

Claudine L. Lefferts and Joel A. Lefferts

2.1 Biology of Nucleic Acids

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are macromolecules that convey genetic information. Both DNA and RNA are made up of nucleotides, molecules that in turn are composed of a nitrogenous base, a sugar, and one or more phosphate groups. There are five nitrogenous bases found in nucleic acids: adenine, guanine, cytosine, thymine, and uracil. Adenine, guanine, and cytosine are found in both DNA and RNA; thymine is found only in DNA and uracil is found only in RNA. A second difference between DNA and RNA is the sugars that are incorporated into the nucleotides. The sugar in the nucleotides of DNA is deoxyribose, RNA nucleotides contain ribose. Finally, DNA molecules are double-stranded, while RNA molecules are usually single-stranded.

2.1.1 Composition and Structure of DNA

DNA is formed by the linear polymerization of nucleotides. The four nitrogenous bases found in DNA are either purines (adenine or guanine) or pyrimidines (cytosine or thymine), and the backbone of the DNA polymer is formed by linkage of these bases via deoxyribose and phosphate groups (Fig. 2.1). The informational content of DNA is governed by the sequential arrangement and primary structure of the nucleotide polymer. The DNA strand is polar, with no nucleotide attached to the 5' position of the deoxyribose at one end (referred to as the 5' end), and no nucleotide attached to the 3' hydroxyl group at the other end (referred to as the 3' end).

The DNA within the eukaryotic nucleus is arranged in a double-stranded helix composed of two strands of opposing

polarity. The helix is stabilized by the formation of hydrogen bonds between complementary bases (A-T and G-C), by pi bonding that occurs when the bases are stacked together, and by the association of proteins [1, 2]. In eukaryotic cells, most of the DNA is in the B-form, a right-handed helix with bases on the inside where they are protected from damage by oxidating or alkylating agents. The Z-form of DNA occurs when a left-handed helix is formed, and is usually associated with portions of the DNA that are highly methylated and are not transcribed actively. Enzymatic reactions within the nucleus are responsible for conversion of DNA from the B-form to the Z-form and vice versa [3].

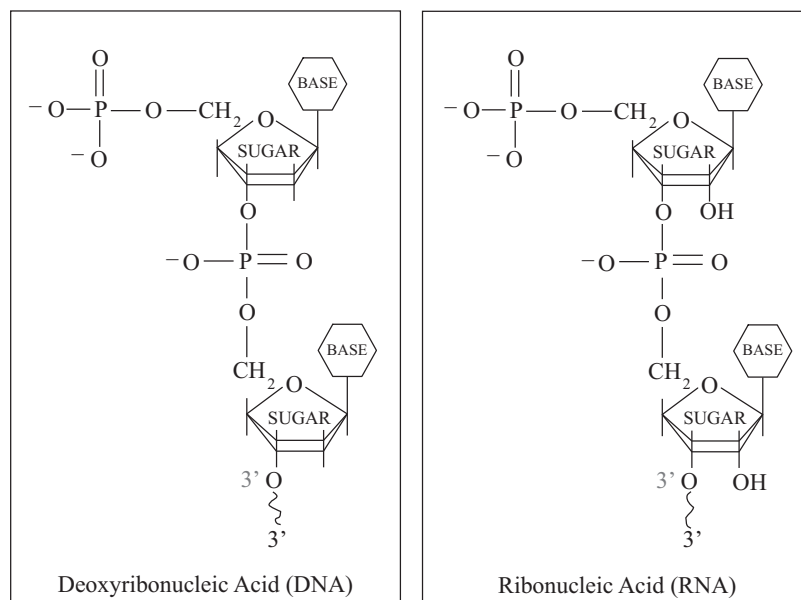
The DNA comprising the human genome is segmented into 46 discrete structural units, termed chromosomes. The DNA in each eukaryotic cell must be compressed to fit within the nucleus, which is only about 10 μm in diameter. Chromosomal DNA is condensed by the formation of nucleosomes, which consist of a group of small basic proteins (histones) with 160–180 base pairs of DNA wrapped around them [4]. Formation of nucleosomes is not sequence-dependent and it occurs in mammalian, bacterial, and viral DNA [5]. Nucleosomes are wound into a left-handed helix for further condensation of the DNA, and higher orders of structure include supercoils and/or rosettes [2]. Ultra-condensed DNA (heterochromatin) is inactive metabolically, and is found primarily in the periphery of the nucleus, while less condensed DNA (euchromatin) is readily accessible by transcription machinery and is located in the center of the nucleus [6].

2.1.2 Gene Organization

The majority of genes that are transcribed into mRNA and translated into cellular proteins exist as two copies in the nucleus of each cell, one maternal and one paternal copy. Some genes are present at a high copy number (100–250 copies) within the genome, including the genes that encode transfer RNA (tRNA), ribosomal RNA (rRNA), and the histone proteins [7]. These tandemly repeated genes are present

C.L. Lefferts, Ph.D. • J.A. Lefferts, Ph.D. (✉)
Department of Pathology and Laboratory Medicine,
Dartmouth-Hitchcock Medical Center, One Medical Center Drive,
Lebanon, NH 03756, USA
e-mail: joel.a.lefferts@hitchcock.org

Fig. 2.1 Chemical structures of DNA and RNA.



on several chromosomes and associate in the nucleus to form a nucleolus [8, 9]. Highly repetitive sequences with thousands of copies, called satellite DNAs, are found at the telomeric ends of chromosomes and around the centromeres [10]. It is likely that the centromeric sequences play a role in the establishment and maintenance of chromosome structure. Telomeric repeats are involved in completing replication of chromosome ends [11], and it has been demonstrated that the length of the telomeric repeat sequences decreases with life-span of cultured human cells [12]. The evidence of telomeric shortening in normal cells along with observations that immortalized or transformed cells display limited telomere degeneration has led to the hypothesis that telomeric shortening is involved in the cellular aging process [12]. Other repetitive sequences, such as the *Alu* sequences, are found throughout chromosomes; their function is largely unknown, but a role in regulation of gene function has been proposed [13].

Simple polymorphic repetitive elements composed of dinucleotide or trinucleotide repeats are present in the human genome and have been associated with cancer and other diseases such as myotonic dystrophy, Fragile X syndrome, Huntington disease, and spinocerebellar ataxia [13, 14]. Symptomatic problems arise when the affected DNA is inappropriately methylated and inactivated (as in Fragile X syndrome) [15], or when the repeats cause detrimental changes in the encoded protein (as in Huntington disease) [16].

2.1.3 DNA Replication and Cell Division

In order for the DNA in a cell to be replicated prior to cell division it must be single-stranded. This is accomplished by an enzyme (helicase), which denatures the DNA and allows DNA-binding proteins to associate with the DNA and prevent

reformation of the DNA helix [17]. A small strand of RNA 10–20 nucleotides in length acts as a primer, initiating synthesis of new complementary strands of DNA from multiple replication starting points. Deoxynucleotide triphosphates (dNTPs) are added to the primer by DNA polymerase, the RNA primers are removed, the gaps are filled in with dNTPs by DNA polymerase, and the nucleotide strands are joined by DNA ligase. The enzymatic action of topoisomerase removes twists generated during denaturation of the helix and allows the helix to re-form [18]. DNA replication is complete when the telomerase enzyme has added the nucleotide repeats to the telomeres at the 5' end of the DNA strands.

As a new strand of DNA is synthesized, dNTPs are selected based on hydrogen bonding to complementary dNTPs in the template strand, which results in an error rate of 1 in 10^4 – 10^5 [2]. Eukaryotic cells employ a proofreading mechanism that removes mispaired dNTPs in the strand before the next dNTP is added, which decreases the error rate to 1 in 10^6 – 10^7 [2, 19]. Prior to cell division, another error correction system recognizes and repairs mismatched nucleotides and decreases the error rate further to 1 in 10^8 – 10^9 [2, 19]. Several inherited disorders are due to dysfunctional DNA damage-repair, including ataxia telangiectasia, Fanconi anemia, and xeroderma pigmentosum [20].

The DNA within the nucleus of a eukaryotic cell can be replicated completely in about 8 h, during the S phase of the cell cycle. Resting cells (G_0) receive a mitotic stimulus, which causes transition into the G_1 phase, where the cell prepares for DNA synthesis (S phase). The G_2 phase occurs after replication but before division, and mitosis (M) involves actual nuclear and cellular division. The cell cycle is pivotal in cellular and organismal homeostasis, so it is tightly controlled by phosphorylation and dephosphorylation of kinases and cyclins, and by two major checkpoints [21,

22]. The first checkpoint occurs between G_1 and S, and can prevent cells with damaged or unrepaired DNA from entering S phase. The second checkpoint occurs between G_2 and M, and can prevent the initiation of mitosis [21].

2.1.4 Structure and Composition of RNA

RNA, or ribonucleic acid, is a linear polymer of ribonucleotides linked by 5′–3′ phosphodiester bonds (Fig. 2.1). RNA differs from DNA in that the sugar group of RNA is ribose, rather than deoxyribose, thymine is replaced by uracil (U) as one of the four bases, and RNA molecules are usually single-stranded. The extra hydroxyl group present on the ribose causes RNA to be more susceptible to degradation by nucleases than DNA. Single-stranded RNA molecules form complex secondary structures, such as hairpin stems and loops, via Watson–Crick base pairing between adenine and uracil, and between guanine and cytosine.

RNA molecules are classified by function and cellular location, and there are three major forms of RNA in eukaryotic cells. Ribosomal RNA (rRNA) is the most stable and most abundant RNA. rRNA is highly methylated, and complexes with proteins to form ribosomes upon which proteins are synthesized [23]. In eukaryotic cells, two major species of rRNA are present, 28S and 18S, as well as two minor species of 5.8S and 5S. Transfer RNA (tRNA) is responsible for carrying the amino acid residues that are added to a growing protein chain during protein synthesis. All tRNAs form secondary structures consisting of four stems and three loops, and many bases found in tRNA are modified by methylation, ethylation, thiolation, and acetylation [24, 25]. Messenger RNA (mRNA) mediates gene expression by carrying coding information from the DNA to the ribosomes, where the mRNA molecule is translated into protein. Messenger RNA is the most heterogeneous type of RNA, and also has the shortest half-life.

Many other RNA species exist in eukaryotic cells. Heterogeneous nuclear RNAs (hnRNA) are the precursors to mature mRNA [26]. Small nuclear RNAs (snRNAs) associate with proteins to form small nuclear ribonucleoprotein particles, which participate in RNA processing [27]. Small interfering RNAs (siRNAs) and microRNAs are double-stranded RNAs 21–25 nucleotides in length involved in regulation of gene expression at the translational level [28].

2.1.5 Transcription of RNA

Even simple eukaryotic organisms, such as yeast, contain a large number of genes (~2000), and higher eukaryotes, such as mammals, have ~20,000–25,000 protein encoding genes [29]. Clearly, the proteins encoded by all genes are not expressed simultaneously at any given time. Transfer of genetic information from DNA to protein begins with syn-

thesis of RNA molecules from a DNA template by RNA polymerase, a process termed transcription. The RNA polymerase holoenzyme works processively, building an RNA chain with ribonucleoside triphosphates (ATP, GTP, CTP, UTP) [30]. Initiation of transcription involves association of the transcription machinery (RNA polymerase and transcription factors) with the DNA template and the synthesis of a small ribonucleotide primer from which the RNA strand will be polymerized [31]. Initiation of transcription is not random, but occurs at specific sequences called promoters that are located at the 5′-end of genes. Every gene initiates transcription independently at its own promoter, therefore the efficiency of the process varies greatly depending on the strength of the promoter. Once RNA polymerase binds to a promoter, the DNA helix is opened and an RNA primer is synthesized. Elongation occurs as the RNA polymerase moves along the DNA strand, opening the DNA helix and conducting DNA-directed RNA synthesis until the gene is transcribed [30]. Termination of transcription is poorly understood in eukaryotes, but takes place at sites that include a stretch of Ts on the non-template strand of the gene [32].

2.1.6 RNA Processing

The majority of eukaryotic RNAs require extensive modifications before they attain their mature structure and function. RNA strands may be modified by (1) the removal of RNA sequences, (2) the addition of RNA sequences, or (3) the covalent modification of specific bases. The long, relatively unstable mRNA precursor strand (hnRNA) is synthesized and remains in the nucleus where it is subjected to several stability-enhancing processes. With the exception of mitochondrial mRNAs, the 5′-ends of eukaryotic mRNA precursor molecules are capped, which involves removal of the terminal phosphate group of the 5′-nucleoside triphosphate and subsequent linkage of the 5′-diphosphate group to a GTP molecule [33]. The cap structure is covalently modified by methylation of the newly added guanine. The hnRNA is also modified at the 3′ end by the addition of a poly-A tail, a string of 50–250 adenine residues. The poly-A tail serves to extend the life of the mRNA by protecting the 3′-end of the molecule from 3′-exonucleases, and may also act as a translational enhancer [34]. The mRNA molecule is stabilized further by the association of a ~70 kDa protein with the poly-A tail [35].

The majority of protein-encoding eukaryotic genes contain intervening sequences, termed introns, which do not encode any portion of the protein. These introns, which may be between 65 and 10,000 nucleotides in length, are maintained during transcription of the hnRNA molecule, resulting in production of a long hnRNA that must be modified in order to become a continuous template for synthesis of the encoded protein. Maturation of the hnRNA requires a splicing event in which the introns are removed in conjunction with the joining

of the coding sequences, termed exons. Three short consensus sequences are necessary for splicing to occur; two are found at the intron–exon boundary at both the 5′ and 3′ of the intron and the third is found within the intron near the 3′-end [36]. The consensus sequences mark splice sites and act as targets for the spliceosome, a large multisubunit protein complex comprised of 45 proteins and thousands of snRNA [36]. The spliceosome catalyzes removal of introns and rejoining of exons, ultimately resulting in the formation of a protein-encoding mRNA. Some genes encode for more than one protein, which is accomplished by alternative splicing of the primary hnRNA transcript. One mechanism of alternative splicing involves removal of one or more exons during splicing when the spliceosome ignores one or more intron–exon boundaries [37]. Alternative transcripts may also be generated by the use of a secondary polyadenylation site [38].

RNA processing is not limited to mRNA. Eukaryotic tRNAs are modified posttranscriptionally, as are eukaryotic rRNAs [39]. Introns present within rRNAs are classified as Group I or Group II introns. Group I introns are removed as linear molecules by a self-splicing mechanism that requires magnesium and guanosine as cofactors [40]. Group II introns are also self-splicing, are removed as a lariat structure, and require spermidine as a cofactor [41].

2.2 Basic Molecular Analysis and Interpretation

Investigations into the molecular mechanisms of disease depend on the analysis of cellular and often times microbial DNA and RNA to identify and characterize the genes involved. Target genes can be identified, localized to specific chromosomes, amplified by cloning, and subjected to sequence analysis. DNA analyses are used practically in the identification of individuals for forensic or parentage testing, detection of gene mutations, amplifications, and deletions, associated with disease and RNA analyses to characterize gene expression in various forms of human cancer.

2.2.1 Isolation of Nucleic Acids

In theory, nucleic acids can be isolated from any tissue that contains nucleated cells. Common starting material for nucleic acid isolations include blood and other body fluids, cultured cells, buccal swabs, and a variety of fresh, frozen, or formalin-fixed, paraffin-embedded (FFPE) tissue specimens. Although isolation may need to be optimized for any particular tissue type or sample and application, there are basic steps in the extraction process that are universal: the cells must be lysed, proteins and cellular debris must be removed, and the nucleic acid must be made available in solution.

FFPE tissue is the most common specimen type used when interrogating solid tumors for somatic changes in the clinical

setting. FFPE tissue is readily available following tumor resection or biopsy, is easily stored at room temperature and the tissue being used for molecular analysis can be viewed microscopically to determine the precise histological diagnosis and to determine the adequacy of the specimen (percent tumor cells, exclude excessive necrosis, etc.). Nucleic acid derived from FFPE tissue, however, can be very problematic because the fixation process degrades nucleic acid and creates cross-links between nucleic acids and proteins. Molecular analysis of FFPE tissue, therefore, requires special attention during isolation and the design of downstream molecular applications. For example, the majority of an FFPE DNA sample is fragmented to 100–200 bp and will not work at all in most assays designed for high quality DNA. Assays specifically designed to detect these smaller fragments must be used in when FFPE-derived DNA is needed.

Before nucleic acids can be separated from the rest of the cellular contents, the cells must be lysed. This is often accomplished in some type of lysis buffer that can contain agents such as proteinase K and sodium dodecyl sulfate which are needed to digest structural proteins and to disrupt the cellular and nuclear membranes. Some samples such as solid tissues often require additional steps first, such as mincing, homogenization, or freezing the sample in liquid nitrogen followed by pulverization with a mortar and pestle to obtain a complete lysis. After the sample has been adequately lysed, the protein component in the sample must be removed. This can be accomplished in several ways, including: (1) adding a phenol–chloroform–isoamyl alcohol mixture to the sample which contains an organic phase and an aqueous phase whereby proteins and lipids dissolve into the organic phase and nucleic acids are retained in the aqueous phase of the mixture; (2) precipitation of proteins; (3) absorbance of nucleic acids to a column or magnetic bead and washing away proteins.

After the DNA has been separated from the rest of the other cell components, it must be transferred to a suitable buffer such as tris–EDTA (TE) or deionized water. This can be accomplished by adding ethanol or isopropanol which causes the DNA to precipitate out of solution. The DNA can then be pelleted by centrifugation, and resuspended in deionized water or a buffer. Alternatively, nucleic acids can be eluted directly from columns or magnetic beads in some of the more automated methods.

Due to the time involved and the toxicity and waste disposal issues associated with phenol, many laboratories choose other isolation protocols that remove proteins by other means. After lysis of the cells a protein precipitation solution can be added followed by centrifugation to separate the DNA from the cellular proteins. DNA can then be further purified by the use of silica gel particles in suspension or bound in a column. These silica columns can be purchased from numerous companies and are available in a wide range of formats and sizes, including the popular spin columns which rely on centrifugation to move solutions through the column. DNA binds to the silica gel as it passes through the column, allowing unwanted

salts and contaminants to wash through. After the DNA is bound to the column it can be washed and then eluted off the column with water or a low salt buffer such as TE. Similar methods have been developed using magnetic beads capable of binding DNA in solution. DNA can be purified by binding it to these beads in a tube. When these tubes are placed next to a magnet, the beads can be held in place on the side of the tube while the original solution and contaminants are removed. This process can then be repeated with wash solutions and finally an elution buffer that is often heated to facilitate removal of the DNA from the beads.

Although these alternative methods can be faster, they are often more expensive and can result in lower yield and slightly lower quality DNA. However, this is not generally a problem for most molecular techniques commonly used today. Another improvement on traditional nucleic acid isolation methods has been the introduction of automated isolation systems which can greatly reduce the hands-on time needed for isolating DNA. Once loaded, these instruments can go through the entire DNA isolation procedure, often using some form of magnetic separation. Although these automated systems are much more expensive than manual methods, the time they save will often make these systems a valuable investment for laboratories that process larger numbers of samples.

RNA isolation methods are similar to those for DNA. However, when isolating RNA extra care must be taken to avoid RNA degradation due to its susceptibility to RNA-digesting enzymes (RNases). To avoid RNase digestion of the sample, it is often necessary to work quickly and to work in a clean area. Care should be taken to ensure that all solutions purchased for use in these extractions have been designated as “RNase-free” or treated with special chemicals such as diethylpyrocarbonate (DEPC) that inactivate RNases.

2.2.2 Use of Enzymes in Molecular Biological Techniques

DNA or RNA isolation techniques only produce the starting material needed for molecular evaluation of nucleic acid. Almost every method used to examine DNA or RNA relies heavily on the use of enzymes, often isolated from various microbial species. These enzymes can be exploited to perform various tasks such as selectively cutting DNA or RNA into smaller pieces or digesting them completely, modifying the ends of DNA strands to remove or add phosphate groups, ligating two strands of DNA together to form one, and replicating nucleic acids.

2.2.2.1 Nucleases

Nucleases make up a diverse group of enzymes that are able to cleave or digest nucleic acids. This group of enzymes can be divided into various classifications based on the type of nucleic acid they use as substrates and the way in which they digest.

Exonucleases begin digesting nucleic acids at the ends while endonucleases make internal cuts in a nucleic acid molecule. Restriction endonucleases make up a very large group of enzymes found in bacteria that recognize very specific sequences of DNA, often six base pairs in length. For example, the restriction endonuclease EcoRI (isolated from *E. coli*) recognizes the sequence GAATTC and makes a single cut within this recognition site. If the sequence of a particular piece of DNA is unknown, observing the digestion patterns of various restriction enzymes can provide valuable information about its sequence. For example, if a particular mutation occurs within the recognition sequence of a specific restriction enzyme, that enzyme may not cut when the mutation is present.

When complete digestion of DNA or RNA is desired, for example when DNA is being isolated and contaminating RNA is not desirable, RNase can be used to selectively digest the RNA, leaving a pure population of DNA. In the same way, DNases can be used to digest unwanted DNA molecules in an isolation of RNA.

2.2.2.2 DNA Ligases, Kinases, and Phosphatases

In addition to digesting DNA, enzymes are also available which facilitate the joining or ligation of two or more DNA fragments. Enzymes with this capability are called DNA ligases and commercially, T4 DNA ligase is widely available and most often used in molecular laboratories. Since DNA ligases work by creating a phosphodiester bond between the 5' phosphate and 3' hydroxyl termini of two DNA strands, the presence of the 5' phosphate is essential. DNA phosphatases, such as calf intestinal phosphatase or shrimp alkaline phosphatase, remove the phosphate group on the 5' end of a DNA strand, preventing ligation; a DNA kinase, T4 polynucleotide kinase, can restore the 5' phosphate in order to allow ligation.

2.2.2.3 Polymerases

DNA and RNA polymerases are enzymes that synthesize new strands of DNA or RNA, normally using a preexisting strand of DNA (or RNA) as a template. Reverse transcriptase is another polymerase that is unique in its ability to use RNA as a template for synthesizing DNA. DNA polymerases are essential for the polymerase chain reaction (PCR), a technique described below that has revolutionized molecular biology.

2.2.3 Electrophoretic Separation of Nucleic Acids

Gel electrophoresis of DNA is widely used in procedures such as Southern blotting and PCR in which separation or sizing of a population of DNA fragments is an essential step in the analytical process. In some cases, a significant amount of information can be gleaned from relatively simple electrophoretic procedures that take advantage of the various properties of DNA molecules (i.e., charge, size, and conformation).

Fig. 2.2 Agarose gel electrophoresis of human genomic DNA.

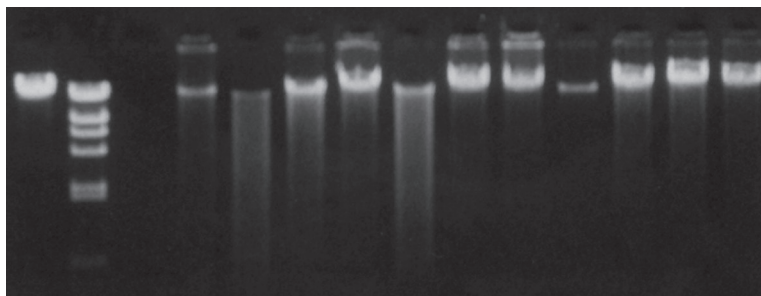
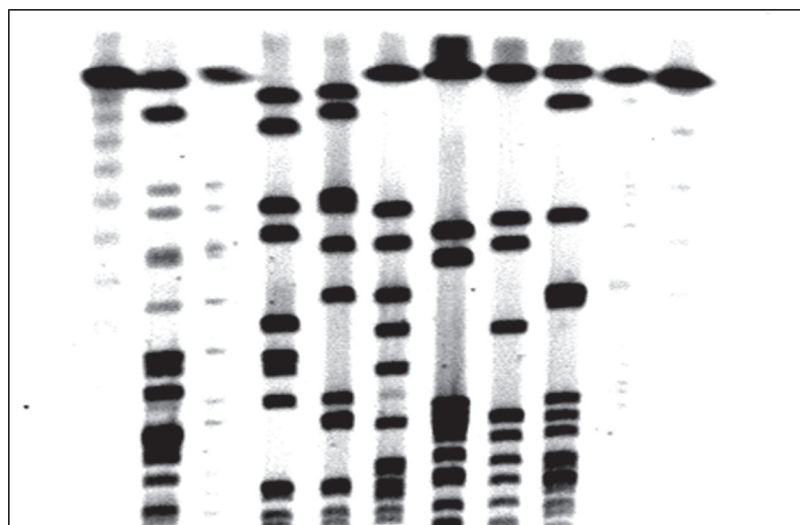


Fig. 2.3 Pulse field gel electrophoresis for strain typing of *Staphylococcus aureus*.



Gels employed in electrophoretic techniques are generally composed of agarose or polyacrylamide, which act as a matrix through which DNA (or RNA) will travel when an electric current is passed through the gel. Generally, larger fragments of DNA will migrate more slowly through the gel and smaller fragments will migrate more quickly. DNA of a given size will travel through the gel at the same rate and form a band that is visualized by staining the gel with a dye such as ethidium bromide, which binds to DNA and fluoresces when viewed under UV light (Fig. 2.2). Agarose gels are useful in most traditional molecular techniques for visualizing DNA greater than 100 base pairs. However, when smaller DNA fragments are to be analyzed, special agarose formulations or polyacrylamide gels are needed for proper resolution and visualization.

When larger fragments of DNA need to be resolved, a technique called pulsed-field gel electrophoresis can be utilized. Conventional gel electrophoresis techniques are not useful for separation of extremely long pieces of DNA, because the constant current eventually unravels the DNA strands completely so that they travel, end first, through the gel at a rate that is independent of their length. Pulsed-field gel electrophoresis (PFGE) overcomes this challenge by periodically switching the orientation of the electric fields

with respect to the gel, thus preventing the DNA strands from losing secondary structure and allowing long strands to be size-differentiated [42]. PFGE is often used in the identification of pathogens [43], i.e., differentiating between strains of bacteria (Fig. 2.3). Other applications include chromosomal length polymorphism analysis, and large-scale restriction and deletion mapping in DNA that is hundreds or thousands of kb in length. The effectiveness of PFGE is dependent on high-integrity starting material, and DNA that is degraded or sheared will not yield informative results. High quality DNA is often generated by embedding the cells of interest in agarose plugs, lysing cell membranes with detergent, and removing proteins enzymatically, leaving intact DNA which can be digested by restriction enzymes *in situ* and easily loaded into a gel apparatus [42, 44]. In a typical analysis, PFGE will produce a pattern of DNA fragments that range in size from 10 to 800 kb.

Electrophoresis of RNA to check the quality of a sample or for blotting techniques is typically accomplished by electrophoresis through a 1–2 % agarose gel containing formaldehyde, which maintains the denatured state of the RNA strands. When secondary structure of an RNA molecule is to be investigated, samples are subjected to non-denaturing polyacrylamide gel electrophoresis.

2.2.4 Denaturation and Hybridization

The majority of techniques used in molecular analyses require single-stranded nucleic acids as starting material. Prior to hybridization with a complementary nucleotide sequence such as an oligonucleotide primer or a nucleic acid probe, double-stranded DNAs and single-stranded RNAs must be denatured to generate single strands and eliminate secondary structure. Denaturation of nucleic acids is rapid, and can be induced by various conditions, including extremes of pH ($\text{pH} < 4$ or $\text{pH} > 10$), hydrogen-bond disrupting agents (such as urea or formamide), or heat [45]. Heat is the most commonly used means of disrupting the hydrogen bonds that occur between complementary strands of double-stranded DNA to produce single-stranded DNA. The melting temperature (T_m) of a specific double-stranded DNA sequence is reached when the disruption of the hydrogen bonding in a population of DNA causes 50% of the DNA to become single-stranded. Since more hydrogen bonding occurs between G and C as compared to A and T, the T_m will be greater for DNA containing a larger percentage of G and C nucleotides. The denaturation process is easily monitored by spectrophotometry, since the absorbance of the DNA at 260 nm increases as denaturation progresses [45].

Hybridization of nucleic acid strands is a relatively slow process and the rate is governed by the relative concentration of strands with complementary sequences and by the temperature. When two complementary strands are aligned properly, hydrogen bonds form between the opposing complementary bases and the strands are joined. Target nucleic acid sequences can hybridize to a complementary DNA (cDNA) or RNA strand, or to other complementary sequences such as oligonucleotide primers or nucleic acid probes. Hybridization between two complementary RNA molecules is strongest, followed by RNA–DNA hybrids, and DNA–DNA hybrids [46].

2.2.5 Concepts and Applications of Southern Blotting

Southern blotting is a technique that relies heavily on the concepts of denaturation and hybridization of DNA. Southern blotting has been used in clinical or forensic settings to identify individuals, determine relatedness, and to detect genes associated with genetic abnormalities or viral infections. Southern blot analysis is also used in basic scientific research to confirm the presence of an exogenous gene, evaluate gene copy number, or to identify genetic aberrations in models of disease. Although this technique has been replaced to a large extent by newer, faster techniques, it still has its place in a molecular laboratory.

The first step in successful Southern blotting is to obtain DNA that is reasonably intact (Fig. 2.4). DNA that has been

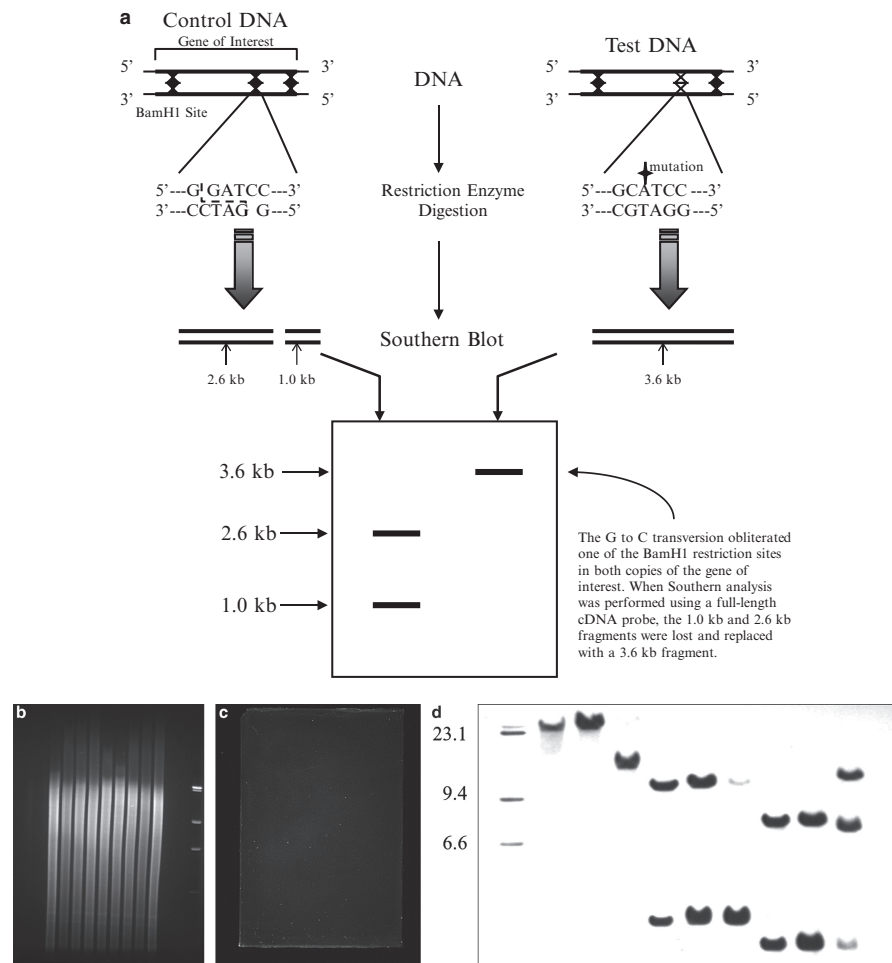
degraded by excessive exposure to the elements or mishandling will not produce a good quality Southern blot because it cannot be fragmented uniformly prior to the blotting procedure. The test DNA must be fragmented with restriction enzymes, which cut the double strands of DNA at multiple sequence-specific sites, creating a set of fragments of specific sizes which represent the regions of DNA between restriction sites. The fragmented DNAs are size-fractionated via agarose gel electrophoresis and are subsequently denatured, which enables them to later be hybridized to complementary nucleic acid probes. The DNA from the gel is transferred to a solid support such as a nylon or nitrocellulose membrane via capillary action or electrophoretic transfer and is bound permanently to the membrane by brief UV crosslinking or by prolonged incubation at 80 °C. Blots at this stage may be stored for later use or may be probed immediately. For detailed protocols, the reader is referred to other sources [46, 47].

Interpretation of the Southern blot is easy when the question is whether or not a gene is present in a particular sample, as long as appropriate positive and negative controls are included (a sample of DNA known to be positive and a sample of DNA known to be negative). The presence of multiple copies of a gene indicates gene amplification, which may occur in oncogenes during cancer development [48, 49]. Amplifications are obvious on a Southern blot as a band or bands that are more intense than the normal single-copy control; numerical values that reflect intensity may be assigned to bands using a densitometer. Structural aberrations in a gene of interest can be detected by Southern blotting, including the insertion or deletion of nucleotides or gene rearrangements (Fig. 2.4). When nucleotides are mutated, inserted, or deleted, the ladder of fragments produced may be abnormal due to the obliteration of restriction sites, the generation of novel restriction sites, or alterations in fragment size due to an increase or decrease in the number of nucleotides between restriction sites. The majority of these aberrations are apparent on Southern blots as abnormal banding patterns. The Southern blot remains a useful and reliable way to obtain definitive data on gene structure.

2.2.6 Polymerase Chain Reaction (PCR)

The development of PCR has increased the speed and accuracy of DNA analysis, and has resulted in the rapid development of new and creative techniques for detecting, replicating, and modifying DNA. Since it was described originally [52, 53], PCR has evolved to encompass an enormous array of specific applications. This section will cover the basic concepts of PCR and several applications that are useful in molecular analyses of cancer. For a complete technical description of PCR techniques, the reader is referred to a more detailed source [50].

Fig. 2.4 Southern blotting. Variations in DNA sequences that alter restriction endonuclease recognition sites can be detected by Southern blotting (a). After digesting genomic DNA with restriction endonucleases, the DNA is subjected to agarose gel electrophoresis and then transferred to a membrane. The agarose gel can be imaged before (b) and after (c) the transfer has taken place to confirm adequate amounts of digested DNA and proper transfer of the DNA from the gel to the membrane. The membrane is then incubated with radiolabeled or chemiluminescent DNA probes which hybridize to specific DNA fragments. Exposing the membrane to film allows for visualization and sizing of these fragments (d).



2.2.6.1 Principles of PCR

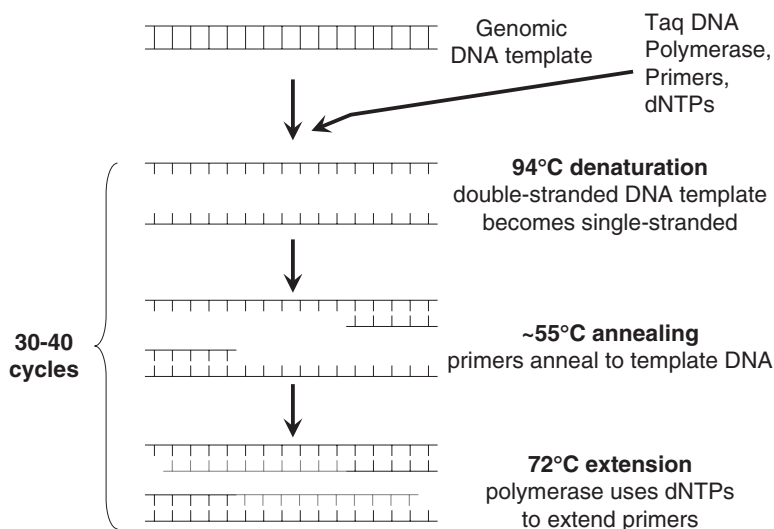
Every PCR reaction must contain several key components: a small amount of target DNA to be used as a template; a thermostable polymerase, such as *Taq* polymerase that is not denatured at high temperatures; a pair of single-stranded DNA oligonucleotide primers; and deoxynucleotides (dNTPs), the building blocks for the DNA to be amplified. The DNA may be genomic DNA isolated directly from experimental or patient material, or it may be cDNA that has been synthesized from an RNA template by reverse transcriptase. The target for any PCR reaction is dictated by the specific oligonucleotide primers used. Two primers are designed to anneal to sites at either end of the region of interest, on opposite template strands (Fig. 2.5). The primers are extended in the 5'–3' direction by DNA polymerase to yield overlapping copies of the original template. PCR is a cyclic process, consisting of three steps: denaturation of template (94 °C), annealing of primers (temperature is sequence dependent; often between 50 and 60 °C), and extension of primers (72 °C). The three steps are repeated, with each cycle resulting in amplification of the target sequence. By the end of the third cycle, a new double-stranded molecule is formed in which the 5'- and 3'-ends coincide exactly with

the primers [51, 52]. The copies that are produced of this targeted DNA sequence are referred to as the amplicon. These double-stranded molecules accumulate exponentially during subsequent cycles of PCR, so that the majority of products are of a defined size and are seen clearly as a sharp band upon electrophoretic separation. Due to the incredible sensitivity of PCR, even a miniscule amount of DNA can be amplified, which makes it a powerful tool but also mandates that precautions are taken to avoid introduction of contaminating DNA which could result in misinterpretation.

2.2.6.2 Design of Primers for PCR

When constructing primers for PCR, it is important to keep in mind a few basic concepts. Primer length can influence target specificity and efficiency of hybridization. As a general guideline, primers should be 20–30 nucleotides in length. Whenever possible, both primers should be the same length because primer length is considered when calculating an appropriate annealing temperature. The base composition of the primers is also important, since annealing temperature is governed in part by the percent G+C content of the primers. Ideally, G+C content is between 40–60%, and the percent G+C should be the same in any primer pair. A simple formula can be used to

Fig. 2.5 The polymerase chain reaction (PCR). Specific sequences of DNA are amplified by PCR using pairs of oligonucleotide primers. Genomic DNA, primers, *Taq* DNA polymerase, and deoxynucleotides (dNTPs) are combined in a suitable buffer and subjected to repeated cycles of alternating temperatures which denature the template DNA, allow for primer annealing to the template and finally extend the primer to create a new strand of DNA.



calculate an appropriate annealing temperature for any given primer: $T_m = 69.3 + 0.41(\%G + C) - (650/L)$, where L = primer length in bases [53]. Repetitive or palindromic sequences should be avoided in a primer to avoid self-hybridization, and primer pairs should not contain sequences complementary to each other to avoid primer dimers.

2.2.6.3 The Role of Polymerase in PCR

A DNA polymerase enzyme is essential for the primer extension step of PCR. Early PCR experiments employed the Klenow fragment of *E. coli* DNA polymerase I, but this enzyme is heat labile and must be replenished with each amplification cycle. The developments of thermostable DNA polymerase and commercially available thermal cyclers have greatly improved the efficacy of PCR methodology. *Taq* DNA polymerase was isolated from *Thermus aquaticus*, and is characterized by its 5′–3′ exonuclease activity, thermostability, and optimum performance at 70–80 °C [54, 55]. Temperature, pH, and concentration of Mg^{++} influence the activity of *Taq* polymerase. Lower divalent cation (Mg^{++}) concentrations decrease the rate of dissociation of enzyme from template by stabilizing the enzyme-nucleic acid interaction [56]. The optimum pH for a given PCR reaction will be between 8.0 and 10.0 (usually ~8.3), but must be determined empirically. While *Taq* DNA polymerase is ideal for routine PCR, there are many other DNA polymerases with unique qualities which make them useful for special PCR applications such as amplification of long stretches of DNA or high-fidelity amplification [56].

2.2.6.4 Detection of PCR Products

Once the PCR is complete, the products must be analyzed and interpreted. Amplification products of routine PCR reactions can be separated by standard agarose gel electrophoresis and visualized by staining with ethidium bromide or other

DNA dye. When finer resolution is needed, such as in the analysis of very small (<100 bp) products, polyacrylamide gel electrophoresis is standard. In addition, newer technologies such as bead arrays, microarrays, and capillary electrophoresis can be used to detect PCR products (Fig. 2.6).

2.2.6.5 Contamination Issues

Due to the extreme sensitivity of PCR, care must be taken to avoid exposing PCR reagents, set-up areas, and equipment to DNA that is not intended to be a part of the PCR reaction, especially amplicons from previous PCR runs. Unwanted false-positive results can occur if issues of contamination are not addressed. One laboratory practice that is often used to avoid this type of contamination issue is designating separate pre- and post-PCR work areas. After a PCR run is complete, all handling of the amplicon, such as for gel electrophoresis, is performed in an isolated work area with equipment that is designated for post-PCR use only. Pipettes, pipette tips, and buffers, for example, used in post-PCR handling should not be outside of the area designated for post-PCR work. Gloves used in this area should also be discarded before returning to other work areas in the lab. Additionally, the use of filtered pipette tips, especially in the pre-PCR work area, can help in eliminating PCR contamination issues.

2.2.7 Modifications and Improvements of PCR

Many variations to the original PCR procedure have been developed to improve its utility, ease, sensitivity, and specificity. Many companies now offer PCR master mixes which contain a DNA polymerase and dNTPs in buffers optimized to amplify most target sequences. This often eliminates the need for exten-

sive optimization of PCR components and cycling conditions and simplifies PCR set-up by only requiring the addition of a DNA template and primers to the master mix. In order to eliminate some types of PCR contamination, some master mixes are available that contain the enzyme uracil N-glycosylase (UNG) and the nucleotide dUTP along with the normal dNTPs needed for PCR. If a PCR is contaminated by an amplicon from a previous PCR made with dUTP, a pre-amplification incubation will degrade the contaminating amplicon, preventing it from being used as a template in the current PCR.

2.2.7.1 Hot Start PCR

Hot start PCR was developed to reduce background from nonspecific amplification by preventing polymerization of new DNA during the set-up and the initial phase of the reaction when nonspecific binding may occur between primers and other DNAs in the mixture [57, 58]. Hot start may be achieved by limiting the initial Mg^{++} , dNTP, or enzyme concentration, or by separating the components with a barrier, such as wax beads that melt as the mixture is heated. Alternatively, hot start polymerases bound by antibodies or

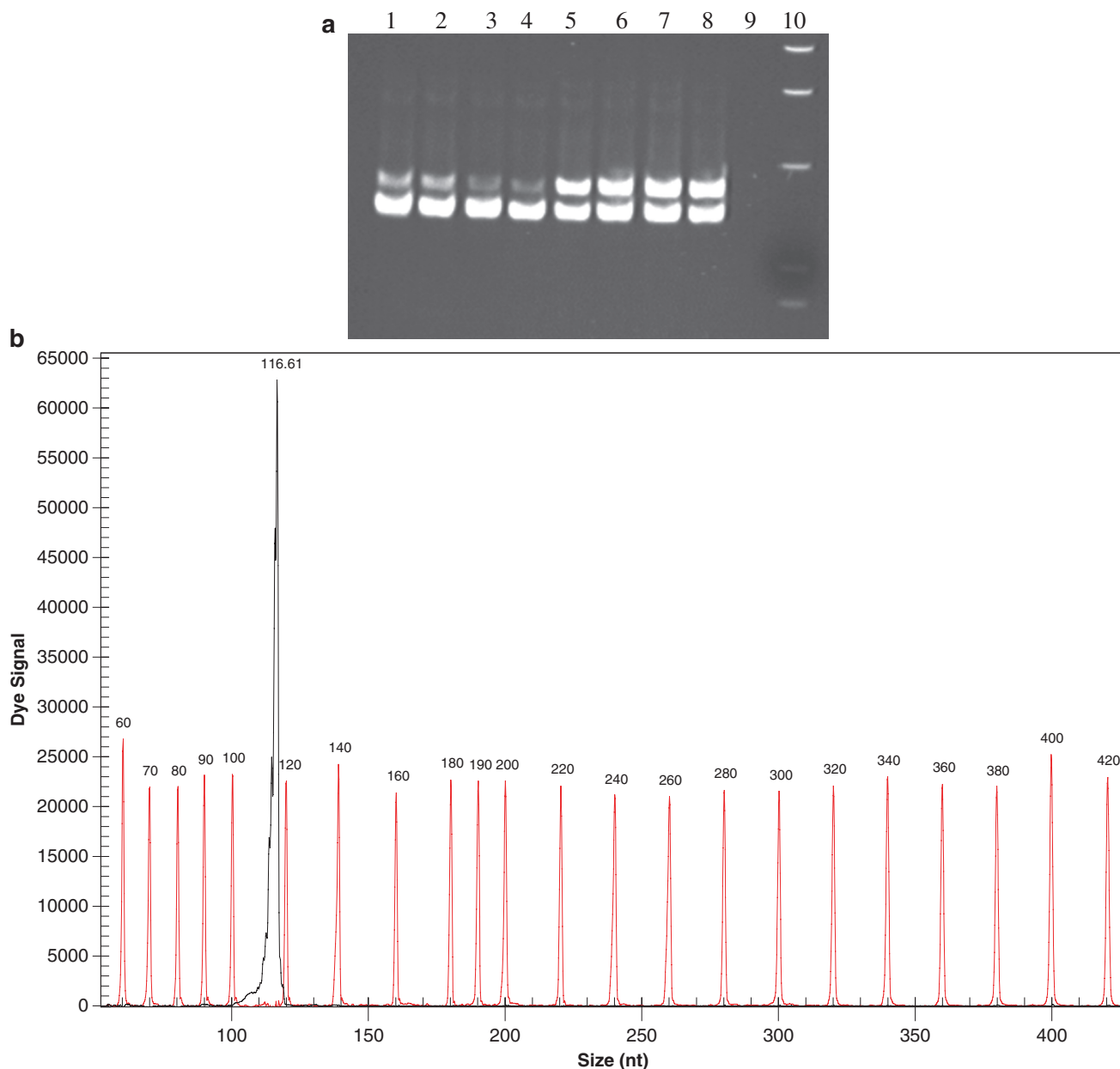


Fig. 2.6 Methods for analysis of DNA amplified by PCR. Gel electrophoresis separates DNA fragments based on size since smaller DNA fragments migrate at a faster rate through the gel (a). When DNA of an unknown size is run alongside a “ladder” containing DNA fragments of known sizes, the length of the unknown DNA can be estimated. More precise sizing of DNA can be made when using capillary electrophoresis

(b), in which fluorescently labeled DNA (*black peak*) is moved through thin capillary tubing along with a DNA sizing standard (*red peaks*) to obtain higher resolution than possible with standard gel electrophoresis. Alternatively PCR-amplified DNA can be analyzed based on sequence due to its ability to hybridize with probes in microarrays and bead arrays (c).

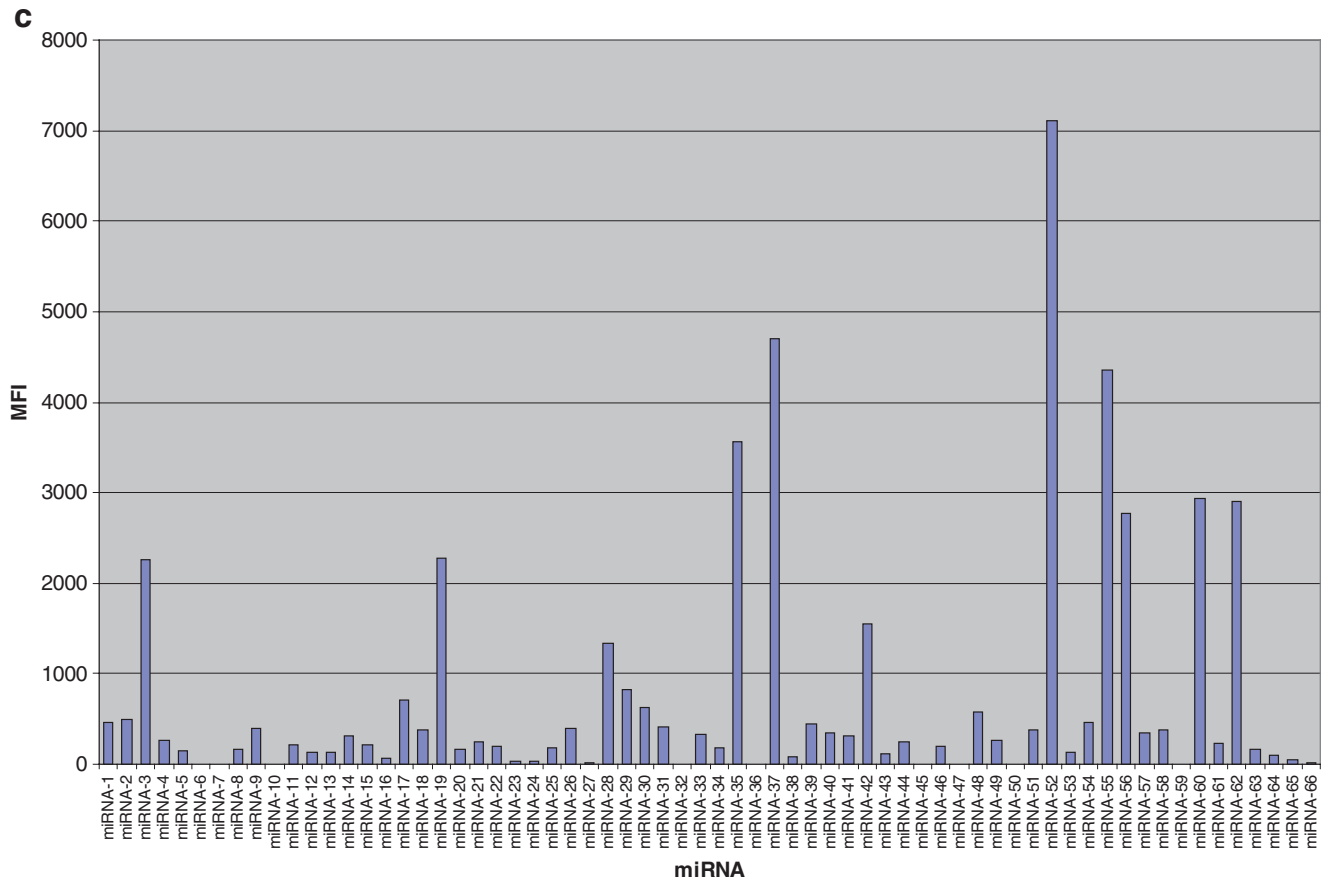


Fig. 2.6 (continued)

other molecules are enzymatically inactive until a high temperature incubation (immediately preceding normal PCR cycling) activates the polymerase, thereby preventing premature or nonspecific DNA amplification in the time between mixing the components of the reaction and the PCR cycling.

2.2.7.2 Touch-Down PCR

Touch-down PCR was developed to enhance amplification of the desired target sequences while reducing amplification of artifacts [59, 60]. The initial cycle begins with an annealing temperature that is greater than the expected T_m of the primer and the annealing temperature is lowered progressively with each cycle. As a result, the desired amplicon will accumulate preferentially while the amplification of undesired products is minimal.

2.2.7.3 Nested PCR

Performing nested PCR can increase both the sensitivity and specificity of amplification [52]. The amplification product(s) generated in the first PCR reaction are used as the template for a second PCR reaction, in which primers are used that are internal, or nested, within the first primer pair. Nonspecific products that are produced within the first round of PCR are not likely to contain sequences complementary to the nested primers, so that spurious products are eliminated during the

second round of PCR. Extremely rare target sequences can be detected using nested PCR, since the first round of PCR effectively amplifies the specific template for the second round of PCR. Due to the sensitive nature of nested PCR, special care must be taken to avoid contamination.

2.2.7.4 Long-Distance PCR

It is possible to amplify sequences as large as 50 kb using long-distance PCR (LD-PCR) [61]. One step toward successful LD-PCR is the use of thermostable, long-life polymerases that are capable of generating long strands of cDNA. The first LD-PCR was accomplished by using a 5'-endonuclease-deficient, N-terminal deleted variant of *Taq* DNA polymerase in combination with Pfu DNA polymerase in a 180:1 ratio [62]. Many special DNA polymerases capable of performing well in LD-PCR are now available commercially. Other prerequisites for successful LD-PCR are high quality DNA for use as template, and carefully constructed primers with matching melting temperatures.

2.2.7.5 Reverse Transcriptase PCR (RT-PCR)

Specific RNA sequences can also be detected using PCR. An enzymatic reaction using reverse transcriptase creates a complementary DNA (cDNA) copy of the RNA, which can then be amplified using standard PCR conditions. RT-PCR is

a useful tool for detecting the presence or absence of mRNAs for specific genes and can be performed quantitatively or semiquantitatively to detect variations in gene expression between samples.

2.2.7.6 Multiplex PCR

Often it is desirable to amplify more than one target sequence at a time. The use of more than one pair of PCR primers in a single reaction is often possible, although these multiplexed reactions become increasingly difficult to optimize with increasing numbers of amplicons. Designing primers with similar melting temperatures helps ensure successful multiplexed PCRs and using Mg^{++} concentrations that are higher than normal PCRs is usually necessary.

2.2.7.7 Quantitative PCR

Quantitative PCR provides a quick and simple alternative to Southern blot analysis for the evaluation of gene copy number or gene expression levels [63]. The underlying premise for quantitative PCR is that the accumulation of amplified products occurs exponentially and follows a predictable curve. The overall profile of product accumulation throughout the course of a reaction may be reproducible enough to extrapolate the amount of starting material. Accurate quantification requires that the analysis be done during the exponential (sloped) part of the amplification curve, and not during the plateau phase when the DNA amplification rate has leveled off. Accurate quantitative PCR experiments using traditional (endpoint) PCR methods must include control template fragments, which may be synthesized or may be isolated from other sources. The control fragments should have priming sites and secondary structure that is identical to the test DNA, but should be sufficiently different in size that they can be discriminated upon electrophoretic separation. In a typical quantitative PCR reaction, replicate tubes are prepared with a fixed concentration of test DNA; then known quantities of control DNA are added in a range of concentrations, PCR is conducted, and the samples are subjected to gel electrophoresis. When one template is in excess, it will yield a greater abundance of PCR product; but when the concentration of both the control and test templates are equal, amplification will occur at equal rates, ultimately producing two bands of equal intensity on the gel. Quantitative approaches to PCR are useful when careful attention is paid to experimental reproducibility [63, 64]. Although this method can achieve quantitative results, it is rather tedious and has been replaced in most situations by one of the most useful modifications of the basic PCR method: real-time PCR.

2.2.8 Real-Time PCR

Previously, DNA amplified by PCR was analyzed in a post-PCR process. Newer technologies allow users to observe the amplification of DNA as each cycle occurs, in real-time.

Real-time PCR eliminates the need for post-amplification analysis, often saving hours of analytical time and minimizing the risk of contamination. Real-time PCR also increases the ease with which quantitative results can be obtained and thus be applied to gene dosage and expression analysis. For this reason the term “quantitative PCR” (qPCR) has become synonymous with real-time PCR.

Real-time PCR is made possible by the use of fluorescent signals which increase as the target sequence is amplified. This can be accomplished in several ways, but the two main methods include the addition of SYBR Green (or similar dyes that emit fluorescence in the presence of double-stranded DNA) to a standard PCR and using fluorescent probes specific for sequences between PCR primers.

SYBR Green is a dye that fluoresces only in the presence of double-stranded DNA. As the amount of amplicon increases during each PCR cycle, the amount of fluorescence also increases. Measuring the amount of fluorescence at each PCR cycle makes it possible to graph the amplification that occurs during the PCR (PCR cycle number on the *X*-axis and fluorescence on the *Y*-axis) showing the extent of amplification over time (Fig. 2.7) [65]. Since any double-stranded DNA will be detected by SYBR Green, precautions must be taken to ensure any increase in fluorescence represents the desired amplicon from the target sequence. In standard PCR this is accomplished using gel electrophoresis to make sure the length of the amplicon matches the expected size and that no nonspecific PCR products appear on the gel.

A second and slightly more complex method for detecting amplification in real-time PCR uses oligonucleotide probes that are fluorescently labeled. The exact design of these fluorescently labeled probes can vary but they all contain a sequence complementary to a region of the target sequence between the forward and reverse primers. In one example of a fluorescent probe-based real-time PCR chemistry, 5' hydrolysis probes or TaqMan® probes an oligonucleotide probe is designed with a fluorescent molecule such as FAM bound on one end and a quencher molecule on the other end. The fluorescent signal is usually not detected due the close proximity of the quencher. During the annealing step of each PCR cycle this probe anneals to the target sequence along with the primers (Fig. 2.8). During the extension stage of each PCR cycle, the 5'–3' exonuclease activity of *Taq* polymerase digests the probe, separating the quencher dye from the fluorescent dye, which results in a detectable increase in fluorescence [66, 67]. This fluorescence can be used to create an amplification curve similar to the one produced when SYBR Green is used.

2.3 Microarray Technology

DNA and RNA microarrays have greatly expanded the amount of data that can be obtained in a single experiment, often using the same basic principles of hybridization as Southern blotting.

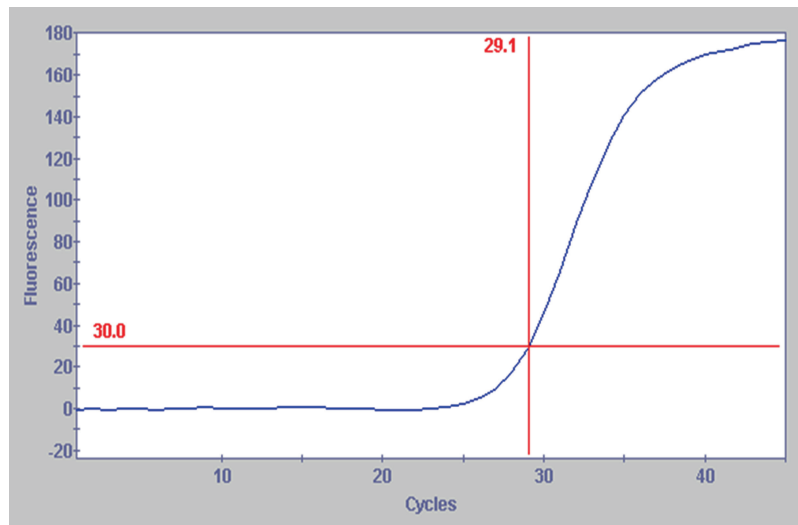
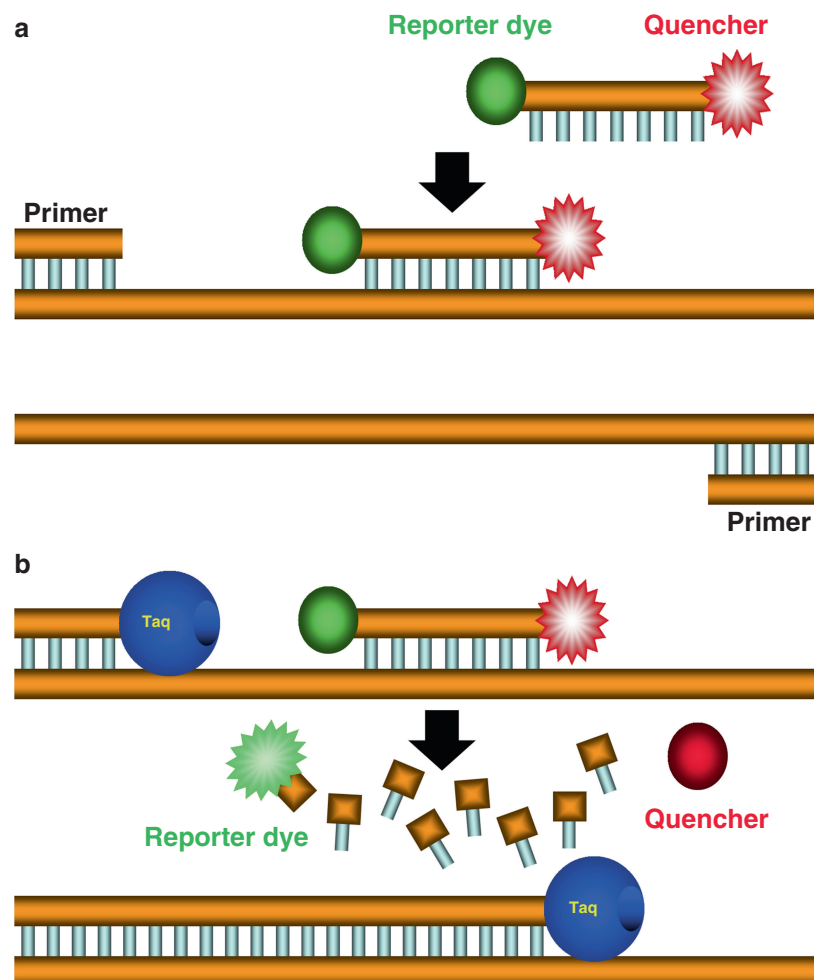


Fig. 2.7 Real-time PCR. Fluorescent detection of DNA being amplified by PCR can be in “real-time” by measuring the fluorescence emitted at each cycle. This fluorescence can be produced by adding a dye such as SYBR Green to the PCR buffer which fluoresces when bound to double-stranded DNA. Alternatively, fluorescence can be produced by adding sequence-specific probes that can be fluorescently labeled in various ways

to produce fluorescence at specific points during each PCR cycle. The exact or relative amount of starting template can be inferred by determining when the amplification curve (*blue*) crosses over a certain threshold (30.0) of fluorescence. The cycle when this occurs is designated as the cycle threshold value or Ct (29.1).

Fig. 2.8 Real-time PCR using hydrolysis (TaqMan) probes. PCR primers are used to amplify target DNA as in standard PCR. During the annealing step of each PCR cycle a fluorescently labeled oligonucleotide probe, complementary to one strand of the template DNA, anneals along with the primers (**a**). The TaqMan probe contains one detection dye (shown in *red*) whose fluorescence is absorbed by another quencher dye (shown in *green*) when the probe is intact. During the extension step of each PCR cycle (**b**) the TaqMan probes are digested. This physical separation of the dye from the quencher allows for increased fluorescence during the extension step of each PCR cycle.



Microarray designs vary greatly among manufacturers but in the more traditional arrays contain many thousands or millions of unique DNA probe sequences that appear as separate dots or features on a fixed surface. Each of these features contains a unique nucleotide sequence that represents a specific gene, chromosomal location or other nucleic acid sequence being interrogated. Labeled nucleic acids, prepared from the samples being analyzed, are incubated on the surface of the microarray to allow hybridization between the microarray probes and the sample-derived nucleic acid. A washing step is then used to remove any labeled sample that is not hybridized to reduce background signaling on the chip when scanned.

Numerous formats and applications for microarrays exist but three main microarray applications predominate in both the research and clinical laboratory setting: gene expression arrays, chromosome arrays (CMA), and genotyping arrays (Fig. 2.9). These three applications differ in the design of the microarrays and in the type of nucleic acid being interrogated [68–77]. In gene expression microarray applications, mRNA or cDNA is applied to a microarray containing features for large groups of genes, often all known genes expressed in a particular organism. The levels of gene expression can be determined by the intensity of the signals at each spot. Microarray results using mRNA isolated from two or more groups of cells or tissue can

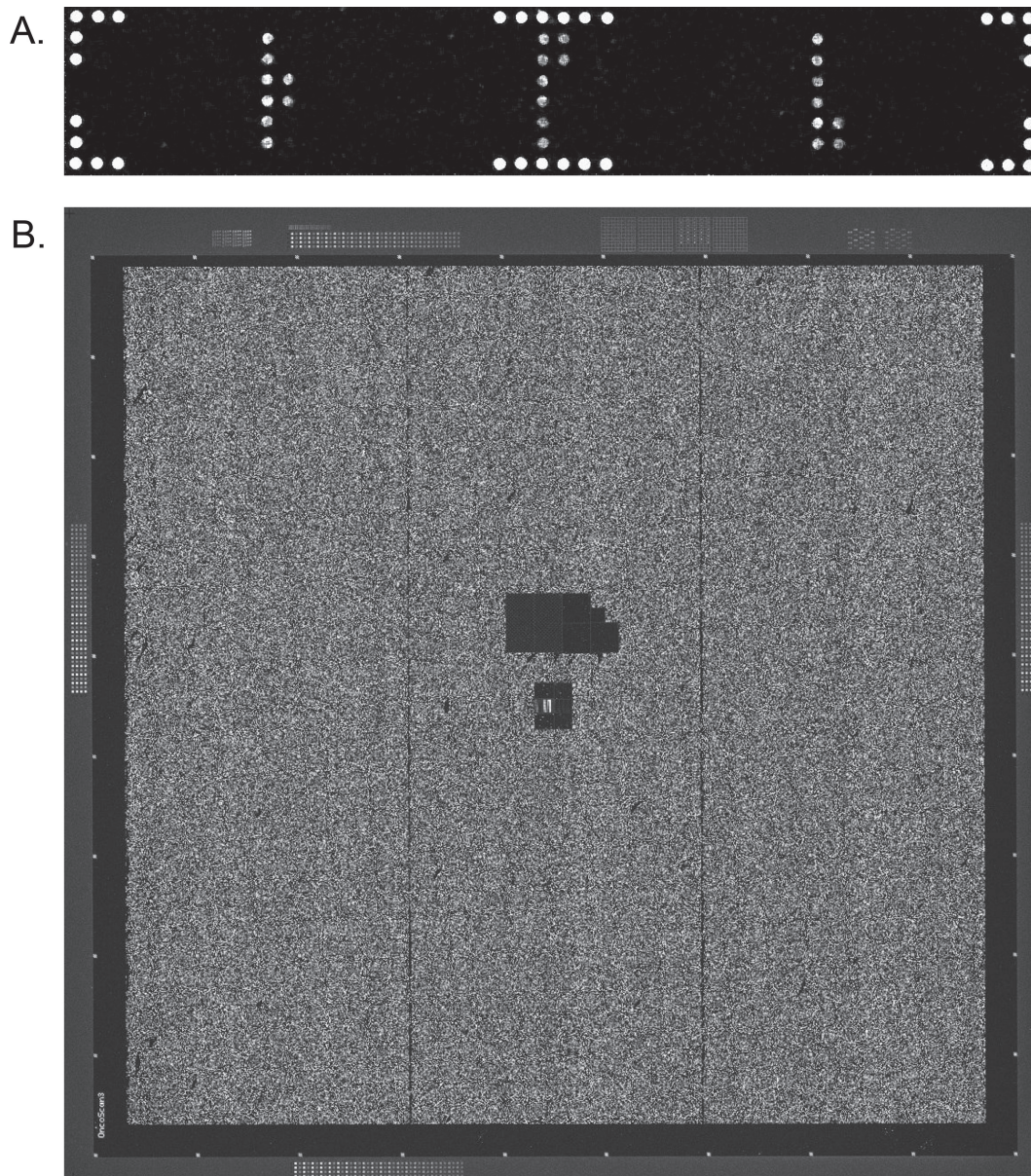


Fig. 2.9 Microarray technology. Genomic DNA or PCR-amplified DNA can be detected by hybridizing with probes that have been attached to specific locations on the surface of a microarray. Interrogation of a few or up to a few million different sequences in a DNA sample can be determined simultaneously. Genotyping of three loci is performed with a low-

density array developed by Nanosphere which can use PCR-amplified or non-amplified genomic DNA (A). High-density arrays such as the OncoScan microarray (Affymetrix, Inc.) can detect changes in copy number and loss of heterozygosity by targeting hundreds of thousands of single nucleotide polymorphism (SNPs) across the genome (B).

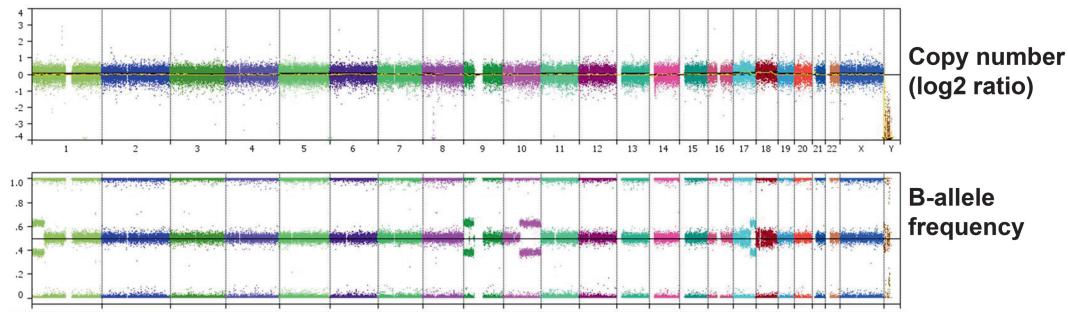


Fig. 2.10 SNP-based microarrays. Genomic DNA derived from formalin-fixed, paraffin-embedded tumor tissue can be used to detect changes in copy number and loss of heterozygosity. In this example, changes in copy number are not detected but regions of copy neutral

LOH can be observed within chromosomes 1, 9, 10 and 17. This example shows microarray data from OncoScan FFPE Assay kit (Affymetrix, Inc) analyzed with using Nexus software.

be compared to look for up- or down-regulation of each gene represented on the microarray [68, 69]. This is often used to compare normal and diseased tissue, varying grades of tumors, or samples treated and untreated with a drug or other agent in the research setting. In the clinical setting, gene expression microarrays are used with metastatic tumor in which the site of the primary tumor is unknown despite the available imaging and pathological findings. The expression profile of a tumor with an unknown primary is compared to known expression profile signatures of tumors in which the site of the primary tumor is known. Clinical expression profile tests may be helpful in this setting for predicting the site of origin and choosing the most appropriate form of therapy [70, 71]. In addition to mRNA from protein-coding genes, expression microarrays can also measure the expression of a large number of regulatory noncoding microRNAs or miRNAs.

While some microarrays provide information about the presence, absence and abundance of a particular target, genotyping microarrays are able to differentiate between two alleles at a given locus. This type of technology allows for the genotyping of many thousands or millions of single nucleotide variants (SNVs), also called single nucleotide polymorphisms (SNPs), in a given sample for a relatively small cost. The methodology varies between platforms but in some cases can be as simple as having two nearly identical oligonucleotide probes for each locus being genotyped. The sequence of each probe will differ near the middle of the probe at the position of the SNV with the probe at one feature annealing to one allele and the probe at another feature annealing to the other allele. Signal at one or both of these features indicates that a sample is homozygous or heterozygous for an allele with respect to each loci represented on the microarray. High-density genotyping arrays have been used frequently in genome-wide association studies (GWAS) but lower density genotyping arrays are more common in the clinical setting to evaluate a smaller number of clinically relevant loci [72, 73].

Genotyping microarrays can be designed to examine genomic DNA or select PCR-amplified targets for any number of single nucleotide polymorphisms, variants or muta-

tions [72]. In the clinical laboratory, these SNP microarrays have been used as a tool to create low to high density multiplexed genotyping assays [73, 74].

Chromosome microarrays (CMA) detect copy number variations (CNVs), such as deletions, duplications and gene amplifications in genomic DNA samples. These changes in copy number include trisomies or monosomies of entire chromosomes or smaller CNVs affecting only one arm of a chromosome or smaller regions of chromosomes. The dense coverage across each chromosome offered in many chromosome microarrays allows for detection of CNVs down to the range of 10–100 kb, which would be too small for detection by traditional cytogenetic techniques [75–77]. Depending on the design of the microarray used, CMAs can fall into one of two main categories. Copy number microarrays, commonly called array comparative genomic hybridization (aCGH), previously used BAC probes and now shorter oligonucleotide sequences to compare signal intensities across the genome of an unknown or patient same to that of a normal control genomic DNA sample. Alternatively, CMAs based on high-density SNP arrays similar to the high-density genotyping microarrays described above detect both copy number variations and long contiguous stretches of homozygosity. These stretches of homozygosity may suggest consanguinity, uniparental disomy (UPD) or in cancer, copy neutral loss of heterozygosity (cnLOH) (Fig. 2.10).

2.4 Sequencing Technologies

2.4.1 Sanger Sequencing

For many years DNA sequencing has been an essential tool in both the research and clinical laboratory. Until recently, DNA sequencing has been performed almost exclusively by Sanger sequencing technology in which target DNA, amplified by PCR or cloning into a bacterial vector, is sequenced by the extension of a single oligonucleotide primer by a DNA polymerase in the presence of both dNTPs and labeled dideoxynucleotide triphosphates (ddNTPs). The addition of

a ddNTP blocks the addition of any other dNTPs or ddNTPs, creating a somewhat random assortment of single-stranded DNA molecules of various lengths, each labeled with a fluorescent molecule corresponding to the ddNTP added on the 3' end (A, T, C, or G). These DNA fragments are separated, most commonly by capillary electrophoresis, to create an electropherogram from which the DNA sequence following the sequencing primer can be determined [78].

The sequencing of individual human genes and later, the entire human genome, led to the identification of countless sequence variants with pathological significance in the fields of genetics, hematology, and oncology. Likewise, sequencing portions of or entire genomes of bacterial and viral pathogens provided valuable biological information regarding infectious disease processes and also led to the advent of molecular infectious disease testing. Clinical sequencing of bacterial 16S rRNA genes aids in identification of unknown organisms and sequencing of HIV genes from patient specimens can identify the presence of mutations causing resistance to anti-viral therapies [79, 80]. Information obtained from sequencing also makes it possible to design clinical molecular tests to monitor viral loads in patients infected with HIV, HCV, BKV, and others and to quickly detect organisms that are difficult to grow and identify in culture.

Although traditional DNA sequencing is an invaluable tool in molecular diagnostics, it does have its limitations. A single sequencing reaction can produce a read length of approximately 600 nucleotides and often requires another sequencing reaction of the opposite DNA strand (bidirectional sequencing) for confirmation. When the sequencing of large stretches of DNA is required, the costs and time involved can be limiting. The analytic sensitivity of traditional sequencing methods can also be problematic in some situations. Sanger sequencing can easily detect two alleles when they are present at nearly equal frequencies as in genomic DNA samples that is heterozygous at a given locus. In some situations, however, an allele or sequence may be present at levels below 10–20 % of the total targeted DNA sequence which may not be detectable by Sanger sequence [81]. For example, tumor specimens submitted for somatic mutation testing often contain a significant amount of non-tumor cells. Germline DNA from these cells can dilute the signal from a low-level somatic mutation to the point of where the signal from the mutation is indistinguishable from background noise.

2.4.2 Next-Generation Sequencing

Newer DNA sequencing technologies, normally referred to as next-generation sequencing, have been developed by several companies that can produce massively parallel sequencing data to overcome some of these limitations of more traditional DNA sequencing [82]. Instead of the individual reads of amplified DNA segments obtained by Sanger sequencing, next-generation sequencing can produce large

numbers of shorter reads across an entire gene, group of genes, or an entire genome with each segment of DNA being sequenced tens, hundreds, or thousands of times. These features provide an increased ability to detect low-level variations and also make sequencing across large stretches of DNA more feasible. The cost of sequencing an entire genome is currently too high to make the regular use of whole genome sequencing a reality in the clinical setting but more targeted approaches to these next-generation sequencing technologies are becoming more feasible. Assays capable of sequencing an entire gene or group of genes related to a specific genetic syndrome or cancer have quickly infiltrated clinical laboratories in recent years. In coming years this approach to large-scale sequencing may replace current multiplexed genotyping methods used to screen patients for large numbers of mutations such as those in the *CFTR* gene responsible for cystic fibrosis and the often tedious, large-scale approaches to whole gene sequencing to detect mutations that can predispose carriers to cancers [83]. The coverage or read-depth can also be used to detect low-level somatic mutations found in various tumor types. Although the expenses involved in next-generation sequencing are still quite high the technology is becoming more and more affordable everyday and the typical cost per nucleotide sequenced for laboratories using these technologies on a regular basis is much lower than traditional Sanger sequencing [84].

Although the chemistry and workflow can vary greatly between next-generation sequencing platforms and even between different applications on the same platform, there are a number basic steps that are common to most next-generation sequencing. After isolating DNA from a specimen, the sample must be processed in such a way to enrich for the desired target sequences and to modify the DNA to make it suitable for sequencing. For example, in clinical testing of cancer specimens, this target enrichment may be performed using PCR to specifically amplify exons of cancer genes known to be hotspots for clinically relevant mutations. After PCR amplification various adapter oligonucleotide sequences can be attached to the ends of the amplified DNA. These adapters often include molecular barcodes to uniquely identify sequences obtained from a specific sample. This barcoding allows for pooling of multiple samples in a single sequencing reaction. If PCR-based enrichment is not used, the addition of adapters can be used along with various capture methods to enrich for the desired sequences. Capture-based enrichment is typically preferred when sequencing larger numbers of target sequencing, as is needed in whole exome sequencing [85, 86]. Regardless of the method used, these prepared libraries DNA to be sequenced can be pulled together and processed on next-generation sequencing platform [87]. After the sequencing is complete, the raw sequence data must be submitted to a bioinformatics pipeline to align the sequence reads to a reference sequence, make variant calls and then filter out benign variants and rank the remaining variants based on clinical significance. This pipeline may vary greatly based on platform, application,

types of variants expected and the preference of a particular lab. Much of this pipeline may be automated but manual review and interpretation is always needed to make the final calls and interpretations. Various software programs and algorithms can be used for this alignment and variant calling and care must be taken to select a suitable component of an analysis pipeline. Compared to more focused sequencing, one of the biggest challenges to successfully implementing clinical next-generation sequencing is the interpretation of variants and dealing with a larger number of variants of unknown significance. This becomes more of an issue as one moves from gene panels to whole exome to whole genome sequencing. In some situations where sequencing a larger number of exons or genes may be technically and financially viable, restricting the targeted sequences of an assay may be desired to reduce the possibility of uncertain findings [88-91].

References

- Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-8.
- Schumm DE. Core concepts in clinical molecular biology. Philadelphia: Lippincott-Raven; 1997.
- Schumm DE. Essentials of biochemistry. 2nd ed. Boston: Little & Brown; 1995.
- Thoma F, Koller T. Influence of histone H1 on chromatin structure. *Cell*. 1977;12:101-7.
- Varshavsky AJ, Bakayev VV, Nedospasov SA, Georgiev GP. On the structure of eukaryotic, prokaryotic, and viral chromatin. *Cold Spring Harb Symp Quant Biol*. 1978;42:457-73.
- Lamond AI, Earnshaw WC. Structure and function in the nucleus. *Science*. 1998;280:547-53.
- Long EO, Dawid IB. Repeated genes in eukaryotes. *Annu Rev Biochem*. 1980;49:727-64.
- Jordan EG. The nucleolus. 2nd ed. Oxford: Oxford University Press; 1978.
- Carmo-Fonseca M, Mendes-Soares L, Campos I. To be or not to be in the nucleolus. *Nat Cell Biol*. 2000;2:E107-12.
- Tyler-Smith C, Willard HF. Mammalian chromosome structure. *Curr Opin Genet Dev*. 1993;3:390-7.
- Biessmann H, Mason JM. Telomeric repeat sequences. *Chromosoma*. 1994;103:154-61.
- Counter CM. The roles of telomeres and telomerase in cell life span. *Mutat Res*. 1996;366:45-63.
- Arzimanoglou II, Gilbert F, Barber HR. Microsatellite instability in human solid tumors. *Cancer*. 1998;82:1808-20.
- Sutherland GR, Richards RI. Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci U S A*. 1995;92:3636-41.
- Craig I. Human genetics. Methylation and the fragile X. *Nature*. 1991;349:742-3.
- The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*. 1993;72:971-83.
- Hoffmann-Berling H. DNA unwinding enzymes. *Prog Clin Biol Res*. 1982;102:89-98.
- Kato S, Kikuchi A. DNA topoisomerase: the key enzyme that regulates DNA super structure. *Nagoya J Med Sci*. 1998;61:11-26.
- Heywood LA, Burke JF. Mismatch repair in mammalian cells. *Bioessays*. 1990;12:473-7.
- Auerbach AD, Verlander PC. Disorders of DNA replication and repair. *Curr Opin Pediatr*. 1997;9:600-16.
- Nojima H. Cell cycle checkpoints, chromosome stability and the progression of cancer. *Hum Cell*. 1997;10:221-30.
- Fotedar R, Fotedar A. Cell cycle control of DNA replication. *Prog Cell Cycle Res*. 1995;1:73-89.
- Green R, Noller HF. Ribosomes and translation. *Annu Rev Biochem*. 1997;66:679-716.
- Persson BC. Modification of tRNA as a regulatory device. *Mol Microbiol*. 1993;8:1011-6.
- Sharp SJ, Schaack J, Cooley L, Burke DJ, Soll D. Structure and transcription of eukaryotic tRNA genes. *CRC Crit Rev Biochem*. 1985;19:107-44.
- Goldberg S, Schwartz H, Darnell Jr JE. Evidence from UV transcription mapping in HeLa cells that heterogeneous nuclear RNA is the messenger RNA precursor. *Proc Natl Acad Sci U S A*. 1977;74:4520-3.
- Sharp PA. RNA splicing and genes. *JAMA*. 1988;260:3035-41.
- Wang Y, Stricker HM, Gou D, Liu L. MicroRNA: past and present. *Front Biosci*. 2007;12:2316-29.
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-45.
- Chou KC, Kezdy FJ, Reusser F. Kinetics of processive nucleic acid polymerases and nucleases. *Anal Biochem*. 1994;221:217-30.
- Kollmar R, Farnham PJ. Site-specific initiation of transcription by RNA polymerase II. *Proc Soc Exp Biol Med*. 1993;203:127-39.
- Tantravahi J, Alvira M, Falck-Pedersen E. Characterization of the mouse beta maj globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. *Mol Cell Biol*. 1993;13:578-87.
- Varani G. A cap for all occasions. *Structure*. 1997;5:855-8.
- Munroe D, Jacobson A. Tales of poly(A): a review. *Gene*. 1990;91:151-8.
- Gorlach M, Burd CG, Dreyfuss G. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res*. 1994;211:400-7.
- Balvay L, Libri D, Fiszman MY. Pre-mRNA secondary structure and the regulation of splicing. *Bioessays*. 1993;15:165-9.
- Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*. 1998;92:315-26.
- Edwards-Gilbert G, Veraldi KL, Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res*. 1997;25:2547-61.
- Morrissey JP, Tollervy D. Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. *Trends Biochem Sci*. 1995;20:78-82.
- Cech TR. Self-splicing of group I introns. *Annu Rev Biochem*. 1990;59:543-68.
- Jacquier A. Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem Sci*. 1990;15:351-4.
- Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*. 1984;37:67-75.
- Sader HS, Hollis RJ, Pfaller MA. The use of molecular techniques in the epidemiology and control of infectious diseases. *Clin Lab Med*. 1995;15:407-31.
- Anand R. Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules. *Trends Genet*. 1986;2:278-83.
- Thomas R. The denaturation of DNA. *Gene*. 1993;135:77-9.
- Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
- Presnell SC. Nucleic acid blotting techniques. In: Coleman WB, Tsongalis GJ, editors. *Molecular diagnostics for the clinical laboratorian*. Totowa, NJ: Humana Press; 1997. p. 63-88.
- Boltz EM, Kefford RF, Leary JA, Houghton CR, Friedlander ML. Amplification of c-ras-Ki oncogene in human ovarian tumours. *Int J Cancer*. 1989;43:428-30.

49. Hynes NE. Amplification and overexpression of the erbB-2 gene in human tumors: its involvement in tumor development, significance as a prognostic factor, and potential as a target for cancer therapy. *Semin Cancer Biol.* 1993;4:19–26.
50. Sorscher DH. DNA amplification techniques. In: Coleman WB, Tsongalis GJ, editors. *Molecular diagnostics for the clinical laboratorian*. Totowa, NJ: Humana Press; 1997. p. 89–101.
51. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science.* 1985;230:1350–4.
52. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 1987; 155:335–50.
53. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. Hybridization of synthetic oligodeoxynucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res.* 1979;6:3543–57.
54. Lawyer FC, Stoffel S, Saiki RK, Chang SY, Landre PA, Abramson RD, Gelfand DH. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl.* 1993;2:275–87.
55. Lawyer FC, Stoffel S, Saiki RK, Myambo K, Drummond R, Gelfand DH. Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *J Biol Chem.* 1989;264:6427–37.
56. Fanning S, Gibbs RA. PCR in genome analysis. In: Birren B, Green ED, Klapholz S, Myers RM, Roskams J, editors. *Genome analysis, vol 1. Analyzing DNA*, vol. 1. Plainview, NY: Cold Spring Harbor Laboratory Press; 1997. p. 249–99.
57. Kaijalainen S, Karhunen PJ, Lalu K, Lindstrom K. An alternative hot start technique for PCR in small volumes using beads of wax-embedded reaction components dried in trehalose. *Nucleic Acids Res.* 1993;21:2959–60.
58. Bassam BJ, Caetano-Anolles G. Automated “hot start” PCR using mineral oil and paraffin wax. *Biotechniques.* 1993;14:30–4.
59. Roux KH. Using mismatched primer-template pairs in touchdown PCR. *Biotechniques.* 1994;16:812–4.
60. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. ‘Touchdown’ PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 1991;19:4008.
61. Foord OS, Rose EA. Long-distance PCR. *PCR Methods Appl.* 1994;3:S149–61.
62. Barnes WM. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A.* 1994;91:2216–20.
63. Ferre F. Quantitative or semi-quantitative PCR: reality versus myth. *PCR Methods Appl.* 1992;2:1–9.
64. Raeymaekers L. A commentary on the practical applications of competitive PCR. *Genome Res.* 1995;5:91–4.
65. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques.* 1997;22:130–1.
66. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res.* 1996;6:986–94.
67. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A.* 1991;88:7276–80.
68. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002;347:1999–2009.
69. Pollack JR. A perspective on DNA microarrays in pathology research and practice. *Am J Pathol.* 2007;171:375–85.
70. Pillai R, Deeter R, Rigl CT, et al. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. *J Mol Diagn.* 2011;13:48–56.
71. Erlander MG, Ma XJ, Kesty NC, et al. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. *J Mol Diagn.* 2011;13:493–503.
72. Wang L, Luhm R, Lei M. SNP and mutation analysis. *Adv Exp Med Biol.* 2007;593:105–16.
73. Maurice CB, Barua PK, Simses D, Smith P, Howe JG, Stack G. Comparison of assay systems for warfarin-related CYP2C9 and VKORC1 genotyping. *Clin Chim Acta.* 2010;411:947–54.
74. Lefferts JA, Jannetto P, Tsongalis GJ. Evaluation of the Nanosphere Verigene system and the Verigene F5/F2/MTHFR nucleic acid tests. *Exp Mol Pathol.* 2009;87:105–8.
75. Kearney HM, South ST, Wolff DJ, Lamb A, Hamosh A, Rao KW. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genet Med.* 2011;13:676–9.
76. Lee C, Lafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet.* 2007;39:S48–54.
77. de Ravel TJL, Devriendt K, Fryns J-P, Vermeesch JR. What's new in karyotyping? The move towards array comparative genomic hybridisation (CGH). *Eur J Pediatr.* 2007;166:637–43.
78. Shendure JA, Porreca GJ, Church GM, Gardner AF, Hendrickson CL, Kieleczawa J, Slatko BE. Overview of DNA sequencing strategies. *Curr Protoc Mol Biol.* 2011;96:7.1.1–7.1.23.
79. Woo PC, Lau SK, Teng JL, Tse H, Yuen KY. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect.* 2008;14:908–34.
80. Dunn DT, Coughlin K, Cane PA. Genotypic resistance testing in routine clinical care. *Curr Opin HIV & AIDS.* 2011;6:251–7.
81. Franklin WA, Haney J, Sugita M, Bemis L, Jimeno A, Messersmith WA. KRAS mutation: comparison of testing methods and tissue sampling techniques in colon cancer. *J Mol Diagn.* 2010;12:43–50.
82. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. *Anal Chem.* 2011;83:4327–41.
83. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn.* 2008;10:484–92.
84. Mertes F, Elsharawy A, Sauer S, et al. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics.* 2011;10:374–86.
85. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques.* 2014;56:61–77.
86. Hagemann IS, Cottrell CE, Lockwood CM. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Canc Genet.* 2013;206:420–31.
87. Nguyen L. and L. Burnett, Automation of molecular-based analyses: a primer on massively parallel sequencing. *Clin Biochem Rev.* 2014;35:169–76.
88. Chang F. and M.M. Li, Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genetics.* 2013;206:413–9.
89. Duzkale H., J. Shen, H. McLaughlin, A. et al. A systematic approach to assessing the clinical significance of genetic variants. *Clinical Genetics.* 2013;84:453–63.
90. Rehm H.L., S.J. Bale, P. Bayrak-Toydemir, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine.* 2013;15:733–47.
91. Roy S., M.B. Durso, A. Wald, Y.E. Nikiforov, M.N. Nikiforova, SeqReporter: automating next-generation sequencing result interpretation and reporting workflow in a clinical laboratory. *Journal of Molecular Diagnostics.* 2014;16:11–22.

<http://www.springer.com/978-1-934115-18-3>

The Molecular Basis of Human Cancer

Coleman, W.B.; Tsongalis, G.J. (Eds.)

2017, XIX, 873 p. 217 illus., 120 illus. in color.,

Hardcover

ISBN: 978-1-934115-18-3

A product of Humana Press