningful results is still highly dependent on the quality of the clinical information provided. Variant curators get even more upset with blank clinical indication fields than pathologists. Clinical information recording and interrogation has been aided by efforts to standardise phenotype ontologies [74].

Determining if an identified genetic variation is relevant or not is vastly aided by databases that compile genetic variations, such as dbSNP [68]. Although somatic changes are only very briefly covered in this chapter, COSMIC [75] is a useful tool for this area.

The Genome Aggregation Database (gnomAD) [60] is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community. This database is referenced in most genetic sequencing results as it provides a prevalence Figure for the presence of a particular change at a particular point in the genetic code, based on the results from a large cohort of adults (>100,000 individuals) without early onset childhood disease. A genomic change in an affected individual under investigation that is ultra-rare or absent in gnomAD is more likely to be pathogenic.

Variant classification guideline refinement, global standardisation and information sharing platforms, plus proposed numerical-based scoring systems are likely to continue

to allow automation and enhancement of this process [54, 55, 76]. Despite all this, VUS variants, with the widest pathogenic risk profile (5–95%) are likely to remain the most common classification result, much to the chagrin of patients, families and clinicians alike. This is a function of the simple fact that we do not have enough information yet about many of the individual variants flagged for curation. As collective databases grow, in size and breadth, hopefully the number of variants in this category will also decrease.

3.8 Future of Genetic Testing

The provision of results from genetic testing in the fastest possible time will continue to be a major focus for genomics. As more is understood about normal human genomic variation, identification of abnormal human variation becomes easier. Large and rapidly growing databases of normal genomic variation such as gnomAD, which is freely available, enable more precise and rapid diagnoses to be made. Increased automation through bioinformatic pipelines and NGS technology improvements are also speeding up the diagnostic process. There is a strong willingness to freely share technology improvement and genomic data between institutions and countries, as stakeholders acknowledge the power in this sharing process. A good example of this willingness is the Global Alliance for Genomics and Health (GA4GH), a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework [76]. Large international efforts to establish the validity of specific gene-

disease relationships also promise much benefit [58]. The power of (confidential) social networking also shows promise for resolving the rarest of genetic disorders through exchange of individual phenotype and genotype information [77].

The NGS platform is suitable for any nucleic acid based tests, including analysis of the epigenome (miRNA, lncDNA, CpG methylation), an emerging field likely to continue making inroads into the diagnostic realm. There is also evidence that somatic changes may be responsible for some congenital diseases; e.g., in brain development [78]. Methylation arrays have also been developed which can identify regions of the genome with particular methylation “stamps” (DNA methylation epigraphs) which draw attention to the high probability of an underlying variation in a gene known to modify chromatin production, for example. Specific patterns in the methylomes of individuals with defined congenital syndromes have been recognised, with methylation arrays being a potential valuable clinical tool [79].

Given the vast amounts of data generated and computing power required, the whole field has continued to embrace the advantages and caveats associated with cloud computing [80]. As price and availability of WGS continues to improve, potential for more robust data on structural variation, non-coding regions and even CNVs from this technology will likely increase. **Long-read (third generation) sequencing** holds out the promise for better resolution of challenging or previously inaccessible regions of the genome (e.g. repeat regions, pseudogenes, telomeres), more comprehensive methylation characterisation, plus decreased amplification artefacts, sequence assembly and alignment problems, once costs and accuracy improve to be clinically practical [81].

Functional genomics is an evolving field [82]. The ability to predict whether a particular variation in the genetic code will result in abnormal protein production is an incredibly important question that sometimes cannot be answered. Emerging technologies such as RNAseq, which uses NGS to reveal the presence and quantity of RNA in a sample as a marker of the expression of a particular gene or set of genes (transcriptome), is a potentially valuable functional genomics tool. Finding a DNA-based variant which has a proven negative effect on RNA production and therefore protein production is a big step forward in terms of establishing the pathogenicity of the DNA change [83].

Optical genome mapping (OGM) is an emerging technology to watch as it has the potential to markedly impact classical karyotyping, with automation and resolution benefits similar to the impact NGS has had on sequencing [84].

There has been much hope for these new genetic-based diagnostic technologies, with many blue sky promises and much marketing hype behind them. However, the value of the data they generate will continue to be determined by the quality of clinical description—human factors that are

unlikely to be superseded by technology any time soon, but that will be enhanced by continuing attempts for standardising ontologies [74].

There are many ethical considerations already emerging from the new genetic testing regimes, and a variety of guidelines and laws are likely to be created across many different professional and societal jurisdictions [85]. Many individual patients and families have already benefited from powerful new genetic testing technologies, gaining answers previously not able to be found by other diagnostic odysseys. However, it should always be underlined that the analyses underlying these technologies are based on a probabilistic risk model and that imposing overly deterministic applications has many potential avenues for harm. It is imperative that the power of this potentially rich information mine, be harnessed in tandem with primary stakeholders such as patients, families, support organisations and (often vulnerable or historically mistreated) communities, with their health priorities and wishes for implementation and utility remaining the foremost considerations.

References

1. Genomics England Education. Available from: www.genomicseducation.hee.nhs.uk/education/.
2. Preparing Healthcare Providers for Genomic Medicine. Available from: www.pathologylearning.org/trig/resources.
3. MGHA Educational Resources. Available from: <https://learn-genomics.org.au/educational-resources/>.
4. Your Genome. Available from: www.yourgenome.org.
5. Genetics Home Reference: U.S. National Library of Medicine. Available from: <https://medlineplus.gov/genetics/>.
6. Fact Sheets About Genomics: The National Human Genome Research Institute (NHGRI). Available from: www.genome.gov/about-genomics/fact-sheets.
7. Chromatin. Available from: www.genome.gov/genetics-glossary/Chromatin.
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med*. 2015;17:405–24.
9. Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. *Nat Genet*. 2020;52:754–8.
10. HGNC. HUGO Gene Nomenclature Committee; (May 29, 2021). Available from: <http://www.genenames.org/>.
11. DNA Replication. Available from: www.yourgenome.org/video/dna-replication.
12. DNA replication: how genetic information is passed on, 3D animation with narration: The Cold Spring Harbor Laboratory. Available from: www.dnalc.org/view/15530-DNA-replication-how-genetic-information-is-passed-on-3D-animation-with-narration.html.
13. Codons and amino acids: Human Genome Variation Society. Available from: www.hgvs.org/mutnomen/codon.html.
14. Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol*. 2010;220:152–63.
15. Vaz-Drago R, Custodio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet*. 2017;136:1093–111.

16. The Atlas for Genetics and Cytogenetics in Oncology and Haematology. Available from: <http://atlasgeneticsoncology.org/Educ/NomMutID30067ES.html>.
17. HGVS simple: Human Genome Variation Society. Available from: <http://varnomen.hgvs.org/bg-material/simple/>.
18. HGVS General. Available from: <https://varnomen.hgvs.org/recommendations/general/>.
19. Online Mendelian Inheritance in Man®. An Online Catalog of Human Genes and Genetic Disorders OMIM. Available from: www.omim.org.
20. RefSeq NCBI. Available from: www.ncbi.nlm.nih.gov/refseq/.
21. LRG Locus Reference Genomic. Available from: www.lrg--sequence.org.
22. Wright MW. A short guide to long non-coding RNA gene nomenclature. *Hum Genomics*. 2014;8:7.
23. McGowan-Jordan J, Hastings RJ, Moore S. ISCN 2020: an international system for human cytogenomic nomenclature (2020): Karger; 2020.
24. Trent RJ. Molecular medicine: genomics to personalized health-care. 4d. Amsterdam: Elsevier/AP; 2012. p. ix, 336
25. Roadmap Epigenomics Project: NIH Roadmap Epigenomics Mapping Consortium. Available from: www.roadmapepigenomics.org/overview.
26. Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med*. 2014;6:60.
27. Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet*. 2019;20:235–48.
28. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long non-coding RNAs in health and disease. *Cell*. 2013;152:1308–23.
29. Video animation: RNA interference: Nature Publishing Group; (animation of RNAi mechanism). Available from: www.nature.com/news/video-animation-rna-interference-1.9673.
30. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol*. 2014;6:a018382.
31. Long Noncoding RNA Database: Garvan Institute. Available from: <https://lncipedia.org/>.
32. International Human Epigenome Consortium: International Human Epigenome Consortium. Available from: www.ihec-epigenomes.org.
33. HEP: Human Epigenome Project. Available from: www.epigenome.org.
34. Schmidt EE PO, Buhlmann S, Kerr G, Horn T, Boutros M. GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes. Available from: www.genomernai.org.
35. Triplet-primed PCR video. Available from: <https://uofmedicalgenomics.home.blog/2019/12/10/triplet-repeat-diagnostic-techniques/>.
36. Eijk-Van Os PG, Schouten JP. Multiplex Ligation-dependent Probe Amplification (MLPA®) for the detection of copy number variation in genomic sequences. *Methods Mol Biol*. 2011;688:97–126.
37. Sørensen KM, Agergaard P, Olesen C, Andersen PS, Larsen LA, Ostergaard JR, et al. Detecting 22q11.2 deletions by use of multiplex ligation-dependent probe amplification on DNA from neonatal dried blood spot samples. *J Mol Diagn*. 2010;12:147–51.
38. Furtado LV, Paxton CN, Jama MA, Tripp SR, Wilson AR, Lyon E, et al. Diagnostic utility of microsatellite genotyping for molar pregnancy testing. *Arch Pathol Lab Med*. 2013;137:55–63.
39. Beygo J, Buiting K, Ramsden SC, Ellis R, Clayton-Smith J, Kanber D. Update of the EMQN/ACGS best practice guidelines for molecular analysis of Prader-Willi and Angelman syndromes. *Eur J Hum Genet*. 2019;27:1326–40.
40. Brioude F, Kalish JM, Mussa A, Foster AC, Blied J, Ferrero GB, et al. Expert consensus document: Clinical and molecular diagnosis, screening and management of Beckwith-Wiedemann syndrome: an international consensus statement. *Nat Rev Endocrinol*. 2018;14:229–49.
41. Manning M, Hudgins L, Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med*. 2010;12:742–5.
42. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*. 2020;22:245–57.
43. ClinGen Gene Dosage Sensitivity. Available from: <https://search.clinicalgenome.org/kb/gene-dosage>.
44. ClinGen CNV Pathogenicity Calculator. Available from: <https://cnvcalc.clinicalgenome.org/cnvcalc/>.
45. Dharmadhikari AV, Ghosh R, Yuan B, et al. Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,020 clinical exome cases. *Genome Med*. 2019;11:30.
46. Australian Genomics Health Alliance Acute Care Flagship, Lunke S, Eggers S, Wilson M, Patel C, Barnett CP, et al. Feasibility of ultra-rapid exome sequencing in critically ill infants and children with suspected monogenic conditions in the Australian public health care system. *JAMA*. 2020;323:2503–11.
47. Sequencing at Speed. Available from: www.yourgenome.org/video/sequencing-at-speed.
48. Pertile MD. Chapter 7 – Genome-wide cell-free DNA-based prenatal testing for rare autosomal trisomies and subchromosomal abnormalities. In: Page-Christiaens L, Klein H-G, editors. *Noninvasive Prenatal Testing (NIPT)*. Amsterdam: Academic Press; 2018. p. 97–123.
49. Flowers NJ, Burgess T, Giouzeppos O, Shi G, Love CJ, Hunt CE, et al. Genome-wide noninvasive prenatal screening for carriers of balanced reciprocal translocations. *Genet Med*. 2020;22:1944–55.
50. Skotko BG, Allyse MA, Bajaj K, Best RG, Klugman S, Leach M, et al. Adherence of cell-free DNA noninvasive prenatal screens to ACMG recommendations. *Genet Med*. 2019;21:2285–92.
51. Genome Analysis Toolkit (GATK): the Broad Institute. Available from: <https://gatk.broadinstitute.org/hc/en-us>.
52. IGV Integrative Genomics Viewer. Available from: <https://igv.org/app/>.
53. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. 2018;20:1054–60.
54. Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat*. 2020;41:1734–7.
55. Ellard S, Baple EL, Callaway A, Berry I, Forrester N, Turnbull C, et al. ACGS best practice guidelines for variant classification 2020. Association for Clinical Genetics Science (ACGS) UK; 2020.
56. The Human Gene Mutation Database: the Institute of Medical Genetics in Cardiff. Available from: www.hgmd.cf.ac.uk.
57. Gene Reviews. Available from: www.ncbi.nlm.nih.gov/books/NBK11116/.
58. The Gene Curation Coalition. Available from: <https://search.the-gence.org/>.
59. Genomics England PanelApp. Available from: <https://panelapp.genomicsengland.co.uk/>.
60. Genome Aggregation Database (May 29, 2021). Available from: <https://gnomad.broadinstitute.org/>.
61. Varsome. Available from: <https://varsome.com/>.
62. Uniprot. Available from: www.uniprot.org.
63. ClinGen Variant Pathogenicity Calculator. Available from: <https://calculator.clinicalgenome.org/site/cg-calculator>.

64. Metadome – missense constraint. Available from: <https://stuart.radboudumc.nl/metadome/>.
65. VarCards in silico tools aggregator. Available from: <http://159.226.67.237/sun/varcards/>.
66. KEGG: Kyoto Encyclopedia of Genes & Genomes: Kanehisa Laboratories; (a useful compilation of biochemical pathways and biological basis of drug interactions). Available from: www.genome.jp/kegg.
67. UCSC Genome Bioinformatics: University of California, Santa Cruz. Available from: <http://genome.ucsc.edu>.
68. dbSNP Short Genetic Variations: NCBI. Available from: www.ncbi.nlm.nih.gov/snp/.
69. ClinVar. Available from: www.ncbi.nlm.nih.gov/clinvar/.
70. DECIPHER. Available from: www.deciphergenomics.org.
71. Ensembl European Molecular Biology Laboratory (May 29, 2021). Available from: www.ensembl.org.
72. PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov/>.
73. Google Scholar. Available from: <https://scholar.google.com/>.
74. Köhler S DS, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Available from: <https://hpo.jax.org/app/>.
75. Catalogue of Somatic Mutations in Cancer (COSMIC): Wellcome Trust Sanger Institute. Available from: <https://cancer.sanger.ac.uk/cosmic>.
76. Global Alliance for Genomics & Health. Available from: www.ga4gh.org.
77. Matchmaker Exchange. Available from: www.matchmakerexchange.org.
78. Jamuar SS, Lam AT, Kircher M, D’Gama AM, Wang J, Barry BJ, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med*. 2014;371:733–43.
79. Aref-Eshghi E, Kerkhof J, Pedro VP, France GD, Barat-Houari M, Ruiz-Pallares N, et al. Evaluation of DNA methylation epigenatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *Am J Hum Genet*. 2020;106:356–70.
80. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinformatics*. 2014;15:30.
81. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21:597–614.
82. Rare Diseases: Models & Mechanisms Network. Available from: <http://www.rare-diseases-catalyst-network.ca/>.
83. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.
84. Sahajpal NS, Barseghyan H, Kolhe R, Hastie A, Chaubey A. Optical genome mapping as a next-generation cytogenomic tool for detection of structural and copy number variations for prenatal genomic analyses. *Genes (Basel)*. 2021;12:398.
85. Lohmann K, Klein C. Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics*. 2014;11:699–707.
86. 1000 Genomes Browser: National Center for Biotechnology Information (NCBI). Available from: www.ncbi.nlm.nih.gov/variation/tools/1000genomes/.
87. International HapMap Project. Available from: <http://hapmap.ncbi.nlm.nih.gov>.
88. SnpEff. Available from: <https://pcingola.github.io/SnpEff/>.
89. PolyPhen-2. Available from: <http://genetics.bwh.harvard.edu/pph2/>.
90. CADD. Available from: <https://cadd.gs.washington.edu/>.