PREFACE

The practice of medical oncology is in a period of significant positive change that owes primarily to advances in the basic science of oncology. In recent years, developments in molecular biology techniques have substantially increased our ability to detect and characterize genetic defects in human cells, resulting in significant increases in our understanding of the normal molecular mechanisms controlling cellular proliferation and differentiation. The advancement of our comprehension of these basic molecular mechanisms has been paralleled by comparable increases in our understanding of the molecular basis of the processes involved in neoplastic transformation and tumorigenesis. Information gleaned from studies conducted in basic molecular research laboratories is being applied with unprecedented speed to the development of new molecular tests for cancer diagnosis and prediction of clinical outcome, as well as to the development of new strategies for cancer prevention and treatment. Basic scientists, clinical scientists, and physicians have a need for a source of information on the current state of the art of the molecular biology of human neoplastic diseases. In this volume on *The Molecular Basis of Human Cancer* we attempt to provide such a source of current information, as well as provide a look to the future of the discipline and the potential impact of scientific advances on the practice of medical oncology. This book is directed primarily to advanced graduate students and medical students, postdoctoral trainees, and established investigators having basic research interests in the molecular basis of human neoplastic disease. However, it is also well suited for the non-expert with similar interests because it provides a broad overview of general themes in the molecular biology of cancer. To be sure, our understanding of the many processes of neoplasia and their molecular basis is far from complete, but few areas of thematic or conceptual consensus have developed. We have made an effort to integrate accepted principles with broader theoretic concepts in an attempt to present a current and comprehensive view of the molecular basis of human cancer. We hope that *The Molecular Basis of Human Cancer* will accomplish its purpose of providing students and researchers who already possess strong but diverse basic science backgrounds with unifying concepts, so as to stimulate new research aimed at furthering our understanding of neoplastic disease.

William B. Coleman
Gregory J. Tsongalis

2 Essential Concepts and Techniques in Molecular Biology

SHARON COLLINS PRESNELL, PHD

THE BIOLOGY OF NUCLEIC ACIDS

COMPOSITION AND STRUCTURE OF DNA

Deoxyribonucleic acid (DNA) is formed by the linear polymerization of nucleotides, which are composed of a nitrogen-containing base and a phosphorylated sugar. 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 phosphate groups (Fig. 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 the 5´ end, and no nucleotide attached to the 3´ hydroxyl group at the 3´ end.

The DNA within the 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 in each eukaryotic cell must be compressed to fit within the nucleus, which is only about 10µm in diameter. The DNA is segmented into 46 discreet structural units, termed chromosomes. Chromosomal DNA is condensed by the formation of nucleosomes; a group of small basic proteins (histones) around which 160–180 base pairs of DNA are wrapped *(4)*. Formation of nucleosomes is not sequence-dependent and it
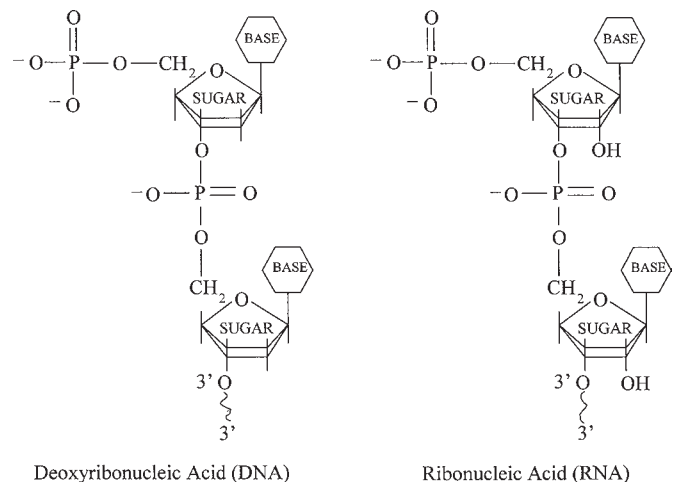
Fig. 1. DNA and RNA structure. Strands of nucleic acids are formed by the linkage of nitrogenous bases (purines and pyrimidines) via their phosphate groups. Note the presence of the extra hydroxyl group on the ribose component of RNA.

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, whereas less condensed DNA (euchromatin) is readily accessible by transcription machinery and is located in the center of the nucleus *(6)*.

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 for transfer RNA (tRNA), ribosomal RNA (rRNA), and the histone proteins *(7)*. These tandemly repeated genes are present on several

chromosomes and associate in the nucleus to form the nucleolus *(8)*. Highly repetitive sequences with thousands of copies, called satellite DNAs, are found at the telomeric ends of chromosomes and around the centromeres *(9)*. 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 *(10)*, and it has been demonstrated that the length of the telomeric repeat sequences decreases with life-span of cultured human cells *(11)*. 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 *(11)*. 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 *(12)*.

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 muscular dystrophy, Fragile X syndrome, Huntington's disease (HD), 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 HD) *(16)*.

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 telangectasia (A-T), Fanconi's anemia, and xeroderma pigmentosum (XP) *(20)*.

The DNA within the nucleus of a eukaryotic cell can be replicated completely in about eight h, during S phase of the cell cycle. Resting cells (G_0) receive a mitotic stimulus, which causes transition into 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)*.

STRUCTURE AND COMPOSITION OF RNA

RNA, or ribonucleic acid, is a linear polymer of ribonucleotides linked by 5´-3´ phosphodiester bonds (Fig. 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. Because of the extra hydroxyl group present on the ribose, RNA is more susceptible than DNA to nucleases. 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 proteins are synthesized. mRNA is the most heterogeneous type of RNA, and also has the shortest half-life. Other RNA species in eukaryotic cells include heterogeneous nuclear RNAs (hnRNA), which are the precursors to mature mRNA *(26)*, and small nuclear RNAs, which are involved in the synthesis/processing of mRNA *(27)*.

TRANSCRIPTION OF RNA

Even simple eukaryotic organisms, such as yeast, contain a large number of genes (~2,000), and higher eukaryotes, such as mammals, have ~100,000 genes *(28)*. 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 synthesis 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) *(29)*. 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 *(30)*. 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 *(29)*. Termination of transcription is poorly understood in eukaryotes, but takes place at sites that include a stretch of T's on the nontemplate strand of the gene *(31)*.

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 *(32)*. 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 *(33)*. The mRNA molecule is stabilized further by the association of a ~70 kDa protein with the poly-A tail *(34)*.

The majority of protein-encoding eukaryotic genes contain intervening sequences, termed introns, which do not encode any portion of the protein. These introns are maintained during transcription of the hnRNA molecule, which results 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 removal of the introns in conjunction with the joining of the coding sequences, termed exons. Introns may be between 65–10,000 base pairs, and short consensus sequences are found at either end and within the intron near the 3′-end *(35)*. 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 snRNAs *(35)*. 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 *(36)*. Alternative transcripts may also be generated by the use of a secondary polyadenylation site *(37)*.

RNA processing is not limited to mRNA. Eukaryotic tRNAs are modified post-transcriptionally, as are eukaryotic rRNAs *(38)*. 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 *(39)*. Group II introns are also self-splicing, are removed as a lariat structure, and require spermidine as a cofactor *(40)*.

CONTROL OF GENE EXPRESSION

Gene expression in eukaryotes may be controlled at the level of mRNA stability, translational frequency, splicing, or protein stability. However, the predominant mechanism for expression control is at the level of transcription. Some genes, called housekeeping genes, are needed by cells at all times regardless of the metabolic needs of the cell. These genes are transcribed at about the same rate at all times, and are therefore expressed constitutively. Other genes are controlled stringently at the level of transcription, and their expression may be induced or repressed depending on the metabolic state of the cell. Some experimental models have taken advantage of this phenomenon by constructing expression vectors in which a promoter that can be modulated metabolically drives expression of a specific gene. For example, vectors have been constructed in which specific genes are placed under control of the metallothionein promoter, which is zinc-inducible *(41)*. When cells containing the metallothionein-promoted expression vector are exposed to Zn^{++}, the level of transcription of the gene of interest is induced, so that the effects of expression of that gene on the host cells can be evaluated in a controlled environment.

BASIC MOLECULAR ANALYSIS AND INTERPRETATION

Investigations into the molecular mechanisms of disease depend on the analysis of cellular 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 and in the search for gene mutations, amplifications, and deletions, and RNA analyses are employed frequently in the characterization of gene expression.

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 (pH <4.0 or pH >10.0), hydrogen-bond disrupting agents (such as urea or formamide), or heat *(42)*. The melting point (T_m) of a specific double-stranded DNA sequence is reached when 50% of the hydrogen bonds are disrupted, and is related linearly to the percent G + C content of the DNA. The denaturation process is easily monitored by spectrophotometry, because the absor-

bance of the DNA at 260 nm increases as denaturation progresses *(42)*.

Hybridization of nucleic acid strands is a 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 *(43)*.

CONCEPTS AND APPLICATIONS OF SOUTHERN BLOTTING The technique of Southern blotting is widely used in a clinical or forensic setting 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.

The first step in successful Southern blotting is to obtain DNA that is reasonably intact. 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 that 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 DNAs are transferred to a solid support such as a nylon or nitrocellulose membrane via capillary action or electrophoretic transfer and are bound permanently to the membrane by brief ultraviolet (UV) crosslinking or by prolonged exposure to temperatures 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 *(43,44)*.

The analysis of Southern blots requires the grasp of a few simple concepts. Interpretation 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 were 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 *(45,46)*, or in genes such as the multidrug resistance gene (*MDR1*) during treatment with pharmacological agents *(47)*. 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 or phosphorimager. Structural aberrations in a gene of interest can be detected by Southern blotting, including the insertion or deletion of nucleotides or gene rearrangements. 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 obvious on Southern blots as an abnormal banding pattern on the autoradiogram (Fig. 2). The Southern blot remains a useful and reliable way to obtain definitive data on gene structure.

THE 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 *(48,49)*, 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)*.

Principles of PCR Any PCR reaction must start with a DNA template. The DNA may be genomic DNA isolated directly from experimental or patient material, or it may be cDNA that has been synthesized from a DNA or RNA template by polymerase or reverse transcriptase. The target for any PCR reaction is dictated by specific oligonucleotide primers that anneal to two sites at either end of the region of interest, on opposite template strands. 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), 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 *(48,49)*. 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. Accumulation of target molecules reaches a plateau eventually; the initial number of target sequences and the efficiency of primer extension determine the upper limits of amplification. Due to the incredible sensitivity of PCR, even a miniscule amount of contaminating DNA can result in misinterpretation.

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. A long primer may be more specific for the target sequence, but is less efficient at hybridization, whereas a short primer is efficient at hybridization but less specific for the target sequence. 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, because 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 calculate an appropriate annealing temperature for any given primer: $T_m = 69.3 +$
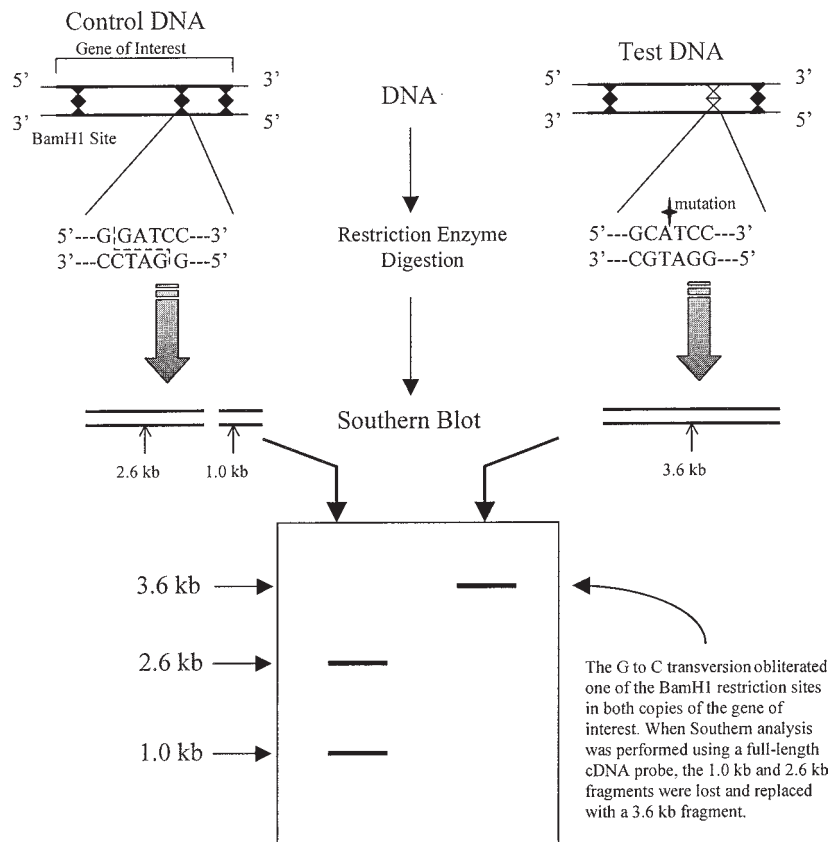
Fig. 2. Southern blot analysis. Double-stranded DNA of interest is fragmented with restriction enzymes and size-fractionated by agarose gel electrophoresis. The resulting ladder of fragments is transferred to a nylon or nitrocellulose membrane, which is then probed with a complementary RNA or DNA probe specific for the gene(s) of interest. In this example, the G‡C mutation present in the test DNA resulted in the loss of one of the *Bam*H1 restriction sites in the gene of interest, resulting in the loss of the 2.6 and 1.0 kb bands and appearance of a 3.6 kb band.

$0.41(\% G + C) - (650/L)$, where L = primer length in bases *(51)*. Repetitive or palindromic sequences should be avoided in a primer, and primer pairs should not contain sequences complementary to each other.

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 $5´\rightarrow3´$ exonuclease activity, thermostability, and optimum performance at 70–80∞C *(52,53)*. Temperature, pH, and concentration of Mg^{++} may influence the activity of *Taq* polymerase, and extremely high denaturation temperatures (>97∞C) will significantly reduce its active lifetime. Lower divalent cation (Mg^{++}) concentrations decrease the rate of dissociation of enzyme from template by stabilizing the enzyme-nucleic acid interaction *(54)*. 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

that make them useful for special PCR applications such as amplification of long pieces of DNA or high-fidelity amplification *(54)*.

Visualization of PCR Products Once the PCR is complete, the products must be visualized for analysis and interpretation. 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 (PAGE) is standard. Recent technological advances have resulted in the development of gel-free quantification methods that are useful in high-throughput applications *(55)*.

Modifications and Improvements of PCR Hot-start PCR was developed to reduce background from nonspecific amplification by preventing polymerization of new DNA during the initial phase of the reaction when nonspecific binding may occur between primers and other DNAs in the mixture *(56,57)*. 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, antibodies to the polymerase enzyme can be added to the reaction mixture; they prevent premature polymerization and are destroyed once the denaturation temperature is

reached. The key concept in hot-start PCR is to achieve a temperature greater than the annealing temperature before the reaction components are allowed to initiate polymerization.

Touch-down PCR was developed to enhance amplification of the desired target sequences while reducing amplification of artifacts *(58,59)*. 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.

Performing nested PCR can increase both the sensitivity and specificity of amplification *(49)*. 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, because 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 in order to avoid contamination.

It is possible to amplify sequences as large as 50 kb using long-distance PCR (LD-PCR) *(60)*. 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 *(61)*. 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 (T_m).

Quantitative PCR provides a quick and simple alternative to Southern- or Northern-blot analysis for the evaluation of gene copy number or gene expression levels *(62)*. 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 curve, and not during the plateau. Accurate quantitative PCR experiments 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, replica 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. For other aspects of quantitative PCR, see the following references *(62,63)*.

Cloning PCR Products Some PCR reactions are intended to provide information that is gleaned from gel electrophoresis and visualization alone. However, PCR products can be cloned and used for sequence analysis, construction of molecular probes, mutation analysis, in vitro mutagenesis, studies of gene expression, and many other applications. PCR fragments can be introduced into suitable vectors via several methods *(54)*, where they can be expanded and/or manipulated. A cloning strategy should be planned prior to the PCR reaction because modifications of the product, such as the insertion of restriction sites, are sometimes necessary to allow insertion of the product into the intended vector. Care should also be taken to verify the size and purity of the PCR product prior to cloning. With careful planning and the inclusion of appropriate controls, the conventional techniques of recombinant DNA technology can be replaced almost entirely with PCR-based methods.

DNA LIBRARIES In order to identify mutations or genetic alterations that are associated with disease, but are not detectable by Southern blot or PCR-based mutational analyses, the DNA of interest must be amplified so that a sufficient quantity is available for sequence analysis. One method for amplifying DNA is to clone the DNA of interest from a genomic DNA library or a cDNA library. Conceptually, genomic libraries represent the complete DNA from a specific source that has been divided into relatively small pieces ("books") that can be examined individually. cDNA libraries are constructed by synthesizing cDNA from the mRNAs that are expressed by a specific tissue or cell line, so that only the DNA that encodes expressed genes is represented. When screening a DNA library, the first task is to explore the library and identify the "book" that contains the DNA sequence of interest. Then, that particular piece of DNA must be isolated and amplified so that it may be examined in detail.

Construction of DNA Libraries DNA libraries may be constructed in bacteriophage vectors or cosmid vectors using genomic DNA that has been fragmented into large pieces by partial digestion with a restriction enzyme and ligated into the vector DNA and ultimately packaged into bacteriophage particles *(43)*. Cosmid vectors are constructed by modification of a plasmid by addition of *cos* DNA sequences that enable the vector to be packaged in bacteriophage λ particles *(64)*. The bacteriophage act simply as vehicles for delivery of the DNA into the host, where it replicates as a plasmid. Cosmid libraries are more difficult to construct using genomic DNA because the vector DNA often becomes concatenated and fails to incorporate the foreign eukaryotic DNA, and the foreign DNA sequences frequently undergo recombination, which results in rearrangement or loss of cloned segments of eukaryotic DNA. Although problems with cosmid vectors may be alleviated somewhat by additional steps such as the use of multiple restriction enzymes and treatment with phosphatases to prevent vector concatemerization, the additional steps required make cosmid vectors undesirable for routine cloning *(43)*. Nonethe-

less, cosmid vectors are useful for cloning and propagation of single, complete eukaryotic genes that contain a large number of introns and are therefore too large (33–45 kb) to be propagated in bacterophage λ vectors *(64)*. Cosmid vectors may also be used to analyze a region of eukaryotic DNA that contains a gene family *(65)*. Gene families are typically spread over 70–300 kb, and can be cloned by using a segment of nonrepetitive DNA isolated from one end of a recombinant as a probe to identify recombinants containing the adjacent sequence. This process, called chromosome walking, is repeated until all recombinants comprising the region of interest are identified. Because the maximum size of foreign DNA that can be carried in cosmid vectors (~45 kb) is twice the maximum size for bacteriophage λ vectors, chromosome walking is expedited greatly by cosmid cloning.

Bacteriophage λ vectors are employed commonly in the construction of genomic and cDNA libraries because they are easy to manipulate and screen *(66)*. Bacteriophage λ infect *Escherichia coli* by adsorbing to receptors that transport maltose into the cells. In the majority of hosts the 50 kb double-stranded λ genome is transcribed extrachromosomally, generating the structural units of mature bacteriophage particles, which are ultimately responsible for lysis of the host bacterium and release of progeny *(43)*. However, a small percentage of bacterial hosts incorporate the λ DNA into their genome and do not undergo lysis. When constructing a library, there are many bacteriophage vectors to choose from (Charon, λgt, EMBL, and λZap, to name a few). The choice is made based on which restriction enzymes will be used, the size of foreign DNA fragments, and whether or not the foreign DNA is to be expressed by the bacterial host.

It is possible for DNA libraries to provide a means for both propagation and expression of inserted foreign DNA sequences *(67)*. Some plasmid-based and bacteriophage-based vectors, such as pUR, pEX, λgt11, or λORF8, allow expression of foreign cDNA sequences by carrying a portion of the *E. coli* β-galactosidase gene including the elements necessary for its expression. Foreign DNA is inserted into the carboxy-terminal coding region of the β-galactosidase gene, resulting in a chimeric gene that encodes for a fusion protein between β-galactosidase and the protein encoded by the foreign DNA. This approach ultimately provides a means of producing antisera to the foreign DNA-encoded protein and allows the expression library to be screened by immunological rather than nucleic acid probes.

Screening of DNA Libraries Once a DNA library has been generated, the sequence or gene of interest must be identified among all of the recombinants. The methods of screening recombinants vary depending on the type of vector that is used, but in theory they are similar. When cosmids are used, host bacteria are plated on agar plates containing an antibiotic that corresponds with the antibiotic resistance gene included on the plasmid vector. Ideally, the bacterial colonies should be far enough apart that it is simple technically to distinguish one from another. Grunstein and Hogness have described a methodology that allows *in situ* lysis of bacterial colonies onto nitrocellulose or nylon membranes, which can then be probed for sequence(s) of interest using labeled nucleic acid probes *(68)*.

When positive signals are obtained, the signal is traced back to a specific colony on the bacterial plates, which is then expanded and used for amplification of the foreign DNA sequences of interest.

Screening of bacteriophage λ DNA libraries involves infection of host bacteria with recombinant phage and plating them onto the surface of an agar plate. The bacteria grow to form a lawn on the surface of the plate and those that are infected with bacteriophage will undergo lysis, releasing progeny that in turn infect and lyse surrounding bacteria. The end result is the formation of cleared plaques on the lawn of bacteria, with each plaque representing the genetic material from a single recombinant *(43)*. Again, it is important that the number of infective bacteriophage used is adjusted so that the plaques are sufficiently separate from each other to be distinguished individually. Phage DNA within the plaques can be transferred easily by absorption onto a nitrocellulose or nylon membrane, and can be probed using labeled nucleic acid probes. Positive signals are traced back to a specific plaque, which is picked from the plate and used to infect more host bacteria, thus amplifying the recombinant phage DNA containing the foreign DNA sequence(s) of interest.

Screening of expression libraries is accomplished by using antibodies to identify specific proteins expressed by recombinant clones in either plasmid or bacteriophage vectors *(69)*. When bacteriophage λ is used, the debris in the plaques (which contains the expressed protein) is transferred directly to nitrocellulose filters and screened for immunoreactive material using antibodies to the protein of interest. Because plasmid libraries are maintained as bacterial colonies, they must be lysed prior to fixation on nitrocellulose membranes. Therefore, plasmid vectors may be advantageous because the conditions for lysis can be altered so that the structure and conformation of the expressed protein is preserved.

CLONING STRATEGIES FOR LARGE DNAS: YACS, BACS, AND PACS

The Yeast Artificial Chromosome System While cosmid and bacteriophage λ vectors are powerful tools for isolation of cDNAs and some eukaryotic genes, limitations on insert size (<45 kb) have precluded cloning and analysis of larger eukaryotic genes or chromosomal regions. Several advances were imperative in the development of the yeast artificial chromosome (YAC) system: 1) *cis*-acting elements required for chromosomal stability in yeast were identified and characterized by Szostak and Blackburn *(70)*; 2) a system was developed that enabled yeast to be transformed at high efficiency *(71)*; and 3) pulsed-field gel electrophoresis was developed, which provided a method for resolving DNA fragments up to 10,000 kb. The YAC library is constructed by ligating large fragments of DNA to the two arms of a YAC vector, which is then introduced into yeast via transformation *(72,73)*. The YAC vector carries antibiotic resistance as well as DNA sequences that function as telomeres, a centromere, and an origin of replication. Transformants that take up and stably maintain an artificial chromosome can be identified as colonies on agar plates that possess characteristics (i.e., antibiotic resistance, etc.) encoded on the YAC vector. Some problems associated with the use of YAC vectors include insert instability, the presence of chi-

meric clones within the YAC libraries, and difficulties in DNA manipulation.

The Bacterial Artificial Chromosome System The bacterial artificial chromosome (BAC) cloning system is based on the *E. coli* fertility plasmid (F-factor), into which foreign DNA sequences up to 300 kb are inserted *(74)*. The F-plasmids are introduced into host *E. coli* via electroporation where they are maintained at a low copy number, which limits detrimental recombination events. Although the BAC system offers some advantages over YAC vectors, there are significant limitations, such as the lack of positive selection for recombinants and low yields of recombinant DNA due to limited copy number of the F plasmid *(74)*.

The P1 Artificial Chromosome System Successful cloning of very large (up to 300 kb) pieces of DNA can be accomplished by employing a bacteriophage P1-based cloning system, or PAC *(75)*. The vector was created by combining a modified pAd10SacBII vector and a modified pUC19 plasmid, resulting in a vector that 1) allows insertion of large DNA fragments, 2) provides control of insert copy number, 3) allows discrimination of recombinant vs nonrecombinant vectors, and 4) provides stable recombinants with little to no occurrence of chimerism *(75)*. High molecular weight foreign DNA is modified by the ligation of specific restriction endonuclease sites to both the 5´- and 3´-ends, enabling the foreign DNA to be inserted into cloning sites within the PAC. The primary limiting step in PAC cloning is integrity of the foreign DNA inserted into the vector.

DNA SEQUENCE ANALYSIS DNA that has been isolated by screening of a genomic DNA or cDNA library or by PCR can be subjected to sequence analysis by either of two methods. Maxam-Gilbert analysis involves chemical cleavage at a specific base, followed by electrophoresis of the DNA, which produces a series of bands of various sizes, each ending in the targeted base *(76)*. The fragmentation procedure is done four times (once for each base) and the separate reactions are subjected to PAGE simultaneously, generating a ladder of bands from which the DNA sequence is read top to bottom.

In the Sanger sequencing method, a labeled primer is annealed to the DNA to be sequenced and the DNA is replicated in a reaction that contains a limiting concentration of one of the four nucleotides *(77)*. The reaction contains a dideoxynucleotide corresponding to the specific nucleotide that is in low concentration, resulting in termination of replication upon incorporation of the dideoxynucleotide, which has no 3´-hydroxyl group upon which another nucleotide may be added. The result is a set of DNA fragments, each of which ends in a specific base. The reaction is done four times, one reaction with a limiting concentration of each nucleotide, followed by denaturation and PAGE analysis. A ladder of bands is apparent as in the Maxam-Gilbert method, and the sequence is read bottom to top. Note that the sequence obtained by the Sanger method is the cDNA sequence, whereas the Maxam-Gilbert method yields the direct DNA sequence. When adequate normal controls are included, sequence analysis can reveal point mutations as well as other anomalies such as insertions and deletions. It is also possible to discern whether an individual is heterozygous or homozygous for a specific mutation by evaluating whether both a normal and

mutated allele are present in the sequence, which would indicate heterozygosity.

Recent advancements have significantly improved sequence analysis by eliminating the tedious task of reading sequencing autoradiographs. Fluorescent dyes have replaced radiolabeled isotopes used in the Sanger sequencing method *(78)*, and instruments have been devised that can image fluorescent dye-labeled DNA fragments during electrophoresis *(79,80)*. Automated DNA sequencing instruments shuttle sequence data into a computer, where it is assimilated and output. The development of improved DNA-labeling techniques and the introduction of more efficient DNA polymerases continue to streamline DNA sequencing procedures *(78,81)*, improving the quality and efficacy of high-throughput sequence analysis.

ELECTROPHORETIC SEPARATION OF NUCLEIC ACIDS

Standard Electrophoresis of Nucleic Acids Gel electrophoresis of DNA is widely used in procedures such as Southern blotting and PCR, where separation 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). Gels employed in electrophoretic techniques are generally composed of agarose or polyacrylamide. Agarose gels are useful for routine Southern blotting and analysis of PCR products >100 base pairs. However, when very small DNA fragments are to be analyzed, special agarose formulations (MetaPhor®, NuSieve®) are available with enhanced sieving properties that aide in the resolution of small fragments and/or fragments that differ in size by as little as 1%. Some mutational analyses require special gels, such as MDE™ (mutation detection enhancement) gels, that can resolve DNA fragments based on differences in conformation. MDE gels can be nondenaturing, which maintains single-stranded nucleic acids in their native conformation, or denaturing agents, such as formamide, hydroxide ions, or urea, can be added to the gel or sample buffer, resulting in elimination of secondary structure in the nucleic acid strand. The following sections discuss several widely used electrophoretic techniques that are capable of extracting valuable information from genomic DNA samples or from cDNA samples that were generated by PCR.

Separation of RNA for Northern-blot analysis 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 nondenaturing PAGE.

Pulsed-Field Gel Electrophoresis 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 *(82)*. PFGE is often used

in the identification of pathogens *(83)*, i.e., differentiating between strains of bacteria. 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 that can be digested by restriction enzymes *in situ* and easily loaded into a gel apparatus *(82,84)*. In a typical analysis, PFGE will produce a pattern of DNA fragments that range in size from 10–800 kb.

Field-inversion gel electrophoresis (FIGE) is a type of PFGE that allows clear, reliable separation of large DNA fragments from 100 kb to several megabases *(85)*. FIGE devices are constructed so that periodic reversals of the electric field cause the DNA to alternate between forward and reverse migration through the gel *(86)*. Net forward movement of the DNA through the gel is achieved by making the duration of the forward pulse longer than the duration of the reverse pulse, or by using a greater voltage for the forward current than for the reverse current. Resolution of fragments less than 200 kb is achieved best by using greater voltage for the forward current rather than longer forward pulse duration *(86)*.

MUTATION ANALYSIS DNA sequencing is the most reliable way to detect a mutation, but it is an expensive and complex screening method. Mutations that involve single base changes are detectable by Southern-blot analysis or restriction fragment length polymorphism (RFLP) analysis when the change destroys or creates a restriction site *(87)*, and mutations may be identified directly by PCR when primers are designed to detect a specific mutation *(88)*. However, the success of these methods relies on some knowledge of the mutation, i.e., the restriction site that is altered or the location of the mutation in the gene sequence. Broader mutational analyses may be conducted using PCR and electrophoretic separation; the most commonly used techniques include heteroduplex analysis (HA), single-strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), and thermal gradient gel electrophoresis (TGGE).

SSCP Analysis SSCP analysis can detect mutations when the nucleotide sequence influences the secondary structure, or conformation, of a single strand of DNA *(89)*. Conformation polymorphisms may be detected as a change in the relative mobility of the DNA in a gel. Initially, the DNA of interest is purified and amplified via PCR using end-labeled primers. After denaturation, which generates single DNA strands, the DNA is subjected to nondenaturing PAGE. Mutations in sequences that govern the secondary structure of the single-strand result in conformation polymorphisms. Polymorphisms are evident on the gel as bands that migrate slower or faster than the control normal DNA, which must be included in the assay for interpretation (Fig. 3).

Heteroduplex Analysis HA is accomplished by mixing denatured DNA fragments from a wild-type (control) sample and a test DNA sample. The fragments reanneal, producing
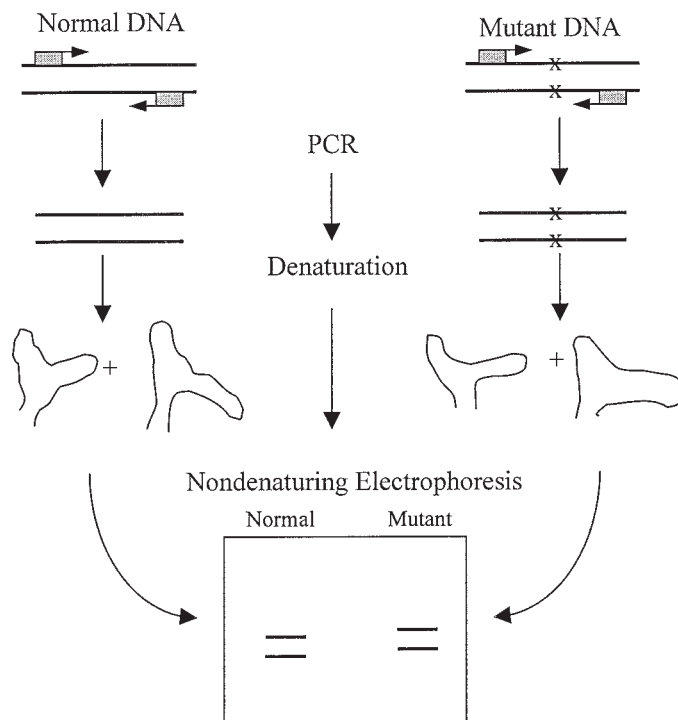

Fig. 3. SSCP analysis. Control and test DNA samples are subjected to PCR to amplify the region of interest. Single strands are generated by denaturation, and electrophoretic mobility of the strands are analyzed by nondenaturing PAGE. In this example, mutant DNA strands have altered conformations that are evident on the gel as a slight decrease in electrophoretic mobility when compared to normal control DNA.

either 1) wild-type:wild-type homoduplexes, 2) test:test homoduplexes, or 3) wild-type:test heteroduplexes. If a mutation is present in the test DNA, the heteroduplex will contain a bulge (in the case of an insertion or deletion) or a bubble (in the case of a point mutation) *(90)*. Bulge heteroduplexes are easily observed by agarose gel electrophoresis or PAGE, where they migrate slower than homoduplexes and result in the formation of a doublet on the gel (Fig. 4). However, bubble heteroduplexes are more difficult to visualize, and usually require the use of specialty gel matrices *(91,92)*.

DGGE and TGGE Analyses Some mutations in DNA can be detected by DGGE *(93)*. Denaturation of double-stranded DNA into single-stranded DNA occurs in steps as temperature rises or concentration of denaturant increases. Regions of DNA with the highest A + T content denature first, and denaturation continues until the region with the highest G + C content denatures, leaving completely single-stranded DNA. Mutations that change the denaturation profile of a region of DNA may be detected by DGGE. As test DNA migrates through a polyacrylamide gel containing a transverse gradient of denaturant, domains that denature yield partially single-stranded DNA fragments that have a decreased rate of migration *(94)*. The sensitivity of DGGE analysis may be increased by generating heteroduplexes between the test DNA sample and control DNA fragments of known sequence (Fig. 5). A mismatch between the test DNA and control DNA causes the heteroduplex to melt early compared to control DNA (Fig. 5). Once the area contain-
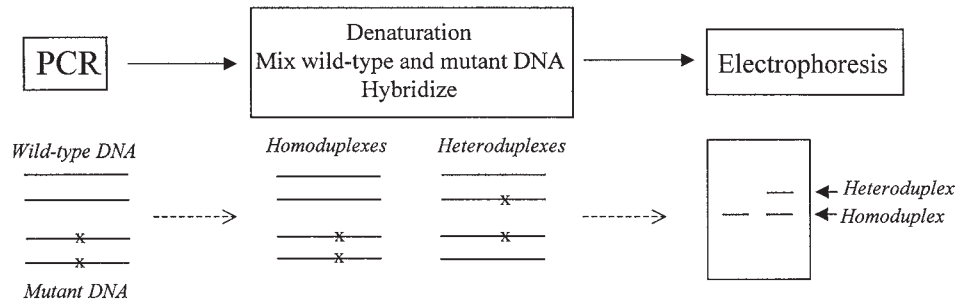
Fig. 4. Heteroduplex analysis. When wild-type DNA and test DNA are denatured, mixed, and hybridized, three types of double-stranded molecules are formed: 1) homoduplexes consisting of two strands of wild-type DNA, 2) homoduplexes consisting of two strands of test DNA, and 3) heteroduplexes consisting of one wild-type and one test strand. When mutations, insertions, or deletions are present in the test DNA, heteroduplexes that form do not align perfectly with the wild-type DNA strand, resulting in the formation of double-stranded DNA molecules with bulges or bubbles in the region of unpaired bases. Partially denatured heteroduplexes migrate more slowly than fully hybridized homoduplexes, resulting in the appearance of a doublet upon PAGE analysis. When the sequence of wild-type and test DNA is identical, heteroduplexes migrate at the same rate as homoduplexes and a single band is produced.

ing the mutation is defined and the melting profile generated by the mutation is characterized, DGGE analysis may be simplified by electrophoresing samples through MDE or polyacrylamide gels with a denaturant gradient in the same direction as electrophoresis; heteroduplexes with a mismatch exhibit slower mobility, producing a doublet in the gel similar to that seen in heteroduplex analysis (Fig. 4). TGGE is conducted essentially as DGGE, except the denaturing gradient is established by increasing temperatures rather than increasing concentrations of a chemical denaturant *(95)*.

BASIC RNA ANALYSIS AND INTERPRETATION

Isolating High-Quality RNA The success of any RNA-based molecular analysis rests largely on the quality of the RNA to be tested. Careful handling is essential to ensure that the RNA molecules remain intact during the isolation process. Total cellular RNA may be isolated reliably by centrifugation through cesium chloride gradients, or by acidic phenol extraction of cellular lysates *(96)*. For detailed protocols, *see* Sambrook et al. *(43)*. Contamination with RNAses (ubiquitous enzymes that readily degrade RNA) is frequently the cause of poor sample integrity. The quality of total RNA can be monitored by agarose gel electrophoresis of RNA samples, followed by ethidium bromide staining and UV visualization. The upper (or 28S) and lower (or 18S) ribosomal bands should be present at 5.0 kb and 1.87 kb, respectively, and there should be little to no degraded RNA evident. Samples lacking a clear upper or lower ribosomal RNA band or containing significant amounts of low molecular-weight (degraded) RNA should be considered suspect.

The poly-A tail present on mRNA strands provides an effective means of purifying mRNA from total cellular RNA, the bulk of which is rRNA. Isolation of mRNA is advantageous because it increases the proportion of mRNAs in the sample while lessening the concentration of other RNAs that may lead to background and interference in RNA analysis. Messenger RNA isolation techniques employ short stretches of thymidine or uracil residues (oligodT or oligodU) that are linked covalently to magnetic beads or resin. The total cellular RNA is exposed to the oligodT, and the poly-A tail of the mRNA hybridizes to the complementary thymidine or uracil residues. Unhybridized tRNA and rRNA are washed away, and the poly-A mRNA is eluted from the oligodT in a low-salt buffer. In typical eukaryotic cells, only 1–5% of the total cellular RNA is mRNA.

Northern Blotting Northern blotting involves the analysis of RNA without first converting the RNA to cDNA. Through Northern analysis, one can evaluate whether a gene is expressed, the relative level of expression, the size of the mRNA, and the presence of alternatively spliced transcripts. Total cellular RNA or mRNA is denatured, electrophoretically separated, and transferred to a solid support, such as nitrocellulose or positively charged nylon. The membrane-bound RNA is then probed with a labeled, complementary DNA or RNA probe. Visualization of positive signals is accomplished by generating autoradiograms, which can be analyzed quantitatively using a scanning densitometer, or, if ^{32}P-labeled probes were used, a phosphoimager. Simultaneous or subsequent probing of the RNA blot with a probe for a constitutively expressed housekeeping gene, such as β-actin or *cyclophilin*, will allow true quantitative assessment of expression of the gene of interest. For a more detailed description of Northern blot analysis and related protocols, see the following references *(43,44)*. Northern blots may be advantageous over other methods of gene-expression analysis, such as reverse transcription polymerase chain reaction (RT-PCR), when true quantitative results are desired, because there are fewer experimental artifacts associated with Northern blot analysis.

RNA ANALYSIS INVOLVING CDNA SYNTHESIS RNA

is inherently unstable in comparison to DNA, therefore techniques have been developed that provide a means of analyzing RNA by first converting it to complementary DNA (cDNA). RNA-derived cDNA is used routinely in many analyses, including construction of gene-expression libraries, PCR amplification, sequence analysis, and vector construction. By taking a population of mRNAs from a cell line and converting them to cDNAs, a gene-expression fingerprint is created, which can
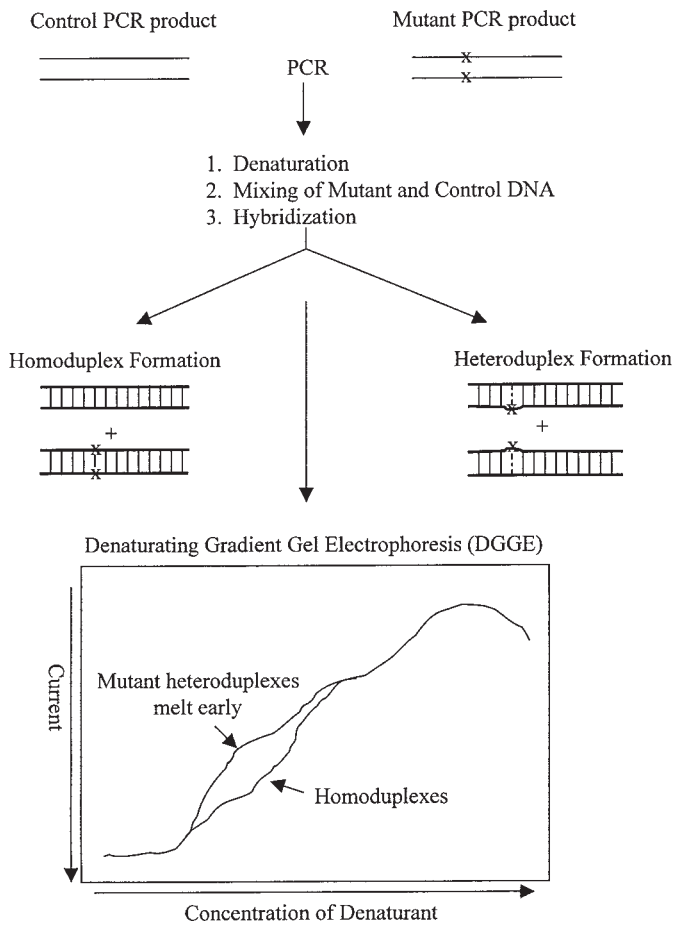
Control PCR product   Mutant PCR product

PCR

1. Denaturation
2. Mixing of Mutant and Control DNA
3. Hybridization

Homoduplex Formation   Heteroduplex Formation

Denaturing Gradient Gel Electrophoresis (DGGE)

Mutant heteroduplexes
melt early

Homoduplexes

Current

Concentration of Denaturant

Fig. 5. DGGE analysis. Mutations may be identified in a test DNA sample by comparing the melting profile of test DNA to control wild-type DNA. In this example, heteroduplexes are generated between test DNA and wild-type DNA. The hybridized mixture is then loaded into a single trough-like well that extends the width of the gel, which contains a transverse concentration of a chemical denaturant. As the sample is electrophoresed, the double-stranded DNA molecules denature, or melt, in domains. The higher the concentration of denaturant, the more rapidly the DNA strands melt. Double-stranded molecules that are hybridized fully migrate rapidly, but electrophoretic mobility of partially denatured strands is slowed. When mutations are present in the test DNA, the melting profile of the heteroduplexes is altered so that the DNA melts at a lower concentration of denaturant when compared to the control homoduplexes, resulting in the formation of a bubble on the gel. Once the DNA strands are denatured fully, they migrate very rapidly through the gel at a rate that is independent of their sequence.

then be compared to other fingerprints. The following sections will discuss the process of cDNA generation and several technical applications.

Generating cDNA from RNA The process of RNA-directed cDNA synthesis relies on an enzyme called reverse transcriptase, which is a retroviral enzyme capable of synthesizing DNA using RNA as a template *(97)*. RNA that is isolated from cells or tissue is incubated in the presence of reverse transcriptase and all four deoxynucleotides, along with an oligonucleotide primer upon which the cDNA will be built. When mRNA is the intended template, the primer can be a short sequence of deoxythymidine residues that targets the poly-A tail.
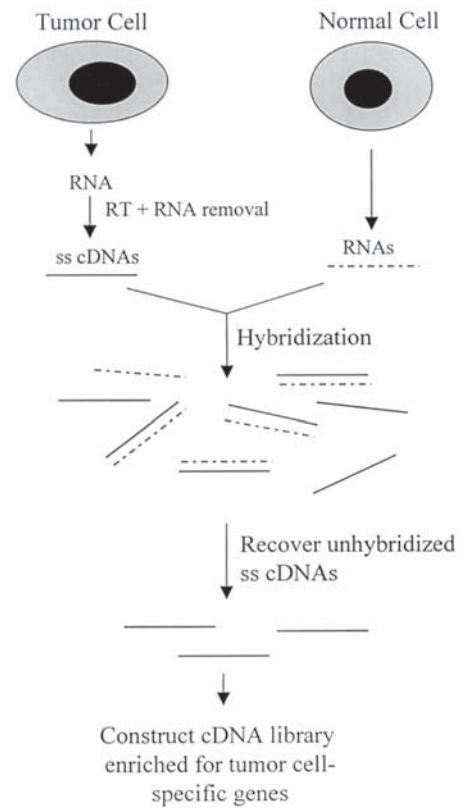
Tumor Cell   Normal Cell

RNA
        RT + RNA removal
ss cDNAs   RNAs

Hybridization

Recover unhybridized
ss cDNAs

Construct cDNA library
enriched for tumor cell-
specific genes

Fig. 6. Constructing a subtraction library. Subtraction libraries are often constructed to enrich for a population of cDNAs that may be involved in a specific biological process, such as malignant transformation. In this example, mRNA was isolated from a normal cell line and from a derived tumor cell line. The tumor cell mRNA was subjected to reverse transcription and the remaining RNAs were eliminated by enzymatic treatment with DNAse-free RNAse, leaving single-stranded cDNA molecules. A hybridization step is performed between the single-stranded cDNAs and mRNA isolated from the normal cells. Unhybridized mRNAs represent expressed genes that are probably unique to the normal cells, whereas unhybridized cDNAs represent genes that are expressed only by the tumor cells. Unhybridized single-stranded cDNAs are recovered and used to construct a cDNA library that is enriched for genes that may be involved in the malignant transformation of this cell line.

Alternatively, primers specific for internal sequences may be used, or, when sequence information is unknown, random oligonucleotide primers are used. The primer provides a 3´-OH group to which the reverse transcriptase enzyme adds deoxynucleotides to the growing cDNA strand. As the reverse transcriptase transcribes RNA into DNA, it degrades the template RNA, leaving a single-stranded DNA that is complementary and in the opposite orientation (5´ to 3´).

Using cDNA to Construct Libraries A population of RNAs isolated from a specific cell type or tissue may be reverse transcribed, yielding a population of cDNAs that is representative of the RNAs expressed. The cDNAs can then be used to construct DNA libraries. These libraries can be screened repeatedly to identify and isolate cDNAs corresponding to genes that are expressed by the cell type or tissue. To enrich for RNAs expressed specifically during a particular biological process, a subtraction library can be prepared *(43,98)*. For example, sup-

pose the goal is to detect genes that are expressed in a particular cell during malignant transformation. In this scenario, mRNA would be isolated from a control normal cell or tissue, and then mixed with cDNA prepared from the transformed cells or tumor tissue. Hybridization occurs between cDNAs and complementary RNAs, leaving unhybridized cDNAs that are expressed uniquely by the transformed cells or tumor tissue (Fig. 6). The unhybridized cDNAs are purified and used to produce a cDNA library that is enriched for genes that may be involved in the transformation process *(43)*. As an alternative approach, subtraction libraries may be constructed from the unhybridized mRNA that remains after hybridization between the cDNA and mRNA *(98)*.

RT-PCR PCR methodology can be exploited in the analysis of gene expression *(99)*. Typically, mRNA is reverse transcribed using an oligo-dT primer, generating a cDNA strand complementary to the RNA sequence beginning at the 3´-end of the mRNA *(100)*. Subsequently, a second strand is synthesized using the cDNA as a template, resulting in a double-stranded DNA molecule. If the mRNA is long, reverse transcription with an oligo-dT primer may not generate a cDNA that extends through the 5´-end of the mRNA. When specific sequence information is known, internal sequence-specific primers may be used for the reverse transcription reaction. Alternatively, when no 5´-sequence information is available, random oligonucleotides may be used to prime the cDNA synthesis *(99)*. Once the desired cDNA is obtained, specific oligonucleotide primers are used to amplify target sequences using the reverse transcribed cDNA as a template. The amplified product(s) can be visualized by agarose gel electrophoresis or PAGE, and information on gene expression can be extracted.

Successful reverse transcription of sample mRNA with oligo-dT primers requires that the poly-A tail be relatively intact. Therefore, it is imperative that high quality mRNA is used as starting material for RT reaction or sufficient cDNA template will not be generated for the subsequent PCR reaction. Inclusion of appropriate controls when setting up a RT-PCR experiment will allow a semi-quantitative assessment of gene expression. Fixed concentrations of control template can be amplified along with the test template to provide standards against which the amount of test amplification product is gauged. Alternatively, amplification of housekeeping genes, such as actin or cyclophilin, may provide a mechanism for quantification similar to the approach used in Northern blot analysis. See Larrick and Siebert for more information on quantitative approaches to RT-PCR *(99)*.

MOLECULAR PROBES FOR NUCLEIC ACID ANALYSIS

There are several methods for generating molecular probes to be used in analysis of DNA or RNA. The central concept behind the use of nucleic-acid probes is hybridization. Probes may be made of RNA or DNA, and can be labeled with colorimetric, fluorescent, enzymatic, or radioactive molecules for visualization purposes. Labeled nucleic acid probes are frequently used in Southern blotting, Northern blotting, and screening of DNA libraries.

Generation of DNA Probes DNA probes are useful for most routine procedures, because they are easy to generate and are very stable. Nick translation is one method for generating labeled DNA probes *(101)*. The first step in nick translation is to generate nicks in the phosphodiester backbone of a double-stranded DNA template by treatment with pancreatic DNAase I, which produces free 3´-hydroxyl termini along the strand. DNA polymerase I extends the 3´-OH termini in the 3´-direction, using its 5´→3´ exonuclease activity to hydrolyze the nontemplate strand. There are many disadvantages associated with nick translation, including the requirement for a large amount (0.5 µg) of DNA template, and the need for strict time and temperature limits in the protocol. Consequently, most DNA probes are generated using alternative methodologies.

An easy and effective method for generating DNA probes is random primer extension. Oligonucleotide primers of random sequence are annealed to denatured DNA template strands, followed by primer extension by the Klenow fragment of DNA polymerase or by T7 DNA polymerase in the presence of labeled nucleotides *(101)*. Random primer extension is advantageous because of the small amount of template required (~25 ng), the high specific activity of probes generated, the ability to generate probes from very large or very small templates, and the ease in preparation.

Probes generated from double-stranded DNA templates are effective when target sequences are present in sufficient quantity and the probe hybridizes strongly to the target sequences. When hybridization between the probe and target is weak, unwanted hybridization may occur between the complementary sequences in the probe. Single-stranded probes provide a means to detect specific target sequences without the danger of reannealing probe, and are particularly useful when target sequences are rare or are only weakly homologous to the probe *(101)*. Recombinant bacteriophage M13 are composed of single-stranded DNA molecules that can serve as templates for synthesis of single-stranded DNA probes. Oligonucleotide primers are designed and annealed to viral DNA sequences upstream from the inserted template sequences, followed by primer extension with the Klenow fragment of DNA Polymerase I in the presence of labeled dNTPs. Labeled probe is often separated from unlabeled DNA by gel electrophoresis or alkaline chromatography, but the resulting probe has an extremely high specific activity.

Generation of RNA Probes The RNA:RNA hybrid is much more stable than RNA:DNA hybrids or DNA:DNA hybrids. Therefore, single-stranded RNA probes (termed riboprobes) are advantageous when probing Northern blots or when detecting signals that are weak by other probing methods. The template DNA is inserted into a plasmid vector that contains one or more strong bacteriophage promoters (SP6, T7, or T3) that are recognized by bacteriophage-specific, DNA-dependent RNA polymerases *(102,103)*. In the presence of labeled NTPs, these polymerases synthesize a single-stranded riboprobe that is complementary to the target sequence and has a high specific activity. DNA is eliminated from the reaction by treatment with RNAse-free DNAase I. Riboprobes are a more sensitive and efficient alternative to double-stranded or single-stranded DNA probes.

Molecules for the Labeling of Probes The majority of conventional probe-labeling practices involve incorporation of nucleotides that are labeled with a radioactive molecule, such

as ^{32}P. Radioactive nucleotides may be incorporated into the probe during synthesis; three unlabeled (or cold) deoxynucleotides are added to the labeling mixture (dATP, dGTP, and dTTP), along with radiolabeled (or hot) ^{32}P-CTP. The radioactive molecule is incorporated into the nucleic acid probe, and positive signals are detected by autoradiography. As an alternative, radiolabeled dNTPs may be attached to the 3´-termini of a DNA probe using bacteriophage T4 DNA polymerase or to the 5´-termini using bacteriophage T4 polynucleotide kinase (43). In general, a higher specific activity is achieved when the label is incorporated rather than attached to the ends of the probe.

Many nonradioactive labeling alternatives have been developed, reflecting concerns about the environment, safety, and cost. The most widely used methodologies involve incorporation or end-labeling of probes with molecules such as digoxigenin, biotin, or fluorescein, that can be detected using immunological methods or visualized by chemiluminescence (104,105). This methodology generates an autoradiograph identical to those produced by traditional radioactive probes (106).

PRACTICAL MOLECULAR BIOLOGY IN CANCER RESEARCH AND CLINICAL LABORATORY SCIENCE

In some cases, the tools of molecular biology are used to provide simple answers that aid in the diagnostic and/or prognostic evaluation of patient material or experimental specimens. Examples of such analysis include the analysis of genes that are associated with an increased risk of development of breast cancer such as *BRCA1* and *BRCA2 (107,108)*, or analysis of genes, that are associated with the development of a subset of colon cancers, such as *MSH2* and *MLH1 (109)*. For these types of analysis, PCR, Southern-blot analysis, or Northern-blot analysis may suffice. However, in many cases, especially in experimental models of cancer where there are many unknowns, the quest for information is not straightforward. For example, if one wishes to know the molecular differences between normal and cancerous tissue, or between a normal cell line and one that has been transformed by a carcinogenic treatment. Answers to such questions are dependent on the development of efficient techniques for screening large numbers of genes from two or more sources simultaneously (Fig. 7). While Southern blots, Northern blots, PCR, and RT-PCR may be used for such large-scale comparisons, it is time-consuming, and it requires that some information about potential target genes be known. The development of micro array technology has provided a new way to evaluate expression of a large number of genes in a small amount of time. In addition, techniques such as differential display RT-PCR (RT-PCR/DD) and comparative genomic hybridization (CGH) have created methods for comparing the genetic composition and gene expression patterns of two or more populations simultaneously.

MICROARRAY TECHNOLOGY Microarrays provide a means for screening samples for presence or expression of very large numbers of genes simultaneously (110). Several approaches have been developed for construction of arrays. In one approach, hundreds of cDNA targets are immobilized
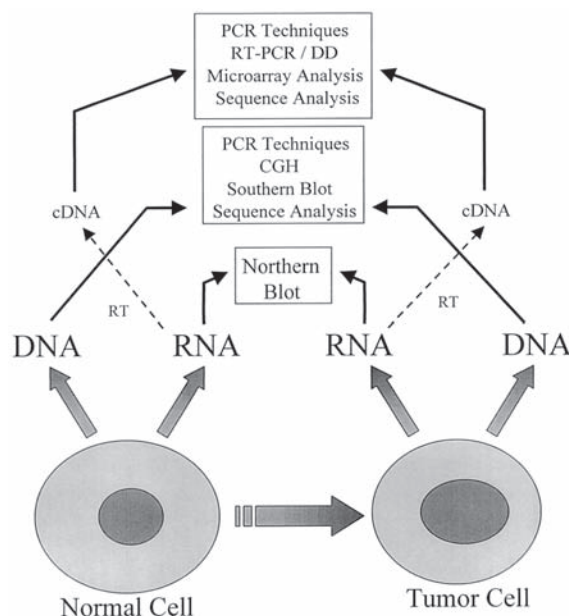

Fig. 7. Comparative analyses in molecular biology. In vivo and in vitro models of cancer, and the analysis of patient samples, often involve a comparative analysis between normal cells and tumor cells. Cellular RNA and DNA may be compared through a large variety of methods, including several (RT-PCR/DD, microarray analysis, CGH) that allow a broad comparative survey of genetic differences.

robotically on glass slides and hybridized to fluorescent ssDNA probes produced by reverse transcription from total mRNA of interest (111). Positive hybridization signals are identified and compared to an internal standard with a laser scanner and displayed as a colored map, where the colors are indicative of relative expression level. A larger-scale analysis may be accomplished by immobilization of up to 30,000 cDNA or genomic DNA targets on large nylon membranes using automation (112). Test DNA or cDNA is used to generate short ^{32}P-labeled oligonucleotide probes that are hybridized to the nylon membranes and imaged with a phosphoimager. In a similar approach, GeneChips have been constructed that contain nearly 100,000 probes per chip (10^6 probes per cm^2). Test DNA samples are labeled fluorescently, hybridized to the GeneChip, and analyzed by confocal epifluorescence microscopy and computer digitization (113).

DNA microarrays have been employed successfully in studies of gene expression in yeast, plants, and mammalian cells. De Risi et al. investigated the molecular basis of tumorigenicity in a human melanoma cell line (UACC-903) by comparing tumorigenic and nontumorigenic derivatives of the cells by microarray technology. This approach resulted in the identification of several genes that are associated with suppression of tumorigenicity in UACC-903 cells (114). In addition, a GeneChip containing 96,000 oligonucleotide probes was used successfully to detect heterozygous mutations in exon 11 of the *BRCA1* gene in patient samples (113). Microarray technology is advancing rapidly, and has proven to be effective in diagnostics, gene-expression studies, gene mapping, and gene discovery.

DIFFERENTIAL DISPLAY RT-PCR RT-PCR/DD was introduced in 1992 as a technique for the comparison, identification, and isolation of genes that are expressed uniquely among
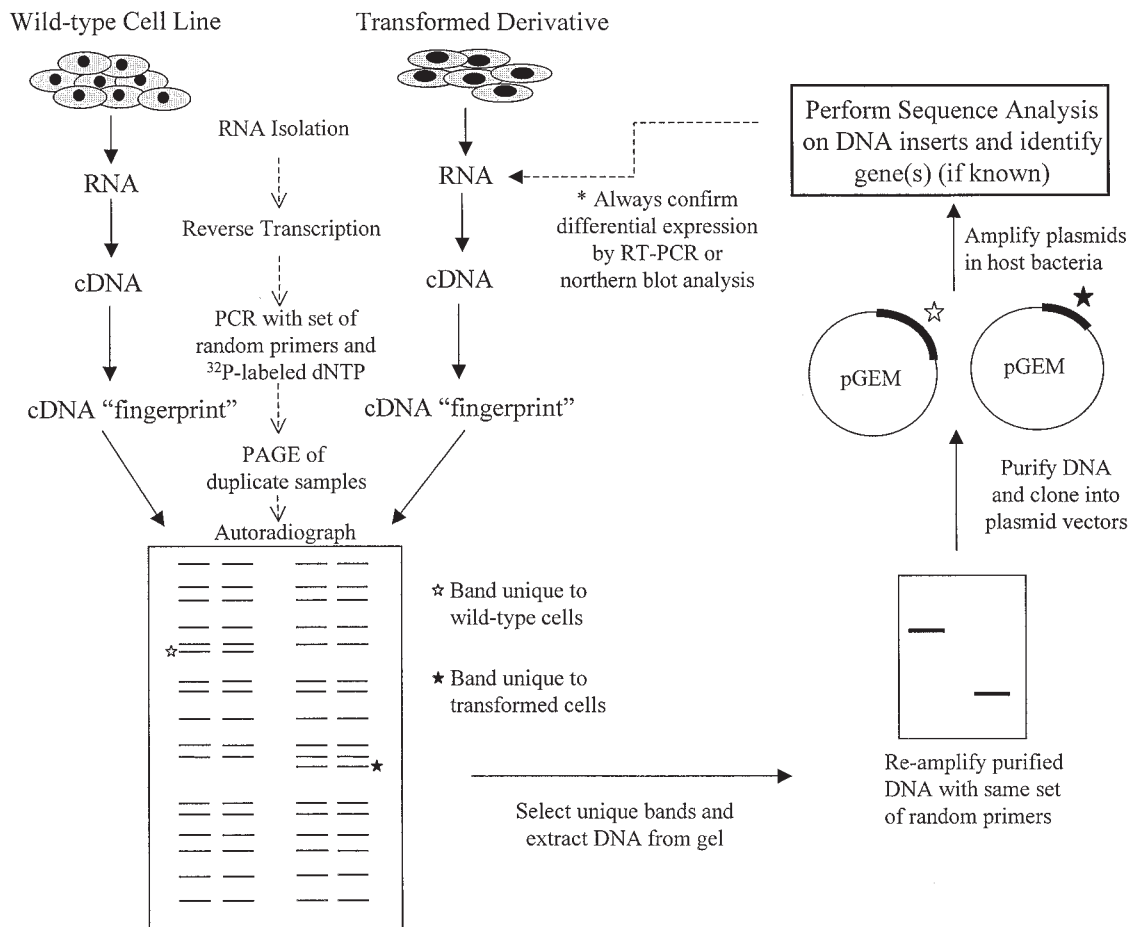
Fig. 8. Differential display RT-PCR analysis. RT-PCR/DD is a useful technique for evaluating differences in gene expression between two or more related cell lines. mRNA is isolated from the cells to be compared and subjected to reverse transcription with oligo-dT anchored primers. PCR is then conducted with a set of random oligonucleotide primers in the presence of ^{32}P-dNTP, generating a cDNA fingerprint composed of a ladder of labeled fragments that represents a subset of the genes expressed by the cell. The cDNA fingerprints from each sample to be compared are loaded onto a polyacrylamide gel and subjected to electrophoretic separation and autoradiography. Unique bands are extracted from the gel, purified, reamplified, and cloned into plasmid vectors where they can be amplified further, manipulated, stored, and sequenced. Sequence analysis often identifies the differentially expressed genes, and differential expression should always be confirmed by quantitative RT-PCR or Northern-blot analysis.

two or more cell populations *(115)*. RT-PCR/DD has been used successfully in the identification of differentially expressed genes in cancers and other diseases, in cells after treatment with specific growth factors, and in many other situations *(116)*. The basic scheme of RT-PCR/DD is shown in Fig. 8. Briefly, high-quality cellular RNA is reverse-transcribed into complementary single-stranded DNA using an anchored oligo-dT primer. Subsequently, PCR is conducted on the DNA in the presence of a labeled nucleotide (^{35}S-dNTP or ^{32}P-dNTP) using an anchored oligo-dT primer in combination with an arbitrary decamer, thus amplifying the 3´-ends of mRNAs. The labeled cDNAs produced by PCR are displayed on a gel such as a DNA sequencing gel, with each set of primers producing between 100–200 bands that are usually between 100–500 base pairs in length. Bands of interest are eluted from the gel and reamplified, after which they can be cloned and sequenced or used as probes in Northern blot analyses or cDNA library screening.

Since the original description of RT-PCR/DD, several improvements have been made that have increased the efficacy of the technique. The anchored oligo-dT RT primer may be modi-

fied regarding the base composition of the anchor or the number of anchored bases, allowing subdivision of the pool of mRNAs into several groups, thereby simplifying the screening of an entire mRNA pool *(117)*. Also, restriction sites may be incorporated into the 5´-ends of the PCR primers, thus facilitating manipulation of the amplified cDNAs during cloning *(117)*. Lastly, the use of fluorescein as an alternative label for RT-PCR/DD yields fluorescent products that can be read on an automated DNA sequencer *(118)*.

Although RT-PCR/DD is used widely for comparative analyses of gene expression, there are several caveats associated with the technique that should be considered before choosing the method. The major drawback to RT-PCR/DD is the large number of false-positives that may appear; on average, ~25–30% of bands are not differentially expressed when evaluated by Northern blot analysis or quantitative RT-PCR, which emphasizes the importance of confirming differential expression before proceeding with further analysis *(119)*. False-positives are frequently due to the inherent randomness of the technique paired with the extreme sensitivity that accompanies

the 35–40 cycles of PCR. The number of false-positives can be reduced by including 3–4 independent replicates of each sample and by employing a stringent band-selection strategy.

Another potential problem is that the 3´-untranslated region is preferentially amplified using RT-PCR/DD *(119)*. Thus, the sequence of a selected band may not match any known sequence in the database because the 3´-untranslated sequence of previously identified genes may not be included in the database. When a differentially expressed band does not match a sequence in the database, it may be identified by screening an appropriate cDNA library. However, library screening is a relatively inefficient means of identifying large numbers of differentially expressed genes.

Lastly, the selection of appropriate samples for RT-PCR/ DD comparison is an important step in ensuring meaningful results. Two unrelated samples, such as a prostate cancer cell line and a patient prostate tumor sample, would be poor candidates for RT-PCR/DD due to the large number of genetic differences between individuals. Companion RT-PCR/DD samples should be related, for example, patient tumor tissue and surrounding normal tissue from the same patient, or a cell line +/- a growth factor. Comparisons between multiple individuals or between multiple cell lines derived from different species will likely produce too many differentially expressed bands to analyze.

A related method, RNA arbitrary primed PCR (RAP-PCR), has been established as an alternative strategy for differential analysis of gene expression among multiple samples (120). In RAP-PCR, an alternative primer is used for cDNA synthesis that does not rely on recognition of the poly-A tail of mRNA. The arbitrary primer used for reverse transcription, usually ~20 base pairs in length, amplifies palindromic sequences typically found in large mRNAs *(120)*. A set of arbitrary primers are then used to amplify regions of the cDNA, as in RT-PCR/DD. With RAP-PCR, there are fewer mRNA bands generated, and the sequences that are amplified are more likely to correspond to mRNA coding regions than those amplified by traditional RT-PCR *(116)*. In addition, a typical RAP-PCR consists of a few initial cycles of low stringency annealing followed by cycles of high-stringency annealing, which increases reproducibility of the bands amplified. The main drawback of RAP-PCR is that mRNAs are not preferentially selected in the RT step, so that rRNAs and tRNAs may be amplified when total RNA is used as starting material *(116)*.

COMPARATIVE GENOMIC HYBRIDIZATION The technique of CGH was developed by Kallioniemi et al. in 1992, and has gained popularity as a tool for the detection of losses and gains in DNA copy number in the entire genome *(121)*. Tumor DNA is labeled with a green fluorochrome, normal reference DNA is labeled with a red fluorochrome, and the two samples are hybridized simultaneously to normal metaphase cells affixed to a slide *(121)*. Using a sensitive camera equipped with fluorescence detection and a computerized analysis system, the green:red ratios can be established along the length of each chromosome. Amplifications in tumor DNA result in an increase in the green:red ratio, while losses of DNA will decrease the green:red ratio of a specific region. For an amplification to be detected by CGH, the amplified DNA must total at

least 2 Mb in size (i.e. amplicon size × degree of amplification ≥ 2 Mb) *(122)*. CGH has been used successfully to document chromosomal gains and losses in astrocytic tumors, to show amplification of the androgen receptor gene in recurrent prostate cancer, and to identify a large number of amplified cellular oncogenes in human neoplasms *(123)*.

REFERENCES

1. Watson, J. D. and Crick, F. H. C. (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**:737, 738.
2. Schumm, D. E. (1997) *Core Concepts in Clinical Molecular Biology*. Lippincott-Raven, Philadelphia.
3. Schumm, D. E. (1995) *Essentials of Biochemistry,* 2nd ed. Little & Brown, Boston.
4. Thoma, F. and Koller, T. (1977) Influence of histone H1 on chromatin structure. *Cell* **12**:101–107.
5. Varshavsky, A. J., Bakayev, V. V., Nedospasov, S. A., and Georgiev, G. P. (1978) On the structure of eukaryotic, prokaryotic, and viral chromatin. *Cold Spring Harbor Symp. Quant. Biol.* **42**:457–473.
6. Lamond, A. I. and Earnshaw, W. C. (1998) Structure and function in the nucleus. *Science* **280**:547–553.
7. Long, E. O. and Dawid, I. B. (1980) Repeated genes in eucaryotes. *Annu. Rev. Biochem.*, **49**, 727–764.
8. Jordan, E. G. (1978) *The Nucleolus,* 2nd ed. Oxford University Press, Oxford.
9. Tyler-Smith, C. (1993) Mammalian chromosome structure. *Curr. Opin. Genet. Dev.* **3**:390–397.
10. Biessmann, H. and Mason, J. M. (1994) Telomeric repeat sequences. *Chromosoma* **103**:154–161.
11. Counter, C. M. (1996) The roles of telomeres and telomerase in cell life span. *Mutation Res.* **366**.
12. Hanke, J. H., Hambor, J. E., and Kavathas, P. (1995) Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8 alpha cell-specific enhancer. *J. Mol. Biol.* **246**:63–73.
13. Arzimanoglou, I. I., Gilbert, F., and Barber, H. R. (1998) Microsatellite instability in human solid tumors. *Cancer* **82**:1808–1820.
14. Sutherland, G. R. (1995) Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA* **92**:3636–3641.
15. Craig, I. (1991) Methylation and the fragile X. *Nature* **349**:742–743.
16. (HDCRG) Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is unstable on Huntington's disease chromosomes. *Cell* **72**:971–983.
17. Hoffmann-Berling, H. (1982) DNA unwinding enzymes. *Prog. Clin. Biol. Res.* **102**:89–98.
18. Kato, S. and Kikuchi, A. (1998) DNA topoisomerase: the key enzyme that regulated DNA super structure. *Nagoya J. Med. Sci.* **61**:11–26.
19. Heywood, L. A. and Burke, J. F. (1990) Mismatch repair in mammalian cells. *Bioessays* **12**:473–477.
20. Auerbach, A. D. and Verlander, P. C. (1997) Disorders of DNA replication and repair *Curr. Opin. Pediat.* **9**:600–616.
21. Nojima, H. (1997) Cell cycle checkpoints, chromosome stability and the progression of cancer. *Human Cell* **10**:221–230.
22. Fotedar, R. and Fotedar, A. (1995) Cell cycle control of DNA replication. *Prog. Cell Cycle Res.* **1**:73–89.
23. Green, R. and Noller, H. F. (1997) Ribosomes and translation. *Ann. Rev. Biochem.* **66**:679–716.
24. Persson, B.C. (1993) Modification of tRNA as a regulatory device. *Mol. Microbiol.* **8**:1011-1016.
25. Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J., and Soll, D. (1985) Structure and transcription of eukaryotic tRNA genes. *Crit. Rev. Biochem.* **19**:107–144.
26. Goldberg, S., Schwartz, H., and Darnell, J. E. (1977) Evidence from UV transcription mapping in HeLa cells that heterogeneous nuclear

RNA is the messenger RNA precursor. *Proc. Natl. Acad. Sci. USA* **74**:4520–4523.

27. Sharp, P. A. (1988) RNA splicing and genes. *JAMA* **260**:3035–3041.

28. Lewin, B. (1994) Genome size and genetic content. *In*: *Genes V* (Lewin, B., ed.), Oxford University Press, New York, pp. 657–674.

29. Chou, K. C. and Kezdy, F. J. (1994) Kinetics of processive nucleic acid polymerases and nucleases. *Anal. Biochem.* **221**:217–230.

30. Kollmar, R. and Farnham, P. J. (1993) Site-specific initiation of transcription by RNA polymerase II. (Review). *Proc. Soc. Exp. Biol. Med.* **203**:127–139.

31. Tantravahi, J. and Alvira, M. (1993) Characterization of the mouse beta globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. *Mol. Cell. Biol.* **13**:578–587.

32. Varani, G. (1997) A cap for all occasions. *Structure* **5**:855–858.

33. Munroe, D. and Jacobson, A. (1990) Tales of poly(A): A review. *Gene* **91**:151–158.

34. Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994) The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp. Cell Res.* **211**:400–407.

35. Balvay, L., Libri, D., and Fiszman, M. Y. (1993) Pre-mRNA secondary structure and the regulation of splicing. *Bioessays* **15**:165–169.

36. Staley, J. P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**:315–326.

37. Edwalds-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997) Alternative poly(A) site selection in complex transcription units: means to an end? (Review). *Nucleic Acids Res.* **25**:2547–2561.

38. Morrissey, J. P. and Tollervey, D. (1995) Birth of the snRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. *Trends Biochem. Sci.* **20**:78–82.

39. Cech, T. R. (1990) Self-splicing of group I introns. *Ann. Rev. Biochem.* **59**:543–568.

40. Jacquier, A. (1990) Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem. Sci.* **15**:351–354.

41. Suzuki, K. and Watanabe, M. (1994) Modulation of cell growth and mutation induction by introduction of the expression vector of human hsp70 gene. *Exp. Cell Res.* **215**:75–81.

42. Thomas, R. (1993) The denaturation of DNA. *Gene* **135**:77–79.

43. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

44. Presnell, S. C. (1997) Nucleic acid blotting techniques. *In: Molecular Diagnostics for the Clinical Laboratorian.* (Coleman, W. B. and Tsongalis, G. J., eds.), Humana Press, Totowa, NJ, pp. 63–88.

45. Boltz, E. M., Kefford, R. F., Leary, J. A., Houghton, C. E., and Friedlander, M. L. (1989) Amplification of c-ras-Ki oncogene in human ovarian tumours. *Int. J. Cancer* **43**:428–430.

46. Hynes, N. E. (1993) 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.* **4**:19–26.

47. Wang, S. J., Scavetta, R., Lenz, H. J., Danenberg, K., Danenberg, P. V., and Schonthal A. H. (1995) Gene amplification and multidrug resistance induced by the phosphatase-inhibitory tumor promoter, okadaic acid. *Carcinogenesis* **16**:637–641.

48. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich H., et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis of sickle cell anemia. *Science* **230**:1350–1354.

49. Mullis, K. B. and Faloona, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**:335–350.

50. Sorscher, D. H. (1997) DNA Amplification Techniques. *In: Molecular Diagnostics for the Clinical Laboratorian* (Coleman, W. B. and Tsongalis, G. J., eds), Humana Press, Totowa, NJ, pp. 89–101.

51. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res.* **6**:3543–3557.

52. Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., et al. (1993) 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.* **2**:275–287.

53. Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989) Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from Thermus aquaticus. *J. Biol. Chem.* **264**:6427–6437.

54. Fanning, S. and Gibbs, R. A. (1997) PCR in Genome Analysis. *In*: *Genome Analysis. Volume 1: Analyzing DNA.* (Birren, B., Green, E. D., Klapholz, S., Myers, R. M., and Roskams, J., eds.) Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 249–299.

55. Heid, C. A., Stevens, J., Livak, K. J., and Williams, P.M. (1996) Real time quantitative PCR. *Genome Res.* **6**:986–994.

56. Kaijalainen, S., Karhunen, P. J., Lalu, K., and Lindstrom, K. (1993) An alternative hot start technique for PCR in small volumes using beads of wax-embedded reaction components dried in trehalose. *Nucleic Acids Res.* **21**:2959–2960.

57. Bassam, B. J. and Caetano-Anolles, G. (1993) PCR using mineral oil and paraffin wax. *BioTechniques* **14**:30–34.

58. Roux, K. H. (1994) Using mismatched primer-template pairs in touchdown PCR. *BioTechniques* **16**:812–814.

59. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**:4008.

60. Foord, O. S. and Rose, E. A. (1994) Long-distance PCR. *Genome Res.* **3**:S149–61.

61. Barnes, W. M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**:2216–2220.

62. Ferre, F. (1992) Quantitative or semi-quantitative PCR: reality versus myth. *Genome Res.* **2**:1–9.

63. Raeymakers, L. (1996) A commentary on the practical applications of competitive PCR. *Genome Res.* **5**:91–94.

64. Lau, Y. F. and Kan, Y. W. (1983) Versatile cosmid vectors for the isolation, expression, and rescue of gene sequences: studies with the human alpha-globin gene cluster. *Proc. Natl. Acad. Sci. USA* **80**:5225–5229.

65. Lepourcelet, M., Andrieux, N., Giffon, T., Pichon, L., Hampe, A., Galibert, F., et al. (1996) Systematic sequencing of the human HLA-A/HLA-F region: establishment of a cosmid contig and identification of a new gene cluster within 37 kb of sequence. *Genomics* **37**:316–326.

66. Rowen, L., Kobori, J. A., and Scherer S. (1982) Cloning of bacterial DNA replication genes in bacteriophage lambda. *Mol. Gen. Genet.* **187**:501–509.

67. Young, R. A., Bloom, B. R., Grosskinsky, C. M., Ivanyi, J., Thomas, D., and Davis, R. W. (1985) Dissection of Mycobacterium tuberculosis antigens using recombinant DNA. *Proc. Natl. Acad. Sci.USA* **82**:2583–2589.

68. Grunstein, M. and Hogness, D. S. (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**:3961–3965.

69. Wang, S. Z. and Esen, A. (1985) Screening expression libraries with nonradioactive immunological probes. *Gene* **37**:267–269.

70. Szostak, J. W. and Blackburn, E. H. (1982) Cloning yeast telomeres on linear plasmid vectors. *Cell* **29**:245–255.

71. Burgers, P. M. and Percival, K. J. (1987) Transformation of yeast spheroplasts without cell fusion. *Anal. Biochem.* **163**:391–397.

72. Larin, Z., Monaco, A. P., and Lehrach, H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA* **88**:4123–4127.

73. Burke, D. T., Carle, G. F., and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**:806–812.

74. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., et al. (1992) Cloning and stable maintenance of 300-

kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**:8794–9797.

75. Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., et al. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**:84–89.

76. Maxam, A. M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* **74**:560–564.

77. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.

78. Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., and Mathies, R. A. (1995) Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. USA* **92**:4347–4351.

79. Zimmerman, J. H., Voss, H., Schwager, C., Stegemann, J., and Ansorge, W. (1988) Automated Sanger dideoxy sequencing reaction protocol. *FEBS Lett.* **233**:432–436.

80. Mardis, E. R. and Roe, B. A. (1989) Automated methods for single-stranded DNA isolation and dideoxynucleotide DNA sequencing reactions on a robotic workstation. *BioTechniques* **7**:840–850.

81. Tabor, S. and Richardson, C. C. (1995) A single residue in DNA polymerases of the Escherichia coli DNA polymerase 1 family is critical for distinguishing between deoxy- and dideoxy-ribonucleotides. *Proc. Natl. Acad. Sci. USA* **92**:6339–6343.

82. Schwartz, D. C. and Cantor, C. R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**:67–74.

83. Sader, H. S., Hollis, R. J., and Pfaller, M. A. (1995) The use of molecular techniques in the epidemiology and control of infectious diseases. *Clin. Lab. Med.* **15**:407–431.

84. Anand, R. (1986) Pulsed field gel electrophoresis: A technique for fractionating large DNA molecules. *Trends Genet.* **2**:278–286.

85. Turmel, C., Brassard, E., Slater, G.W., and Noolandi, J. (1990) Molecular detrapping and band narrowing with high frequency modulation of pulsed field electrophoresis. *Nucleic Acids Res.* **18**:569–575.

86. Carle, G. F., Frank, M., and Olson, M. V. (1986) Electrophoretic separation of large DNA molecules by periodic inversion of the electric field. *Science* **234**:1582–1585.

87. Pourzand, C. and Cerutti, P. (1993) Genotypic mutation analysis by RFLP/PCR. *Mutation Res.* **288**:113–121.

88. Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., et al. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**:2503–2516.

89. Hayashi, K. (1992) PCR-SSCP: a method for detection of mutations. *Genet. Anal. Tech. Appl.* **9**:73–79.

90. Bhattacharyya, A. and Lilley, D. M. J. (1989) The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.* **17**:6821–6840.

91. Keen, J., Lester, D., Inglehearn, C., Curtis, A., and Bhattacharyya, S. (1991) Improved detection of heteroduplexes on Hydrolink gels. *Trends Genet.* **7**:5.

92. Molinari, R. J., Conners, M., and Shorr, R. G. (1993) Hydrolink gels for electrophoresis. *In: Advances in Electrophoresis,* Vol. 6 (Chrambach, A., Dunn, M. J. and Radola, B. J., eds.), VCH, New York, pp. 44–60.

93. Fodde, R. and Losekoot, M. (1994) Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Human Mutation* **18**:83–94.

94. Myers, R. M., Maniatis, T., and Lerman, L. S. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155**:501–510.

95. Riesner, D., Steger, G., Zimmat, R., Owens, R.A., Wagenhofer, M., Hillen, W., et al. (1989) Temperature-gradient gel electrophoresis of nucleic acids: Analysis of conformational transitions, sequence variations, and protein-nucleic acid interactions. *Electrophoresis* **10**:377–389.

96. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.

97. Kotewicz, M. L., D'Alessio, J. M., Driftmier, K. M., Blodgett, K. P., and Gerard, G. F. (1985) Cloning and overexpression of Maloney murine leukemia virus reverse transcriptase in Escherichia coli. *Gene* **35**:249–258.

98. Schraml, P., Shipman, R., Stulz, P., and Ludwig, C. U. (1993) cDNA subtraction library construction using a magnet-assisted subtraction technique (MAST). *Trends Genet.* **9**:70–71.

99. Larrick, J. W. and Siebert, P. D. (1995) Reverse Transcriptase PCR. *In: Ellis Horwood Series in Pharmaceutical Technology* (Rubinstein, M. H., ed.), Ellis Horwood Ltd., Hertfordshire, U.K.

100. Lai, C. J., Markoff, L. J., Zimmerman, S., Cohen, B., Berndt, J. A., and Chanock, R.M., R.M.C. (1980) Cloning DNA sequences from influenza viral RNA segments. *Proc. Natl. Acad. Sci. USA* **77**:210–214.

101. Kaguni, J. and Kaguni, L. S. (1992) Enzyme-labeled probes for nucleic acid hybridization. *Bioanal. Appl. Enzymes* **36**:115–127.

102. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**:60–89.

103. Krieg, P. A. and Melton, D. A. (1987) In vitro RNA synthesis with SP6 RNA polymerase. *Methods Enzymol.* **155**:397–415.

104. Nakagami, S., Matsunaga, H., Oka, N., and Yamane, A. (1991) Preparation of enzyme-conjugated DNA probe and application to the universal probe system. *Anal. Biochem.* **198**:75–79.

105. Yamaguchi, K., Zhang, D., and Byrn R. (1994) A modified nonradioactive method for northern blot analysis. *Anal. Biochem.* **218**:343–346.

106. Isaac, P. G., Stacey, J., and Clee, C. M., C.M.C. (1995) Nonradioactive probes. *Mol. Biotechnol.* **3**:259-265.

107. Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., et al. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**:2088–2090.

108. Bowcock, A. M. (1993) Molecular cloning of BRCA1: A gene for early onset familial breast and ovarian cancer. *Breast Cancer Res. Treat.* **28**:121–135.

109. Lynch, H. T. and Smyrk, T. (1996) Hereditary nonpolyposis colorectal cancer (Lynch syndrome): an updated review. *Cancer* **78**:1149–1167.

110. Ramsay, G. (1997) DNA chips: State-of-the art. *Nature Biotechnol.* **16**:40-44.

111. Shalon, D., Smith, J. S., and Brown, P. O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6**:639–645.

112. Drmanac, S., Stavropoulos, N. A., Labat, I., Vonau, J., Hauser, B., Soares, M. B., et al. (1996) Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. *Genomics* **37**, 29–40.

113. Hacia, G. H., Brody, L. C., Chee, M. S., Fodor, S. P. A., and Collins, F. S. (1996) Detection of heterozygous mutations in BRCA1 using high-density oligonucleotide arrays and two color fluorescence analysis. *Nature Genet.* **14**:441–447.

114. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **270**:680–686.

115. Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**:967–971.

116. Liang, P. and Pardee, A. B. (1995) Recent advances in differential display. *Current Opin. Immunol.* **7**:274–280.

117. Liang, P., Zhu, W., Zhang, X., Guo, Z., O'Connell, R. P., and Averboukh, L. (1994) Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res.* **22**:5763–5764.

118. Ito, T., Kito, K., Adati, N., Mitsui, Y., Hagiwara, H., and Sakaki, Y. (1994) Fluorescent differential display: arbitrarily primed RT-PCR fingerprinting on an automated DNA sequencer. *FEBS Lett.* **351**:231–236.

119. Sunday, M. E. (1995) Differential Display RT-PCR for identifying gene expression in the lung. *Am. J. Phsiol.* **269**:L273–L284.

120. Welsh, J., Chada, K., Dalal, S. S., Chang, R., Ralph, D., and McClelland, M. M. (1992) Arbitrary primed PCR fingerprinting of RNA. *Nucleic Acids Res.* **20**:4965–4970.

121. Kallioniemi, A., Kallioniemi, O.-P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., et al. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**:818–821.

122. Kallioniemi, O.-P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F. M., Gray, J. W., et al. (1994) Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer* **10**:231–243.

123. Knuutila, S., Bjorkqvist, A.-M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., et al. (1998) DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.* **152**:1107–1123.
.
.