Review Series: Nucleic Acids in Medicine

Nucleic acids in disease: Information launch-pad

RAM H. DATAR

INTRODUCTION

The past decade or so has seen a remarkable emergence of the so-called 'translational medical research', where knowledge in biological research, particularly molecular biology, has allowed transition from the quest of academic mechanistic approaches to actual 'application' in day-to-day clinical management. Thus, technologies of molecular biology can be used in the diagnosis, prognosis, and even therapy of diseases. Physicians can benefit by combining their clinical acumen with knowledge of disease at the molecular level and employing options that are available to diagnose and treat these diseases using the tools of molecular biology. Many clinicians from 'advantaged' establishments in metropolitan cities are already aware of the potential use of these tools. But they are too busy administering the existing clinical management options for the huge number of patients to pause and make conscious efforts to gather the relevant information about the capability and tools of molecular medicine. By compiling relevant information on molecular medicine, this series aims to complement the existing clinical management, which has the potential to substantially reduce medical costs associated with current symptomatic treatment. The overall aim of this series, therefore, is to bring into 'molecular focus' the art that physicians are adept at—managing disease effectively.

Before delving further into the clinical applications of molecular medicine, it must be emphasized that molecular biology (and hence molecular medicine) involves the study and use of nucleic acids and proteins, which form an equally vital cog in the wheel of control of cellular functions. It is essentially because there has been an almost concurrent and explosive growth in the knowledge of both these classes of macromolecules that the scope of this series is restricted to focus on nucleic acids. The subsequent articles in this series will cover various aspects of applications of nucleic acids in diagnosis, prognosis, therapy, medical commerce and future applications. Beginning with a brief recapitulation of basic concepts in molecular biology, this article presents an overview of the molecular technologies that are available for clinical laboratories, followed by a description of disease-associated alterations in nucleic acids that can be used clinically.

RELEVANT CONCEPTS IN MOLECULAR MEDICINE

DNA and RNA structure and function

DNA is a relatively larger and more stable antiparallel, double-stranded polydeoxyribonucleotide molecule where each deoxyribonucleotide contains a phosphate moiety, a deoxyribose pentose sugar, and either one of four possible bases: adenine (A), thymine (T), guanine (G) or cytosine (C). Of these, A and G are the purines (2-ring bases) while T and C are the pyrimidines (1-ring bases). Typical RNA is a relatively smaller, more labile and mostly single-stranded molecule, with its component ribonucleotides containing a phosphate moiety, a ribose-type pentose sugar and one of the four bases with the base uracil in place of thymine in DNA. DNA exists mainly as a nuclear genome, although a small yet crucial DNA complement also exists as a mitochondrial genome. The DNA in the nucleus is complexed with proteins such as histones and non-histones to form chromatin, which is the chemical structural scaffolding of chromosomes (Fig. 1). Technically speaking, each chromosome is a single huge DNA duplex supercoiled via solenoid structures.[1] The RNA, which can be distinguished into three forms (ribosomal or rRNA, transfer or tRNA and messenger or mRNA) is transcribed from the coding sequences along the genomic DNA.

From a molecular perspective, a gene is a specific nucleotide sequence that is transcribed into RNA. Structural genes encode proteins. In humans, as in other eukaryotes, the coding or 'expressed' regions (exons) are interspersed with non-coding intervening sequences (introns). While the primary gene transcripts (heterogenous nuclear RNA or hnRNA) contains both exons and introns, processed mRNA results after splicing together of exons in the correct order. This mRNA is then transported to the cytoplasm where it is translated into proteins, using ancillary support from tRNA molecules which ferry the correct amino acids to the ribosomes (the physical site for protein synthesis) composed of rRNA and ribosomal proteins.

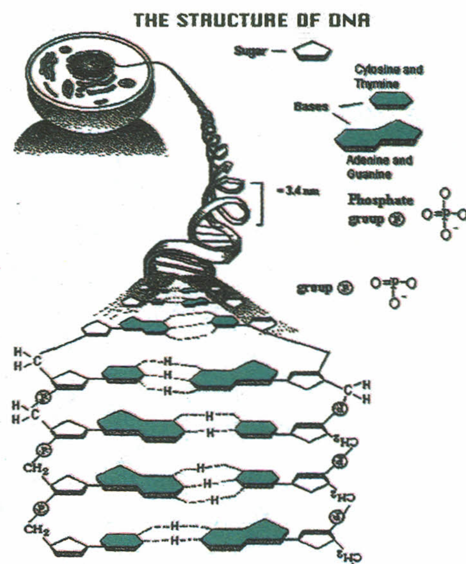

FIG 1. **Basic composition of DNA.** Differences between DNA and RNA structures are detailed in the text.

Department of Clinical Pathology, University of Southern California, Los Angeles, USA

© *The National Medical Journal of India 2000*

Thanks to the global non-profit collaboration through the Human Genome Project (HGP) and the commercial venture by Celera Genomics, the first draft of the human genomic organization was recently completed.[2,3] This is expected to result in a vast amount of useful data about the human genome. Although the predicted number of genes in the human diploid cell ranges from 40 000 to 140 000, not all the genes are transcriptionally active in all the cell lineages at all times. Gene expression differs in a development-specific, differentiation-specific and tissue-specific manner. Let us briefly consider how this fine-tuned control of gene activity takes place.

Gene regulation

The eukaryotic genes have complex regulatory elements in the proximity of (or at some distance from) the actual gene. The trans-acting proteins (transcription factors or TF) recognize and bind specific *cis*-acting sequence elements (Latin *cis*=close or near). The *cis*-acting elements are about 8–10 nucleotides long, and fall into four classes: promoters, enhancers, silencers and response elements. While promoter elements serve to initiate the transcription by binding with RNA polymerase and enhancer elements bind with TF to enhance transcription, silencer elements are negative regulators which reduce transcription. Response elements are found in inducible genes, the expression of which is controlled by external factors such as hormones, growth factors or by internal signalling molecules such as cyclic AMP (cAMP). These sequences are located a short distance upstream of the promoter elements, usually within a thousand nucleotides of the transcription initiation site. A wide variety of response elements for factors such as glucocorticoids, retinoic acid, growth factors, cAMP, etc. are documented in different genes.[4] The inducible genes are transcribed at high levels following binding of the signalling factor to the corresponding response element. Besides these elements regulating individual genes, positive *cis*-acting regulatory elements may occasionally coordinate the expression of genes in *gene clusters*, examples of which are found located 50–60 kilobases upstream of the genes for α and β globin gene clusters for binding with specific TF.[5]

The binding of TF to target regulatory sequences usually results due to extensive correspondence between the surface of a domain of the protein and the surface features of the DNA double helix in a specific region. Typically, two distinct functional domains can be identified in a TF, a DNA binding domain and an activation domain. Many TFs fine-tune their regulatory potential by homodimer versus heterodimer formation. In the latter, each monomer binds to a half-target sequence, allowing combination of unique controls. An example of such a TF combination are the eukaryotic *jun* and *fos* proteins. Thus, while the *jun–jun* homodimer binds weakly to AP-1 promoter, the *jun–fos* heterodimer binds tightly to the same sequence.[6,7] Thus, expression of any gene is a result of the interplay between extracellular and intracellular signalling molecules, *cis*-regulatory elements and TFs.

Cell cycle and apoptosis

Molecular events in the cell cycle and apoptosis have been extensively reviewed elsewhere in the recent literature.[8,9] Brief descriptions of the cell cycle and apoptosis follow.

Growing and dividing cells in eukaryotes traverse through a cell division cycle which is divisible into four phases: the G1 (first gap), S (synthetic phase), G2 (second gap) and M (mitotic phase). The period between two cell divisions (G1+S+G2 phases) is referred to as interphase (Fig. 2). A number of cell division cycle

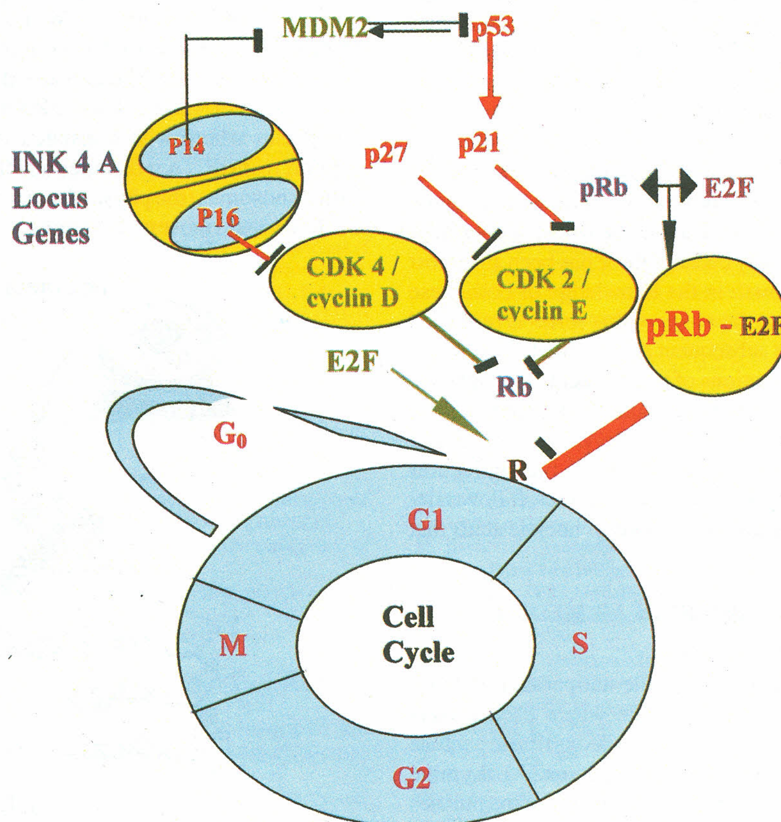

FIG 2. **Cell cycle regulation.** In normal cells, the positive regulators (shown in green) are functionally balanced by negative regulators (shown in red), particularly at the G1–S restriction point (R). This balance may be disturbed in disease.

(cdc) genes act both at the G1 commitment (restriction or R) point, or the boundary between G2 and M phases; and phosphorylation is known to be the general control mechanism for the cdc gene products. The cdc genes encode cyclin-dependent kinases, or CDKs, which complex with cyclin proteins. The cyclin–CDK complex formation results in phosphorylation (and inactivation) of RB (first discovered in retinoblastoma) protein, leading to release and nuclear translocation of the transcription factor E2F. This marks the exit of the cell from G1 into S phase, where E2F induces transcription of a number of DNA synthetic genes. As an additional regulatory mechanism, the phosphorylation function of cyclin–CDK complexes is inhibited by CDK inhibitor proteins (CDKi), such as the *cip/kip* family member genes encoding p27 and p21, and INK4 A family genes encoding p16 and p14/p19. Besides these proteins, p53 plays a dual role in the cell cycle, both at the G1–S checkpoint via upregulation of p21 expression, and at the G2–M transition by inhibiting cdc2 and cyclin B1 production. A normal cell has mechanisms to sense damage to its replicative machinery, and unless these damages are repaired in G1, apoptosis is activated and the cell exits the cell cycle and eventually dies.

Apoptosis is programmed or cell-initiated cell death that is induced in response to physiological or developmental signals to remove cells. The signal for apoptosis comes, in most part, from the 'master switch': p53 protein. In a dividing cell that has accumulated oxidative DNA damage due to viruses, chemotherapeutic drugs or gamma and UVB radiations, p53 recognizes the damaged integrity of DNA in the G1 phase of the cell cycle. It then either chooses to bring about growth arrest by exiting the cell cycle or forcing the cell to undergo apoptosis by exerting a dual effect on *bcl 2* family proteins. Thus, it represses transcription of anti-apoptotic *bcl 2* family genes while simultaneously stimulating transcription of pro-apoptotic BAX family genes. Having received the 'self-destruct' signal, enzymes called pro-caspases are activated to caspases (also referred to as interleukin-1β converting enzyme or ICE family members), which begin a cascade activating a series of proteases, nucleases and other catabolic enzymes. The cells dying by apoptosis usually shrink and their membrane undergoes blebbing. This is accompanied by the characteristic chromatin condensation and DNA ladder formation due to cleavage by Mg- and/or Ca-dependent endonucleases. The small, membrane-bound vesicles generated by cell fragmentation are efficiently engulfed by macrophages in the surrounding tissue and are completely digested, bypassing inflammation, unlike in necrosis.

A number of proteins are required to maintain or prevent the cell cycle, and to induce or inhibit apoptosis. The targets of these events are generally nuclear proteins such as products of oncogenes, TFs, suppressor gene products and products of mismatch repair genes.

Mismatch repair system

Short tandem repeat (STR) sequences or dinucleotide and trinucleotide repeat sequences (microsatellites) in genomic DNA are prone to replication errors due to strand slippage. The mismatch repair (MMR) genes encode a repair system, which detects the mismatches and excises and replaces the incorrect bases.

A total of five human mismatch repair or 'mutator' genes are now identified; hMSH 2 (human homologue 2 of *E. coli Mut s*) located at chromosome 2p21–22, hMLH1, hPMS1 and hPMS2 (homologues of *E. coli Mut l gene*) at chromosomal locations 3p21.3, 2q31–33 and 7p22, respectively, and hMSH6 (2p16).[10] The loss of function of these genes is caused either by mutations or silencing of gene promoters by DNA methylation, and results in accumulation of uncorrected replication errors which manifest as repeat expansions, probably making growth suppressor genes and oncogenes vulnerable to mutations. This is of particular relevance in cancer.

Thus, any lasting interruption in function of one of the crucial control elements of gene expression, cell cycle, apoptosis or mismatch repair has a potential for disrupted cellular pathways, which may potentially result in disease.

GENE REARRANGEMENTS

Mutations

In spite of many safeguards, DNA replication errors do occur, which may affect one or more nucleotides. Besides, mutagens such as viruses, chemicals or gamma and UVB radiations cause permanent nucleotide sequence changes or 'mutations' in either single nucleotide locations (point mutations) or larger chromosomal regions (chromosomal aberration). Since a vast majority of the human genome is non-coding, mutations occurring in this region are most often non-affecting (unless, as is beginning to be understood, they occur in gene regulatory sequences, whereby the binding of TF and other regulators is affected). The mutations occurring in the coding regions or exons, on the other hand, can alter the structure and hence the functions of the proteins. While the germline mutations occurring in gametic cells may be transmitted to the next generation through the gametes, the somatic mutations are not hereditary. Substitution of a purine base by another purine, or that of a pyrimidine by another pyrimidine, is called 'transition'. The replacement of a purine by a pyrimidine, or vice versa, is referred to as 'transversion'. Transitions are more common than transversions. Point mutations are classified as *silent* (when mutation in a base does not change the amino acid sequence), *neutral* (when a base mutation changes an amino acid but the replacement does not affect a protein structure and function), *mis-sense* (when base alteration leads to amino acid change that causes structural and functional protein change) and *non-sense* mutations (when a base change replaces a codon for an amino acid by a translation termination codon, prematurely truncating the protein). Other mutations that lead to either an 'insertion' or a 'deletion' of a nucleotide pair in a sequence lead to *frameshift mutation*, which can either change the amino acid sequence of a protein or may result in protein truncation due to altered reading frame of codons. A mutation in splice donor or splice acceptor sites can result in exon skipping or an intron inclusion in the final mRNA as shown in Fig. 3. Such alternative splice variants may result in dysfunctional proteins, a common feature of neoplastic cells.

Nucleotide sequence alterations due to point mutations can be identified using the amazing tools of molecular biology—the type II restriction endonucleases. These bacterial enzymes cleave DNA in a sequence-specific manner. If a sequence mutation associated with a disease results in either a loss or creation of a new restriction endonuclease cleavage site, it leads to restriction fragment-length polymorphism (RFLP), which can be identified using Southern hybridization or PCR techniques described below. Not only point mutations but alterations over larger stretches of DNA such as large insertions or deletions, and the repeat expansions described above can also characterize or cause disease conditions.[11] These too can be identified using the techniques of Southern hybridization or PCR.

Chromosomal aberrations

Classical cytogenetics has been pivotal in identifying a large number of numerical and structural aberrations in chromosome complements. The mechanisms of numerical changes (aneup-

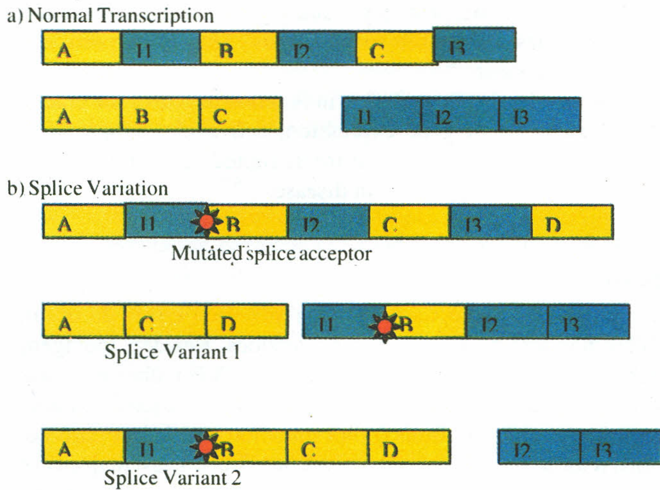

FIG 3. **Splice variants resulting from point mutation**. A point mutation at 'splice-acceptor' site before exon B yield two alternative splice variants, one missing an exon and the other containing an intron. Both these translate into altered protein. (A, B, C and D: exons; I1, I2 and I3: introns)

loidy) and other chromosomal aberrations commonly seen in various diseases have been reviewed by Tucker and Preston.[12] The trisomy of chromosome 21 in Down syndrome, trisomy of chromosome 18 in Edward syndrome, trisomy of chromosome 13 in Patau syndrome, presence of an extra copy of chromosome X in Klinefelter syndrome (47,XXY), etc. are some of the celebrated examples of numerical chromosomal aberrations. Ten- to hundred-fold amplifications of regions of chromosomes are a typical feature of a number of cancers, where highly amplified genes in these regions result in pathological overproduction of proteins. The amplified gene regions mostly exist as 1-megabase-sized DNA circles called *double minutes* in the nucleoplasm, while in 5% of the cells these can be discerned intrachromosomally as distinctive *homogeneously staining regions (HSRs)* with repetitive banding pattern. Examples of genes thus identified are NMYC in neuroblastoma, EGFR in epidermal cancer, HER2/neu in breast cancer, etc. Another phenomenal achievement by classical cytogeneticists is the characterization of various structural chromosomal aberrations such as translocations, inversions and deletions that mark neoplastic disease conditions.[13] All these numerical and structural chromosomal aberrations can now be identified with high accuracy using molecular genetic techniques such as FISH (fluorescent *in situ* hybridization), chromosomal painting and CGH (comparative genomic hybridization) which are described briefly below.

TECHNOLOGIES AND TOOLS

Southern and northern hybridization

This simple technique to transfer size-fractionated denatured DNA or RNA molecules from an agarose gel to a nitrocellulose or nylon membrane filter, developed by Edwin M. Southern[14] in 1975 has served as an invaluable tool for molecular biologists the world over. A radioisotopically or non-isotopically labelled 'probe' specific for a target sequence is hybridized with the template DNA immobilized on the membrane and the hybrids denoting the presence of target sequence are detected by the appropriate method.

Polymerase chain reaction

Initially developed as an *in vitro* alternative for cell-based gene cloning,[15] PCR has become an essential technique of any modern laboratory. Numerous modifications of PCR have evolved over the last two decades, which have spawned off routine applications in clinical diagnoses to detect the presence of pathogen DNA sequences as well as alterations such as mutations and deletions in human gene sequences in disease conditions.[16] Briefly, this technique uses synthetic primer sequences that are complementary to sequences flanking the target DNA sequence. Through repeated cycles of primer annealing, chain extension and denaturation (Fig. 4), the target sequence is amplified several million-fold and detected by agarose gel electrophoresis. The use of thermally resistant DNA polymerases has been an important factor in the development of automated thermal cyclers. PCR has even replaced the classic combination of RFLP and Southern hybridization to detect changes in DNA sequence.

RT-PCR

This is among the most sensitive techniques available for detection of gene transcripts *in vitro*. Here the mRNA template is copied into a complementary DNA (cDNA) transcript using the enzyme reverse transcriptase (RT), followed by amplification of the cDNA using PCR. Three types of RTs are commercially available, which include the mesophilic retroviral enzymes derived from Avian Myeloblastosis Virus (*AMV*) and Moloney Murine Leukemia Virus (*M-muLV*), and a thermostable DNA polymerase *rTth* from the bacterium *Thermus thermophilus*. Three types of sequences can be used to prime reverse transcription reactions; oligo$(dT)_{12-18}$ primers (which bind to the 3' poly-A tail present in most eukaryotic mRNAs), random hexanucleotides (which can bind at any complementary sequence throughout the length of mRNA), or gene-specific oligonucleotides (which are complementary to unique regions towards the 3' end of transcripts).

In situ RT-PCR can be carried out for the detection of mRNAs in fixed cell preparation on slides, giving the molecular identification a morphological correlate.[17] Exonuclease-free DNA polymerase and pretreatment of the cells with T4 ligase to repair the nicks in endogenous DNA are employed to reduce the artifactual false-positive signals typically seen in the apoptotic cells.

Quantitative real-time RT-PCR is the relatively recent modification of *in vitro* RT-PCR that permits precise determination of the absolute abundance of a specific transcript. Real-time PCR employs fluorescent TaqMan probes[18] or molecular beacons[19] to track the kinetic progress of PCR as the amplification proceeds. The same technique can also be applied for detection of extremely low copy viral DNA molecules.[20] Competitive commercial development of automated real-time cyclers and enhancement of fluorochromes for detection will probably lower the currently prohibitive cost.

Mutation detection assays

Commonly used methods for detecting mutations are single-strand conformation polymorphism (SSCP) and heteroduplex analysis (HA). Besides these, a host of other technologies have evolved over the years, yielding increasingly accurate results, many with a promise for high-throughput capability. The prominent ones among these are described in brief below.

In *SSCP*, PCR products (<300 base pairs) are denatured to single strands and allowed to refold on themselves and resolved by non-denaturing polyacrylamide gel electrophoresis (ND-PAGE). A strand differing from another by one or more nucleotide changes assumes a different conformation. Detection is achieved by silver staining, or by autoradiography or fluorescence by virtue of labelled PCR primers. A limitation of SSCP, however, is that the absence of a conformation change may not mean absence of a

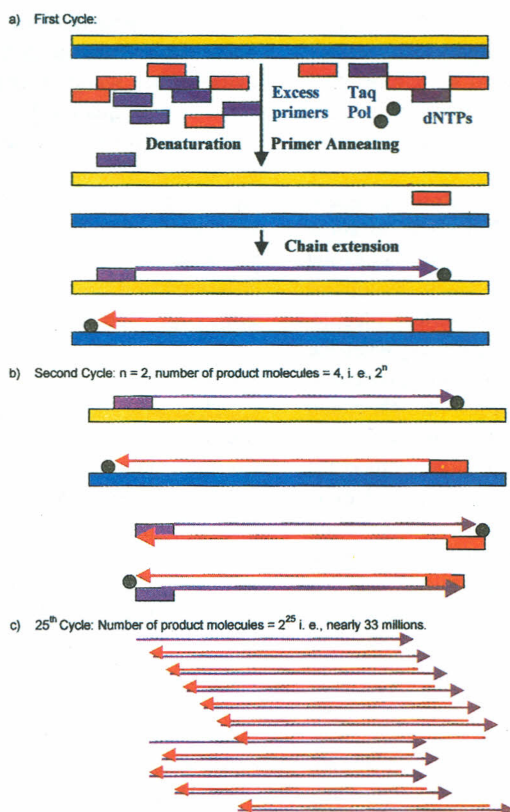

FIG 4. **Polymerase chain reaction (PCR).** *In vitro* enzymatic amplification procedure can yield in excess of a million-fold amplification of the starting target DNA molecule by cyclic repetition of denaturation, primer annealing and chain extension. The excess of primers (P1 and P2) added in the beginning of the reaction are extended as copies of the template DNA molecule under the action of thermostable Taq polymerase enzyme.

mutation. A number of diseases have been subjected to mutation detection via SSCP.[21]

HA detects double-stranded DNA molecules with one or more nucleotide mismatches between the two hybridized strands. Denaturing gradient gel electrophoresis (DGGE) or one of its various modifications are used to detect mutations in clinical samples.[22] During gel electrophoresis, a heteroduplex migrates slower than a homoduplex. Increasing the temperature or denaturant concentration in a resolving gel causes preferential melting at the mismatch region, leading to an abrupt drop in the mobility of the heteroduplex. Complete separation of the two strands is avoided by a GC clamp introduced at one end of the heteroduplex through PCR primers. While SSCP and HA allow mutation detection based on single- or double-stranded DNA conformations, they do not provide any information on the locus of the mutation(s). These diagnostic tests, therefore, often need to be confirmed by direct DNA sequencing.

Enzyme mismatch cleavage (EMC) methods use bacteriophage resolvases (T4 endonuclease VII and T7 endonuclease I) to cleave the site of mismatch in fluorescently tagged PCR products. These then analyse the products by gel resolution to localize the position of cleavage. This method detects deletions more efficiently than point mutations.[23] In contrast, the *chemical mismatch cleavage (CMC)* method uses chemicals (osmium tetroxide and hydroxylamine) that specifically recognize and react with mismatched base(s). Piperidine is used to cleave the modified base, and the resultant products are separated by size to determine the location of mismatched bases. This procedure allows precise detection of mutations.[24]

Dideoxy fingerprinting (ddF) combines dideoxy chain termination with SSCP to detect a mutation.[25] A modification of this method can also use RNA as the starting material.[26] The dideoxynucleotide chain-terminated products are resolved through a non-denaturing gel. The presence or absence of a band with respect to control DNA will locate the position of the polymorphism.

In *glycosylase-mediated polymorphism detection (GMPD)*, limiting amounts of dUTP are incorporated in place of dTTP during PCR amplification of the target DNA. Treatment of the amplified DNA with uracil-N-glycosylase creates an abasic nucleotide at the position of deoxyuridine, which is cleaved with an endonuclease to generate a series of nested fragments.[27] Novel or missing bands at a specific position compared to the control indicate a mutation. Since most of the known human point mutations involve dT, GMPD is probably the most versatile method to detect point mutations.

In the *amplification refractory mutation system (ARMS)*, two PCR reactions are carried out. In the first, a forward primer matching the normal allele is used, whereas the second uses a forward primer matching the mutant sequence, with a mismatch at the 3' end of the primer. The reverse primer is the same in both reactions. With a homozygous mutation, only the second reaction generates a product, while for a heterozygous individual both reactions yield products.[28]

In *allele-specific oligo (ASO) hybridization*, a pair of 15- to 17-mer oligonucleotide probes, one corresponding to the mutant and the other to the normal allele, are hybridized to genomic DNA or cDNA immobilized to a nitrocellulose membrane by dot/slot blotting. The mismatch in the mutant probe is toward the middle. Under high stringency hybridization, a mutant allele will be

discriminated from the normal allele by virtue of the signal.[29] *Reverse ASO hybridization* employs similar oligo probes immobilized to nylon membranes and hybridized to PCR-amplified products of the target sequence.[30] This rather simple and high-throughput method for detecting known mutations is a forerunner of the new DNA microarray technology.

Oligonucleotide ligation assay (OLA) uses the absolute dependence on complementarity with the target DNA so as to allow ligation between adjacently placed oligos. PCR-amplified target DNA is hybridized with an oligo probe labelled with biotin and another mutation-specific oligo with a fluorescent reporter molecule. Ligation reaction follows, where the two probes ligate only when the ligating ends perfectly hybridize to the target DNA. The product of ligation is captured onto a streptavidin-coated microtitre plate, and the detection of fluorescence signals presence of the mutation.[31]

Protein truncation test (PTT) is a highly sensitive method for detecting non-sense frameshift, splice site, insertion, and exon deletion mutations in a gene. In a coupled transcription–translation system, mRNA is translated *in vitro* to detect translation termination or aberrant-sized peptides. A limitation of PTT, however, is that it is unable to detect mis-sense mutations. This technique is utilized for mutation detection in cases of APC (adenomatous polyposis coli), dystrophin, BRCA1 and BRCA2 genes.[32]

Several automated high-throughput methods of mutation detection such as surface plasmon resonance (SPR), mass spectrometry (MS), etc. are evolving.[33,34] The most prominent among these is the use of microarrays.[35] A *microarray* consists of as many as 100 000 sequence-defined 15–20 nucleotide-long oligo sequences ('probes') representing all types of single nucleotide changes, such as substitution, deletion and insertion on a semiconductor chip in a grid pattern. Fluorescent-tagged, PCR-amplified test DNA sequences (called 'targets') are hybridized to immobilized 'probes'. The resulting signals are then compared to the normal profile to determine if mutations are present.[36]

CHROMOSOMAL HYBRIDIZATION ASSAYS TO DETECT CHROMOSOMAL ABERRATIONS

In situ hybridization

In tissue *in situ* hybridization, a suitably labelled complementary RNA (cRNA) probe or riboprobe is hybridized against RNA in tissue sections derived from either paraffin embedded or frozen tissues and intracellular mRNA is localized by bright field microscopy. Non-isotopic labelling has replaced radioactive probes, which in turn has made this technology more amenable to clinical laboratories.[37] The *chromosome in situ hybridization* procedure, on the other hand, uses a suitably labelled DNA probe to hybridize with *in situ* denatured chromosomal DNA to detect the location of the complementary target DNA sequence. Chromosome banding performed either before or after the hybridization allows correlation of the signal with chromosome band pattern. *Fluorescent in situ hybridization (FISH)*[38] is a molecular cytogenetic technique that uses several kilobase-long fluorescently labelled single-stranded probes which are either locus-specific or alphoid (centromeric) repeat DNA probes. While the former type of probes hybridize with and report on specific genetic anomalies such as translocations and amplifications, the latter probes are useful for identification of specific chromosomes. The hybridization signal detection is achieved using an epi-fluorescence microscope. Sophisticated digital image analysis equipment is now routinely used to analyse the results of FISH. *Chromosome painting* is a modification of FISH, which involves using individual chromosome-specific fluorochrome-labelled probes to 'paint' the chromosomes.[39] Chromosome painting can be applied to both metaphase and interphase nuclei and is particularly useful in defining the gene rearrangements in cancer cytogenetics since chromosome preparations from tumours are often of poor quality. Five synthetic fluorescent labels are used in different ratios to label the chromosomes. Variation of ratios allows expansion of the five basic spectral patterns to yield 24 differently fluorescent dyes required for visualizing each chromosome distinctly. Albeit expensive at present, this technique is increasingly being applied to clinical situations to analyse large chromosome deletions and amplifications, and translocations between non-homologous chromosomes.

Comparative genomic hybridization (CGH)

CGH, mostly applied in case of cancers, uses a mixture of matched tumour DNA and normal control DNA labelled with different fluorochromes to hybridize with chromosomes of normal cells in competitive FISH. When both the normal DNA and tumour DNA bind the template normal DNA in equal amounts, a resultant fluorochrome colour pattern is seen. An excess fluorescence of tumour DNA type indicates amplification of that region in the tumour DNA, while predominance of a fluorescence from the normal DNA label indicates loss of the corresponding DNA sequence from the tumour DNA. The 'gain' is indicative of oncogene amplification, whereas the 'loss' suggests deletions of tumour suppression genes, thus revealing all the regions of amplification and losses of genes in a single experiment. The technique of CGH has been able to detect a variety of such gene amplifications and losses in a number of cancer tissues.[40]

MOLECULAR PATHOLOGY OF SOME DISEASES

Infectious diseases

Molecular techniques that have helped to dissect the pathology of infectious diseases are playing an even more important role both in the diagnosis and design of drugs and vaccines to combat infections. The associations of *Helicobacter pylori* with gastroduodenal disorders[41] and cytomegaloviruses and *Chlamydia pneumoniae* with atherogenesis[42] have forced revision of contemporary thinking about the pathogenesis of chronic organ and tissue diseases. Not just oral cancer—which takes a crushing toll in India thanks to tobacco-related habits—but also oral infections such as dental caries and periodontal disease are fast becoming amenable to molecular diagnosis using saliva as a convenient substitute for blood.[43] Similar PCR-based diagnostic assays are available for malaria,[44] pneumonia,[45] hepatitis B and C viral infections,[46] meningitis[47] and sexually transmitted diseases.[48]

Tuberculosis (TB) is a dreaded infectious disease that, in spite of the availability of potent antagonistic drugs, is known to continue to develop into drug-resistant entities in many cases. Early diagnosis of tuberculosis is crucial for early medical intervention and prevention of spread of the bacteria. Molecular diagnostic techniques such as PCR have been shown to be very useful for an effective early detection. The high-sensitivity PCR test can shorten diagnostic duration from several weeks to one or two days.[49] In fact, PCR can offer a definitive diagnosis even in cases where acid-fast bacilli (AFB) are not demonstrable in granulomatous inflammations although the clinical symptoms are suggestive of tuberculosis. It has been shown, for example, that *M. tuberculosis* DNA can be extracted even from paraffin sections and amplified by PCR with the IS6110 primers specific for the Mtb complex and the specific PCR product detected by agarose gel electrophoresis.[50]

Acquired immunodeficiency syndrome (AIDS) has rapidly

become pandemic in India and its toll is rightly being predicted to be devastating.[51] The study of the molecular pathogenesis of AIDS and influenza has delivered a strong boost to drug design strategies revolving around viral structure. Examples of such successful development of drugs are HIV protease inhibitors[52] and influenza neuraminidase inhibitor.[53] Equally important will be the characterization of the human genome through the HGP and elucidation of those single nucleotide polymorphisms (SNPs) which are expected to yield pointers to susceptibility to infectious diseases in some individuals relative to others or their response to a drug.[54]

Genetic diseases

In India haemoglobinopathies such as *thalassaemia* and *haemophilia* are important public health problems. It is estimated that these inherited disorders will affect about 12 000 infants born every year,[55] imposing upon the patients and their families staggering costs by way of frequent blood transfusions and physical and psychological trauma. A preventive genetic approach including carrier detection and prenatal molecular diagnosis coupled with genetic counselling are emerging as important options for couples at high risk for haemoglobinopathies. Molecular diagnostic methods are available for haemophilia,[56] alpha thalassaemia[57] and beta thalassaemia.[58]

In recent years, the molecular basis of neurodegenerative diseases is being elucidated.[59] At least eight neurodegenerative disorders are known to be caused by unstable expansions of the trinucleotide CAG within the related genes,[11,60] Huntington's disease (HD) being the commonest example. Here, the number of CAG repeats in the HD gene of the affected individual jumps to an average of 80 from the normal average of about 27 repeats. Other similar genetic diseases are Fragile X syndrome (CCG repeat expansion in FRAXA or FRAXE genes) and myotonic dystrophy (CTG repeat expansion in DM gene). The unusual clinical genetic properties of these diseases are related to the dynamic nature of CAG repeat expansions, including instability of the repeat expansion in meiosis, particularly meiosis in the male. Thus, there is a strong correlation between the size of the repeat expansion and age of onset of disease; higher the rate of repeat expansion in succeeding generations, the earlier is the onset of disease. This phenomenon is called 'anticipation'. Simple non-isotopic PCR assays with a potential for diagnostic use have been developed that can identify such repeat expansion.[11]

Cancer

Multiple genomic alterations involving oncogenes, tumour suppressor genes and DNA mismatch repair have been characterized in cancers. For example, the K-*ras* gene is somatically mutated in codons 12, 13 and 61 in several sporadic human cancers, including colorectal, pancreatic and lung cancers. Clinical samples used for molecular diagnosis can be stool and lavage fluid, pancreatic and duodenal juices, and sputum and lavage fluids. The diagnostic methods used include Southern hybridization, PCR, or a combination of PCR and RFLP or SSCP.[62]

As noted above, chromosomal rearrangements such as translocations form the hallmark of many cancers, which have been dissected for their molecular basis. These alterations can be identified at the molecular level using techniques such as FISH or PCR. An example of the latter is the development of single-step multiplex nested RT-PCR assays for sensitive screening of patients with myeloproliferative conditions such as chronic myeloid leukaemia (CML) and acute myeloid leukaemia (AML) for the presence of underlying BCR–ABL fusion transcripts that are characteristic of many patients with these diseases.[63] A similar diagnostic test for chimeric PAX3-FKHR transcript resulting from the translocation t(2;13)(q35,q14) has recently become available for childhood rhabdomyosarcoma.[64]

RT-PCR is also being increasingly applied to the identification of occult metastatic cancer cells in peripheral blood, bone marrow and lymph nodes, which has resulted in a significant increase in sensitivity and reduction in tedium associated with histopathological examination.[65] Attempts are being made to standardize the protocols, reagents and choice of molecular targets across clinical laboratories, which should yield the development of globally uniform assays.[66]

In summation, the continuing evolution of the molecular biology of human diseases is having profound effects throughout the field of medicine. Clinical translation of the molecular data and technologies has facilitated development of assays for DNA or RNA sequences to diagnose infectious, malignant and genetic diseases. Many of the assays have been used clinically, and form an integral part of patient care under advanced medicine by effecting cost savings through earlier and more accurate diagnosis. Thus, molecular diagnostic tests can contribute to efficient disease management.

Fiction writer Don DeLillo says in *Underworld*: 'You didn't see because you don't know how to look. And you don't know how to look, because you don't know the names.' The exciting research findings in molecular biology of diseases have helped us 'know the names', which, in turn, is letting us 'see'.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation for the critical review of the manuscript by Dr Ben George, USC, Los Angeles.

REFERENCES

1 Felsenfeld G, Boyes J, Chung J, Clark D, Studitsky V. Chromatin structure and gene expression. *Proc Natl Acad Sci USA* 1996;**93**:9384–8.
2 Marshall E. Rival genome sequencers celebrate a milestone together. *Science* 2000;**288**:2294–5.
3 Pennisi E. Finally, the book of life and instructions for navigating it. *Science* 2000;**288**:2304–7.
4 Fry CJ, Farnham PJ. Context-dependent transcriptional regulation. *J Biol Chem* 1999;**274**:29583–6.
5 Ikuta T, Papayannopoulou T, Stamatoyannopoulos G, Kan YW. Globin gene switching. *In vivo* protein-DNA interactions of the human beta-globin locus in erythroid cells expressing the fetal or the adult globin gene program. *J Biol Chem* 1996;**271**: 14082–91.
6 John M, Leppik R, Busch SJ, Granger-Schnarr M, Schnarr M. DNA binding of Jun and Fos bZip domains: Homodimers and heterodimers induce a DNA conformational change in solution. *Nucleic Acids Res* 1996;**24**:4487–94.
7 Chen L, Glover JN, Hogan PG, Rao A, Harrison SC. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* 1998;**392**: 42–8.
8 Lania L, Majello B, Napolitano G. Transcriptional control by cell-cycle regulators: A review. *J Cell Physiol* 1999;**179**:134–41.
9 Ravi D, Panikkar KR, Krishnan Nair M, Radhakrishnan Pillai M. Apoptosis: Future directions in cancer therapy. *Natl Med J India* 2000;**13**:71–8.
10 Wang Q, Lasset C, Desseigne F, Saurin JC, Maugard C, Navarro C, *et al*. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. *Hum Genet* 1999;**105**:79–85.
11 Vuillaume I, Schraen S, Rousseaux J, Sablonniere B. Simple nonisotopic assays for detection of (CAG)n repeats expansions associated with seven neurodegenerative disorders. *Diagn Mol Pathol* 1998;**7**:174–9.
12 Tucker JD, Preston RJ. Chromosome aberrations, micronuclei, aneuploidy, sister chromatid exchanges, and cancer risk assessment. *Mutat Res* 1996;**365**:147–59.
13 Sanchez-Garcia I. Consequences of chromosomal abnormalities in tumor development. *Ann Review Genet* 1997;**31**:429–53.
14 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;**98**:503–17.
15 Mullis KB, Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* 1987;**155**:335–50.
16 Raj GV, Moreno JG, Gomella LG. Utilization of polymerase chain reaction technology in the detection of solid tumors. *Cancer* 1998;**82**:1419–42.
17 Thaker V. *In situ* RT-PCR and hybridization techniques. *Methods Mol Biol* 1999;**115**: 379–402.

18 Lie YS, Petropoulos CJ. Advances in quantitative PCR technology: 5' nuclease assays. *Curr Opin Biotechnol* 1998;**9:**43–8.

19 Tan W, Fang X, Li J, Liu X. Molecular beacons: A novel DNA probe for nucleic acid and protein studies. *Chemistry-A Eur J* 2000;**6:**1107–11.

20 Kessler HH, Muhlbauer G, Rinner B, Stelzl E, Berger A, Dorr HW, *et al.* Detection of *Herpes simplex* virus DNA by real-time PCR. *J Clin Microbiol* 2000;**38:**2638–42.

21 Jaeckel S, Epplen JT, Kauth M, Miterski B, Tschentscher F, Epplen C. Polymerase chain reaction-single strand conformation polymorphism or how to detect reliably and efficiently each sequence variation in many samples and many genes. *Electrophoresis* 1998;**19:**3055–61

22 Fodde R, Losekoot M. Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Hum Mutat* 1994;**3:**83–94.

23 Deeble VJ, Roberts E, Robinson MD, Woods CG, Bishop DT, Taylor GR. Comparison of enzyme mismatch cleavage and chemical cleavage of mismatch on a defined set of heteroduplexes. *Genetic Testing* 1997–98;**1:**253–9.

24 Ellis TP, Humphrey KE, Smith MJ, Cotton RG. Chemical cleavage of mismatch: A new look at an established method. *Hum Mutat* 1998;**11:**345–53.

25 Shevchenko YO, Bale SJ, Compton JG. Mutation screening using automated bidirectional dideoxy fingerprinting. *Biotechniques* 2000;**28:**134–8.

26 Martincic D, Koury MJ, Gale K, Whitlock JA. Detection of mutations by automated fluorescence/RNA-based dideoxy fingerprinting (ARddF). *Oncogene* 1999;**18:** 617–21.

27 Vaughan P, McCarthy TV. A novel process for mutation detection using uracil DNA-glycosylase. *Nucleic Acids Res* 1998;**26:**810–15.

28 Tonks S, Marsh SG, Bunce M, Bodmer JG, Molecular typing for HLA class I using ARMS-PCR: Further developments following the 12th International Histocompatibility Workshop. *Tissue Antigens* 1999;**53:**175–83.

29 Shuber AP, Michalowsky LA, Nass GS, Skoletsky J, Hire LM, Kotsopoulos SK, *et al.* High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. *Human Mol Genet* 1997;**6:**337–47.

30 Lau J, Tolan DR. Screening for hereditary fructose intolerance mutations by reverse dot-blot. *Mol Cell Probes* 1999;**13:**35–40

31 Shi MM, Bleavins MR, de la Iglesia FA. Technologies for detecting genetic polymorphisms in pharmacogenomics. *Mol Diagnosis* 1999;**4:**343–51.

32 Den Dunnen JT, Van Ommen GJ. The protein truncation test: A review. *Human Mut* 1999;**14:**95–102.

33 Pearson JE, Gill A, Vadgama P. Analytical aspects of biosensors. *Ann Clin Biochem* 2000;**37:**119–45.

34 Garvin AM, Parker KC, Haff L. MALDI-TOF based mutation detection using tagged *in vitro* synthesized peptides. *Nature Biotechnol* 2000;**18:**95–7.

35 Cheng J, Fortina P, Surrey S, Kricka LJ, Wilding P. Microchip-based devices for molecular diagnosis of genetic diseases. *Mol Diagn* 1996;**1:**183–200.

36 Ramsay G. DNA chips: State-of-the art. *Nature Biotechnol* 1998;**16:**40–4.

37 Breier G. *In situ* hybridization with RNA probes. *Methods Mol Biol* 1999;**96:**107–17.

38 Raap AK. Advances in fluorescence *in situ* hybridization. *Mutation Res* 1998;**400:** 287–98.

39 Ried T, Schrock E, Ning Y, Wienberg J. Chromosome painting: A useful art. *Hum Mol Genet* 1998;**7:**1619–26.

40 Rooney PH, Murray GI, Stevenson DA, Haites NE, Cassidy J, McLeod HL. Comparative genomic hybridization and chromosomal instability in solid tumours. *Br J Cancer* 1999;**80:**862–73.

41 Kimmel B, Bosserhoff A, Frank R, Gross R, Goebel W, Beier D. Identification of immunodominant antigens from *Helicobacter pylori* and evaluation of their reactivities with sera from patients with different gastroduodenal pathologies. *Infect Immun* 2000;**68:**915–20

42 Cheng JW, Rivera NG. Infection and atherosclerosis—focus on cytomegalovirus and *Chlamydia pneumoniae*. *Ann Pharmacother* 1998;**32:**1310–16.

43 Murdoch-Kinch CA. Oral medicine: Advances in diagnostic procedures. *J California Dental Assoc* 1999;**27:**773–80,782–4.

44 Ciceron L, Jaureguiberry G, Gay F, Danis M. Development of a *Plasmodium* PCR for monitoring efficacy of antimalarial treatment. *J Clin Microbiol* 1999;**37:**35–8.

45 Khan MA, Farrag N, Butcher P. Diagnosis of *Pneumocystis carinii* pneumonia: Immunofluorescence staining, simple PCR or nPCR. *J Infect* 1999;**39:**77–80.

46 Roth WK, Weber M, Seifried E. Feasibility and efficacy of routine PCR screening of blood donations for hepatitis C virus, hepatitis B virus, and HIV-1 in a blood bank setting. *Lancet* 1999;**353:**359–63.

47 Hosoya M, Honzumi K, Sato M, Katayose M, Kato K, Suzuki H. Application of PCR for various neurotropic viruses in the diagnosis of viral meningitis. *J Clin Virol* 1998;**11:**117–24.

48 van Der Schee C, van Belkum A, Zwijgers L, van Der Brugge E, O'Neill EL, Luijendijk A, *et al.* Improved diagnosis of *Trichomonas vaginalis* infection by PCR using vaginal swabs and urine specimens compared to diagnosis by wet mount microscopy, culture, and fluorescent staining. *J Clin Microbiol* 1999;**37:** 4127–30.

49 Caldarelli-Stefano R, Vago L, Bonetto S, Nebuloni M, Costanzi G. Use of magnetic beads for tissue DNA extraction and IS6110 *Mycobacterium tuberculosis* PCR. *Mol Pathol* 1999;**52:**158–60.

50 Li JY, Lo ST, Ng CS. Molecular detection of *Mycobacterium tuberculosis* in tissues showing granulomatous inflammation without demonstrable acid-fast bacilli. *Diagn Mol Pathol* 2000;**9:**67–74.

51 Hira SK, Srinivasa Rao AS, Thanekar J. Evidence of AIDS-related mortality in Mumbai, India. *Lancet* 1999;**354:**1175–6.

52 Ren S, Lien EJ. Development of HIV protease inhibitors: A survey. *Prog Drug Res* 1998;**51:**1–31.

53 Hayden FG, Atmar RL, Schilling M, Johnson C, Poretz D, Paar D, *et al.* Use of the selective oral neuraminidase inhibitor oseltamivir to prevent influenza. *N Engl J Med* 1999;**341:**1336–43.

54 *http://www.ornl.gov/hgmis/faq/snps.html*

55 Balgir RS. Control and prevention of the genetic load of haemoglobinopathies in India. *Natl Med J India* 1999;**12:**234–8.

56 Montejo JM, Magallon M, Tizzano E, Solera J. Identification of twenty-one new mutations in the Factor IX gene by SSCP analysis. *Hum Mutat* 1999;**13:**160–5.

57 Chong SS, Boehm CD, Higgs DR, Cutting GR. Single-tube multiplex-PCR screen for common deletional determinants of alpha-thalassemia. *Blood* 2000;**95:**360–2.

58 Bravo M, Salazar R, Arends A, Alvarez M, Velazquez D, Guevara JM, *et al.* Detection of beta thalassemia by the technique of refractory amplification of mutation systems (ARMS-PCR). *Invest Clin* 1999;**40:**203–13.

59 Martin JB. Molecular basis of the neurodegenerative disorders. *N Engl J Med* 1999; **340:**1970–80.

60 Nance MA. Clinical aspects of CAG repeat diseases. *Brain Pathol* 1997;**7:**881–900.

61 Pfeilschifter J, Muhl H. Immunopharmacology: Anti-inflammatory therapy targeting transcription factors. *Eur J Pharmacol* 1999;**375:**237–45

62 Minamoto T, Mai M, Ronai Z. K-*ras* mutation: Early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers—a review. *Cancer Detect Prev* 2000;**24:**1–12.

63 Otazu IB, Zalcberg I, Tabak DG, Dobbin J, Seuanez HN. Detection of BCR-ABL transcripts by multiplex and nested PCR in different haematological disorders. *Leuk Lymphoma* 2000;**37:**205–11.

64 Tobar A, Avigad S, Zoldan M, Mor C, Goshen Y, Zaizov R. Clinical relevance of molecular diagnosis in childhood rhabdomyosarcoma. *Diagn Mol Path* 2000;**9:** 9–13.

65 Bustin SA, Dorudi S. Molecular assessment of tumour stage and disease recurrence using PCR-based assays. *Mol Med Today* 1998;**4:**389–96.

66 van Dongen JJ, Macintyre EA, Gabert JA, Delabesse E, Rossi V, Saglio G, *et al.* A standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: Investigation of minimal residual disease in acute leukemia. *Leukemia* 1999;**13:**1901–28.