

Primer on Medical Genomics

Part II: Background Principles and Methods in Molecular Genetics

AYALEW TEFFERI, MD; ERIC D. WIEBEN, PhD; GORDON W. DEWALD, PhD; DAVID A. H. WHITEMAN, MD; MATTHEW E. BERNARD, MD; AND THOMAS C. SPELSBERG, PhD

The nucleus of every human cell contains the full complement of the human genome, which consists of approximately 30,000 to 70,000 named and unnamed genes and many intergenic DNA sequences. The double-helical DNA molecule in a human cell, associated with special proteins, is highly compacted into 22 pairs of autosomal chromosomes and an additional pair of sex chromosomes. The entire cellular DNA consists of approximately 3 billion base pairs, of which only 1% is thought to encode a functional protein or a polypeptide. Genetic information is expressed and regulated through a complex system of DNA transcription, RNA processing, RNA translation, and posttranslational and cotranslational modification of proteins. Advances in molecular biology techniques have allowed accurate and rapid characterization of DNA sequences as well as identification and quantification of cellular RNA and protein. Global analytic methods and human genetic mapping are expected to accelerate the process of identification and localization of disease genes.

In this second part of an educational series in medical genomics, selected principles and methods in molecular biology are recapped, with the intent to prepare the reader for forthcoming articles with a more direct focus on aspects of the subject matter.

Mayo Clin Proc. 2002;77:785-808

BAC = bacterial artificial chromosome; bp = base pairs; cDNA = complementary DNA; cM = centiMorgan; EST = expressed sequence tag; FISH = fluorescence in situ hybridization; FRET = fluorescence resonance energy transfer; kb = kilobase; Mb = million base; mRNA = messenger RNA; mtDNA = mitochondrial DNA; ORF = open reading frame; PCR = polymerase chain reaction; RFLP = restriction fragment length polymorphism; rRNA = ribosomal RNA; RT-PCR = reverse transcription-PCR; snRNP = small nuclear ribonucleoprotein; tRNA = transfer RNA; UTR = untranslated region; VNTR = variable numbers of tandem repeats; YAC = yeast artificial chromosome

The Human Genome Project has provided near-total sequence information about our genes and their surroundings.^{1,2} This monumental accomplishment has been accompanied by equally remarkable progress in biomedical research and computer science.³⁻⁵ Biologists and clinical investigators are now well positioned to perform global analysis of genetic polymorphisms-mutations,^{6,7} gene expression,⁸ protein identity and function (proteomics),⁹ and the 3-dimensional structure of proteins (structural genomics).^{10,11} The resulting information should accelerate the development of novel and targeted therapeutics that are tailored to individual genomic profiles (pharmacogenomics).^{12,13} The massive amount of data generated by genome-wide experiments

requires the development of comprehensive genomic-proteomic databases and software that enable megadata access, transfer, and manipulation (bioinformatics).^{14,15}

For editorial comment, see page 745.

The development of a structural and functional genomic infrastructure will undoubtedly accelerate our understanding of diseases and help refine methods of disease prediction, diagnosis, staging, and treatment.^{16,17} Clinicians, clinician investigators, and students of medicine must be prepared to participate effectively in the “new” genomics revolution and tackle the economical and ethical issues that are integral to all aspects of medical genomics.¹⁸⁻²¹ The intent of the current communication is to begin the process of education in this regard, with a clinician-oriented review of basic principles and methods in molecular biology that are relevant to genomics. In subsequent articles, the aforementioned components of genomics will be discussed and updated periodically.

GENES AND GENE EXPRESSION Genes and DNA

A gene, the original Mendel “factor,”²² is the basic unit of heredity.^{23,24} A gene is defined as a contiguous region of

From the Division of Hematology and Internal Medicine (A.T.), Department of Biochemistry and Molecular Biology (E.D.W., T.C.S.), Division of Laboratory Genetics (G.W.D.), Department of Medical Genetics (D.A.H.W.), and Department of Family Medicine (M.E.B.), Mayo Clinic, Rochester, Minn. The authors are members of the Mayo Clinic Genomics Education Steering Committee.

A glossary of terms frequently used in medical genomics appears on pages 805 through 808.

Individual reprints of this article are not available. The entire Primer on Medical Genomics will be available for purchase from the *Proceedings* Editorial Office at a later date.

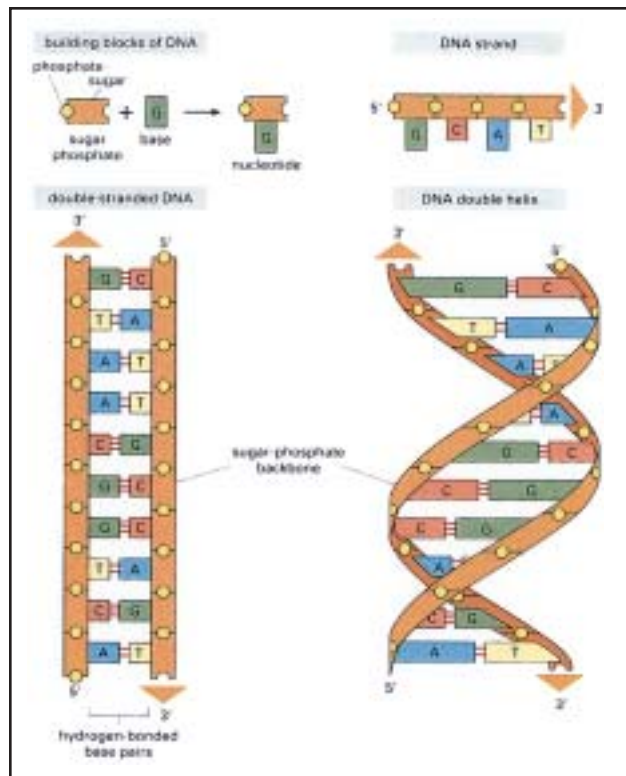


Figure 1. Components and structure of DNA. A = adenine; G = guanine; T = thymine; C = cytosine. See text for further details. Copyright 1998 from Alberts et al.^{30(p186)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

DNA^{25,26} that includes a defined set of exons (protein-coding regions of DNA)²⁷ and introns (DNA regions interspersed between exons).²⁸ The basic unit of DNA is called a *deoxyribonucleotide* (nucleotide for short) and is made up of 1 of 4 nitrogenous bases (adenine [A], guanine [G], thymine [T], cytosine [C]) that is attached to the 1' carbon of a deoxyribose sugar, which in turn is attached to a phosphate group through its 3' and 5' carbons²⁹ (Figure 1). Each DNA strand is a polymer of many nucleotides that are attached to each other through their phosphate and sugar groups (forming a sugar-phosphate backbone of DNA).³¹ There are 2 orientations of sugar-phosphate connections: 5' to 3' (carbon 5' of the sugar moiety of a particular nucleotide in a strand is attached to a phosphate group that is in turn attached to carbon 3' of the next nucleotide) and 3' to 5'.

A DNA molecule is made up of 2 antiparallel (1 strand is oriented 5' to 3' and the other 3' to 5') polymers (strands) of nucleotides (side-by-side chains) in a helical configuration that forms a double helix (Figure 1).³² The 2 strands of DNA are noncovalently (loosely) connected to each other through their nitrogenous bases by hydrogen bonds (A

always bonds with T and G with C). Both strands of DNA are used as templates for new complementary DNA (cDNA) synthesis during DNA replication that precedes cell division. As such, each of the daughter cells contains 1 parental and 1 newly synthesized strand of DNA (ie, DNA replication is semiconservative).³³

Although the accurate number of genes in the human genome remains to be clarified,³⁴ the initial drafts from the Human Genome Project suggest the presence of approximately 30,000 named and unnamed genes and many intergenic DNA sequences.^{1,2} In other words, the human genome represents the entire DNA sequence (approximately 3 billion base pairs [bp] per haploid set of chromosomes) in a human cell and is made up of approximately 1% exons (protein-coding regions of DNA), 24% introns (DNA regions interspersed between exons), and 75% intergenic DNA (DNA regions that flank genes).³⁵ The intergenic DNA consists largely (55% of the total genome) of repetitive sequences, including inverted repeats.³⁶⁻³⁸

In this regard, it is helpful to recall the different categories of DNA that make up the human genome. The most common class of DNA, also the functionally relevant one, is called *single-copy DNA* and appears only once or a few times across the entire genome. In contrast, repetitive DNA sequences appear multiple times and may be found either in a series of adjacent clusters (tandem repeats)³⁹ or scattered at different regions of the genome (interspersed repeats).^{40,41} Tandem repeats make up "satellite" DNA, a band of DNA that separates from other DNA on a cesium chloride density gradient centrifugation.^{42,43} Satellite DNA is usually localized near the centromere.⁴⁴ The terms *minisatellites* (20-70 bp repeats) and *microsatellites* (2-4 bp repeats) are used to refer to tandem repeats of small size.⁴⁵ The total length of such tandem repeats is highly variable among individuals and between cells of different clonal origin. Accordingly, this variability has been used in different investigations, including person identification and cell chimerism.⁴⁶

Scattered DNA sequence repeats can also be short (short interspersed nuclear elements [SINEs], eg, Alu repeats of 100-300 bp long) or long (long interspersed nuclear elements [LINEs], eg, L1 [line-1; long inserted element-1] sequences of 1-5 kilobase [kb] long).⁴¹ An Alu element is a specific DNA sequence that derives its name from the Alu 1 endonuclease because of a single recognition site for the particular enzyme located near the center of the Alu sequence. Approximately one half million Alu repeats are scattered throughout human genome and serve no apparent function. Some DNA sequences can self-replicate, and the additional copies might migrate to different locations within the genome (mobile elements or transposons).⁴⁷ Some mobile elements may have originated from viral

RNA and propagate via an RNA intermediate by using reverse transcriptase (retrotransposons).^{48,49}

Most human DNA is organized into 46 chromosomes that reside in the nucleus of every nucleated cell. However, some DNA resides in mitochondria that are located in the cytoplasm and number in the hundreds per individual cell.⁵⁰ Each mitochondrion has many copies of a circular DNA molecule (the mitochondrial chromosome), which is approximately 16,500 bp long, and is replicated and expressed within the organelle.⁵¹ The mitochondrial DNA (mtDNA) may have coevolved with nuclear DNA and encodes for 2 ribosomal RNAs (rRNAs), 22 transfer RNAs (tRNAs), and 13 functional polypeptides that are involved in oxidative phosphorylation.⁵² Mitochondrial DNA differs from nuclear DNA in terms of not having introns and displaying high mutation rates.^{53,54} Because mitochondria are located in the cytoplasm, inheritance of mtDNA is exclusively maternal.⁵⁵ The combination of uniparental inheritance and high mutation rates has made mtDNA important in the study of human evolutionary genetics.⁵⁶⁻⁵⁸

Each chromosome represents 1 double-stranded DNA molecule that is closely packed with special proteins called histones (DNA + histones + regulatory proteins = chromatin) (Figure 2).^{59,60} Chromatin condenses substantially during the metaphase stage of mitosis and becomes visible by light microscopy. Chromatin that remains condensed during interphase is called *heterochromatin*.⁶¹⁻⁶³ Chromatin that condenses only during metaphase is called *euchromatin*. In general, euchromatin contains more functional genes, and heterochromatin contains more nonfunctional repetitive DNA elements.

In women, the 46 chromosomes in each cell are found in 23 homologous pairs (2 haploid sets of chromosomes): 1 set inherited from the father and the other from the mother. The same is true for men except the twenty-third chromosome pair is only partially homologous, at the point where it pairs an X to a Y chromosome.⁶⁴ The smallest and largest human chromosomes contain about 50 and 250 million bp, respectively. Each human gene is usually more than 10,000 bp long.¹ However, only a tenth of this actually codes for amino acids of the mature protein product.

Gene Expression

The information contained in a given gene is expressed if the DNA sequence of the gene is first transcribed into an RNA molecule that, after being processed in the nucleus, will carry the information into the cytoplasm for translation into a protein.⁶⁵ Unless indicated otherwise, the mechanisms described subsequently pertain primarily to eukaryotic cells.

Transcription.—Transcription starts with formation of an RNA sequence from a DNA template representing a

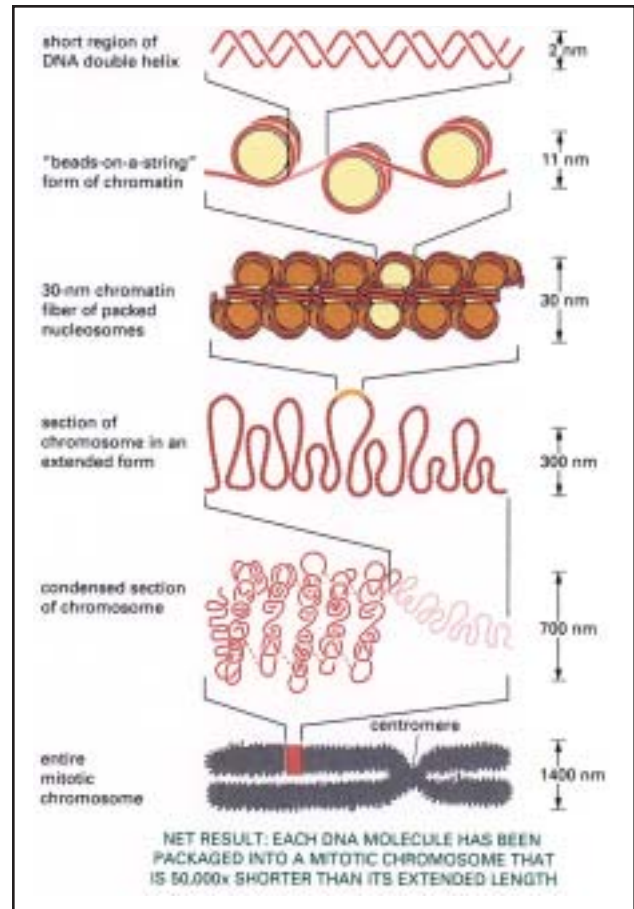


Figure 2. Packing of the DNA molecule with histone proteins to form chromatin packages that are condensed into chromosomes. Copyright 1998 from Alberts et al.^{30(p253)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

gene. The protein-coding gene has both transcribed and nontranscribed regions of DNA.⁶⁶ The transcribed region is first copied onto a nuclear RNA called the *primary transcript* (pre-messenger RNA [mRNA] or heterogeneous nuclear RNA) in a process catalyzed by the enzyme RNA polymerase II.⁶⁷ Part of the nontranscribed region (the promoter sequence) of the gene is the site of the initial contact with RNA polymerase II and controls transcription of the primary transcript; it is usually located upstream (20-100 bp) of the start of the transcribed region.^{68,69} Other regulatory elements of transcription (enhancers-silencers) can be on either side of the transcription start site and be several thousand nucleotides away from the place where transcription actually initiates.⁷⁰ However, both promoters and enhancers-silencers usually reside on the same stretch of DNA (ie, are *cis-acting*).

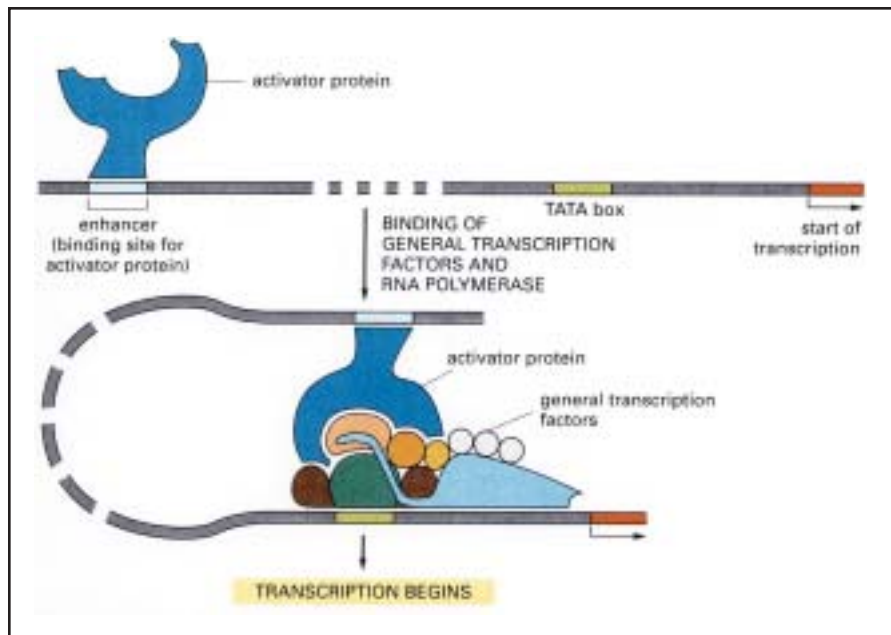


Figure 3. Transcription initiation complex in a eukaryote. See text for details. Copyright 1998 from Alberts et al.^{30(p266)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

Of note, neither the *cis*-acting sequences of DNA nor the RNA polymerase II alone could initiate transcription without the help of proteins called *transcription factors*.⁷¹⁻⁷³ These “*trans*-acting” proteins recognize and bind to the *cis*-acting DNA sequences and together with RNA polymerase II and other coactivators of transcription factors form the transcription complex that ultimately regulates transcription either in a positive or a negative manner (Figure 3).^{74,75} In this regard, promoters often contain characteristic sequences such as TATA and CATT that are recognized by transcription factors.⁷⁶⁻⁷⁹ Transcription factors may be grouped into different classes based on motifs present in their 3-dimensional protein structure (eg, helix-turn-helix, helix-loop-helix, Leucine zipper, zinc finger).^{80,81} These unique motifs allow recognition of the aforementioned specific DNA sequences that are *cis*-acting regulators of transcription.⁸²

During transcription and after completion, the primary transcript (always synthesized from 5′ to 3′ orientation) undergoes several modifications before the mature mRNA is released into the cytoplasm for protein translation.⁸³ These include “capping,” addition of a “poly-A tail,” and “splicing” of the primary transcript.⁸⁴ First, an extra nucleotide (the cap), a 7-methylguanosine (m⁷G-P-P-P) is placed at the 5′ end soon after the initiation of transcription.⁸⁵ This modification is believed to protect RNA degradation and mark the starting point for protein translation by binding to

ribosomes. Subsequently, the end of transcription is heralded when RNA polymerase II reaches a “termination sequence” (TTATT on the DNA template and AAUAAA on the primary transcript), which, together with other proteins, signals cleavage at the polyadenylation site (10-35 nucleotides downstream of the termination sequence).^{86,87} Finally, poly(A) polymerase adds a stretch of approximately 200 adenine nucleotides (poly-A tail) at the 3′ end of the primary transcript.⁸⁸

Once synthesis of the primary transcript is completed, the introns of pre-mRNA are spliced out and the remaining segments rejoined to form the mature mRNA.^{89,90} The latter includes both translated and untranslated regions (UTRs) of exons, the cap, and the poly-A tail (Figure 4).⁸⁶ Splicing occurs in a large complex called a spliceosome, which contains the primary transcript and its associated proteins, several protein splicing factors, and a class of small nuclear RNAs that are present in the cell as small nuclear ribonucleoproteins (snRNPs).⁹¹⁻⁹⁵ There is increasing evidence that both snRNPs and intronic RNA may possess catalytic activity (ie, are ribozymes) and contribute to RNA splicing.⁹⁶⁻⁹⁸ The spliceosome recognizes the 5′ and 3′ splice sites on the exon-intron boundaries via conserved consensus sequences within the intron.⁹⁹ The actual process involves both lariat (loop and tail) formation of the excised intron and transesterification of the 5′ and 3′ splice sites of the exons.¹⁰⁰⁻¹⁰² Alternative splicing occurs not infre-

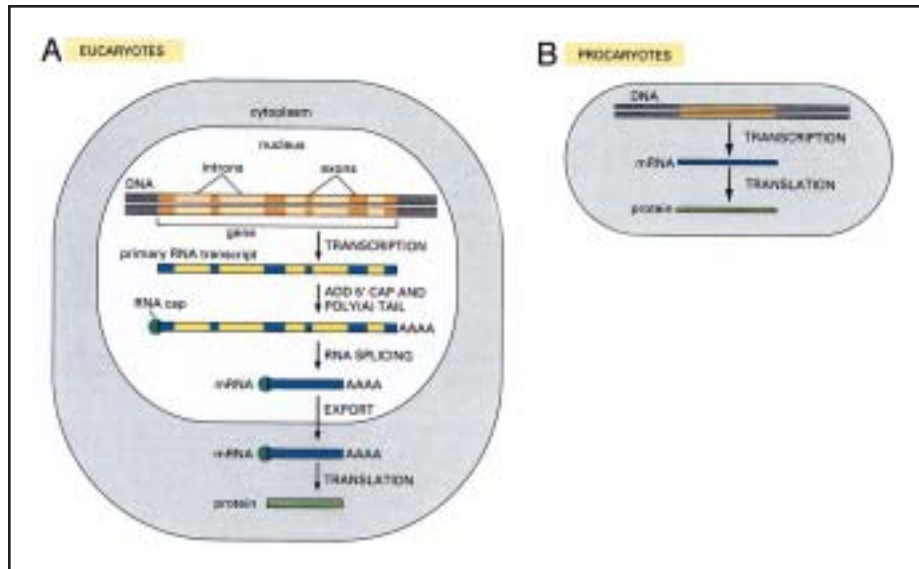


Figure 4. Steps leading from gene to protein for eucaryotes (A) and procaryotes (B). mRNA = messenger RNA. See text for further details. Copyright 1998 from Alberts et al.^{30(p223)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

quently, resulting in different mRNA species and subsequently different proteins from the same primary transcript.¹⁰³⁻¹⁰⁶ Alternative splicing may occur in as many as one half or more of all human genes and sometimes results in altered translation efficiency rather than synthesis of a different polypeptide. The widespread occurrence of this phenomenon increases the coding capacity of genes and contributes to proteomic diversity. The regulation of alternative splicing is an active area of current research, and splicing regulators may include snRNPs.^{104,107-109}

In conventional terms, the final mRNA represents the *sense template*, whereas the cDNA strand that was used as a template for its synthesis is referred to as the *antisense strand*. Regardless, not all the nucleotide sequences of mRNA are ultimately translated into proteins.¹¹⁰ The very first and last transcribed regions of the primary transcript contain sequences (UTRs) that do not code for protein despite being part of the first and last exons.¹¹¹⁻¹¹³ Nonetheless, these sequences are important for the function of the mRNA and often control the stability, subcellular localization, and translational efficiency of an individual mRNA.¹¹⁴⁻¹¹⁷ The 5' UTR, which might be hundreds of bases long, is followed by the protein coding sequence of the mRNA (open reading frame [ORF]) that is in turn followed by the 3' UTR.¹¹⁸

Translation.—The mRNA is translated into protein in the cytoplasm of the cell with the help of ribosomes. Ribosomes are made up of 2 subunits (60S and 40S), each consisting of proteins and rRNA.¹¹⁹ The mature mRNA that

is released into the cytoplasm docks onto rRNA, and its ORF is recognized by another form of cytoplasmic RNA called *transfer RNA*¹²⁰ (Figure 5). A cloverleaf-shaped adaptor molecule, tRNA carries a specific amino acid at its 3' end.¹²² At the opposite end of the amino acid-binding

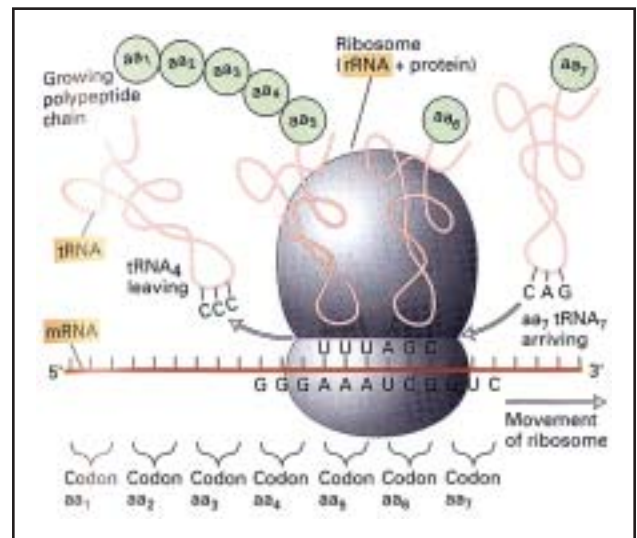


Figure 5. Translation of messenger RNA (mRNA) into amino acids in the cytoplasm of a cell. Process requires the concerted effort of mRNA, transfer RNA (tRNA), and ribosomal RNA (rRNA). See text for further details. Reproduced with permission from Lodish et al.^{121(p117)}

site, the tRNA is equipped with a 3 nucleotide sequence (ie, the “anticodon”) that undergoes complementary base pairing with the mRNA “codon.”^{123,124} This codon-anticodon interaction, which is dictated by the specific nucleotide sequence on the mRNA (the genetic code), aligns anticodon-specific amino acids that subsequently form a growing polypeptide chain through covalent peptide bonding.¹²⁵

Polypeptide synthesis on the ribosome may be subdivided into initiation, elongation, and termination phases. First, elongation factors (eg, eIF4F) recognize the 5′ cap of mRNA, recruit other components of the initiation complex, and subsequently bind to the rRNA on the small (40S) ribosomal subunit.¹²⁶ The initiation complex (40S ribosome, elongation factors such as eIF4, initiator tRNA, etc) rolls along the mRNA sequence until it reaches the “start codon” (AUG). Subsequently, polypeptide synthesis starts with disintegration of the initiation complex and recruitment of the 60S ribosome along with other elongation factors.¹²⁷ As the ribosome continues moving along the mRNA genetic code, successive amino acid–attached tRNAs are recruited, and the polypeptide chain grows until a “stop codon” (UAA, UAG, or UGA) signals release of the finished product.^{128,129} Current evidence suggests that the enzymatic activity required to form peptide bonds between adjacent incoming amino acids may reside in the rRNA, which functions as a ribozyme.¹³⁰

Many newly synthesized proteins are further modified after or during polypeptide synthesis.¹³¹ Posttranslational or cotranslational modifications of proteins ensure proper protein folding, targeting, activation, and stability.¹³² The specific processes include proteolysis, glycosylation, phosphorylation, acylation, metal binding, and assumption of secondary and tertiary structures.¹³³ Chaperones are enzymes that guide proper protein folding by binding and releasing misfolded proteins.¹³⁴ Damaged or surplus proteins are removed through various mechanisms, including lysosomal and nonlysosomal protein degradation. Nonlysosomal protein degradation is facilitated by ubiquitination (a process of tagging proteins with ubiquitin to remove them by proteolysis). Ubiquitin is a small protein (76 amino acids long) that binds and marks damaged or excess proteins for subsequent degradation by the proteasome complex.¹³⁵⁻¹³⁷ The proteasome is a multiunit protease complex that degrades ubiquitin-tagged proteins.¹³⁸ Ubiquitination also influences signal transduction and apoptosis by facilitating controlled degradation of pertinent regulatory molecules.^{139,140}

GENOMICS-RELATED LABORATORY METHODS

Polymerase Chain Reaction

Molecular cloning or DNA amplification often involves synthesis of multiple copies (clones) of a DNA sequence of

interest (the target DNA). This may be accomplished by either inserting the target DNA into a bacterium that is capable of extended cell division (recombinant DNA technology)¹⁴¹⁻¹⁴³ or using the polymerase chain reaction (PCR) for in vitro cloning of DNA (Figure 6).¹⁴⁴⁻¹⁴⁸ Polymerase chain reaction is an in vitro method of replicating relatively small DNA sequences into millions of copies over a short period.

A typical PCR reaction requires (1) two oligonucleotide primers (PCR primers), (2) a thermally stable DNA polymerase, (3) an ample amount of free deoxynucleotides, and (4) a small amount of DNA (the sample) that contains the sequence of interest (ie, the region to be amplified).¹⁴⁹ The PCR primers (short DNA sequences of approximately 20 bp in length) are synthesized in the laboratory, and their nucleotide sequences are designed to be complementary to DNA sequences that flank (lie on either side of) the DNA fragment of interest (the target).¹⁵⁰⁻¹⁵² Therefore, one has to know the nucleotide sequence of the target DNA fragment to use PCR technology. Because DNA synthesis can proceed only in the 5′ to 3′ direction, a PCR primer allows extension of DNA in only 1 direction (ie, the other side is a dead end). Therefore, to replicate both strands of the DNA molecule (the complementary strands), 2 PCR primers (forward and reverse primers) are needed, and they are placed on either side of the DNA sequence of interest on complementary strands (ie, the forward primer on 1 strand of DNA and the reverse primer on the second strand of DNA that is complementary to the first strand). This approach allows amplification of the double-stranded target sequence that is located between the 2 primers.^{145,148}

At the beginning of PCR, the sample DNA is first heated to a temperature greater than 90°C to separate the 2 paired DNA strands (DNA denaturation). The reaction is then cooled to approximately 40°C, in the presence of excess primers, to allow annealing (hybridization) of the specific primers to their complementary sequences on the target DNA (which is now in single strands). After the primers are bound to the DNA sequences that are immediately adjacent to the target sequence, new complementary strands of DNA, which are extensions of the bound PCR primers, are formed by recruiting the free nucleotides in the reaction. This step of primer extension occurs in an intermediate temperature of approximately 70°C with the help of a thermally stable DNA polymerase (an enzyme originally derived from the bacterium *Thermus aquaticus*).¹⁵³⁻¹⁵⁵ The newly synthesized DNA is double stranded, starting at the 5′ end with 1 of the 2 primers (Figure 6). After this first cycle of PCR, the same process is repeated 20 to 30 times (20-30 heating and cooling cycles), and each cycle results in the doubling of the target DNA region, with a final theoretical target amplification of approximately 1 million-fold or more

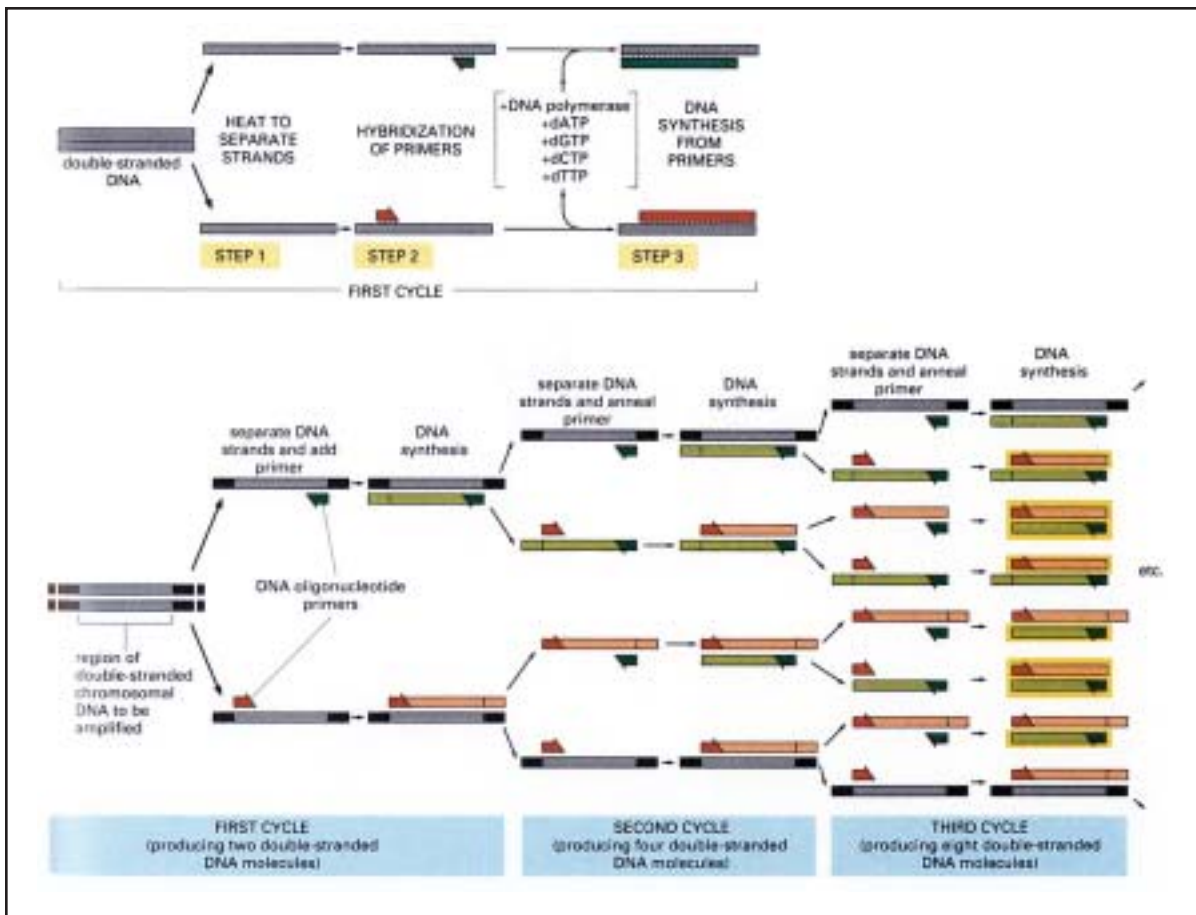


Figure 6. Amplification of DNA using polymerase chain reaction technique. See text for further details. Copyright 1998 from Alberts et al.^{30(p533)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

(Figure 6). Of note, the final amplified product consists mainly of fragments that comprise the region of DNA between the 2 primers (ie, the sequence of interest) (Figure 6).

Several modifications of PCR allow enhanced specificity and broader applicability. For example, the specificity of PCR may be enhanced by subjecting the product of the first PCR to another round of PCR amplification by using a second pair of primers that are placed internal of the first pair. The second pair of primers are called *nested* primers and the reaction *nested PCR*.¹⁵⁶ In another modification, more than 1 DNA sequence of interest may be interrogated in a single PCR by using multiple primers (multiplex PCR).^{157,158} Multiplex PCR promises to be an important diagnostic tool in infectious diseases.^{159,160}

In general, PCR technology works best when the DNA fragment of interest is relatively small (<1000 bp).¹⁴⁵ The technique is easily automated to allow detection and amplification of nanomolar quantities of DNA with a rapid turnaround time (a few hours). Because of the rapidity of the

assay and the small amount of test sample needed (nanogram quantities), PCR is an extremely useful and efficient method of genetic testing in disease diagnosis, monitoring of treatment response, detection of minimal residual disease, and tissue typing.^{156,161-164} Furthermore, the ability to extract DNA from various biological samples combined with PCR technology has substantially affected both forensic medicine and archeological studies.¹⁶⁵⁻¹⁶⁸ However, the extreme sensitivity of PCR makes it vulnerable to the possibility of false-positive test results due to trace amounts of contaminants.¹⁶⁹

Reverse Transcription-PCR.—There are several methods of detecting and quantifying specific mRNA from biological samples, including Northern blotting, nuclease protection assays, and reverse transcription (RT)-PCR. All 3 methods start with acquisition of fresh biological samples (blood, bone marrow, biopsy material, cell line, etc) that allows extraction of either total RNA or purified mRNA (polyadenylated RNA).^{170,171} In Northern blotting a specific

RNA species, among a mixture of RNAs, is first size-separated by gel electrophoresis. The RNA is subsequently transferred to a membrane where hybridization occurs with either radioactively or nonisotopically labeled DNA probes.^{172,173} A nuclease protection assay uses single-strand-specific endonucleases to remove by digestion single- but not double-stranded DNA.¹⁷⁴ In the particular assay, the sample RNA is incubated with excess labeled, single-stranded DNA molecules that are complementary to the specific mRNA. The subsequent hybridization between the mRNA and the labeled DNA strands protects the specific mRNA from being digested with the single-strand-specific endonucleases. In both Northern blotting and nuclease protection assays, the amount of steady-state mRNA level is estimated from the density of the labeling molecule.

In RT-PCR, traditional PCR is combined with reverse transcription to amplify mRNA.¹⁷⁵⁻¹⁷⁸ In reverse transcription, mRNA (sense) in the test sample is converted to a single-stranded cDNA (antisense) in a reaction mediated by reverse transcriptase. Reverse transcriptase (a retroviral enzyme) is an RNA-dependent DNA polymerase that can synthesize a DNA sequence that is complementary to a specific mRNA template.¹⁷⁹ The synthesized cDNA represents only the exons of the corresponding genomic DNA. The subsequent PCR amplification method is called *reverse transcription-polymerase chain reaction* to distinguish it from the standard PCR used to amplify genomic DNA. Reverse transcription-PCR is an extremely powerful and sensitive tool to detect and quantify mRNA (ie, gene expression) from minute amounts of test samples.^{176,180}

In RT-PCR, mRNA quantification may be performed both in relative (relative RT-PCR) and in absolute (competitive RT-PCR) terms. In relative RT-PCR, the amount of the target mRNA is estimated in comparison with a coamplified internal control. The resultant ratio is then compared across different samples. In competitive RT-PCR,¹⁸¹ a known quantity of a synthetic RNA molecule, which is similar in sequence but shorter compared with the target mRNA, is added to the reaction. This results in the synthesis of 2 cDNA templates (target and synthetic) that will compete for the same primers during PCR amplification. As a result, the abundance of the target PCR product is inversely proportional to the amount of the synthetic RNA that is added to the reaction. Therefore, by performing a series of PCRs with addition of serial dilutions of the synthetic RNA, a concentration curve can be constructed that is used to estimate the quantity of the endogenous target product.¹⁸²

Recent medical uses of RT-PCR include detection of lymph node micrometastasis in prostate cancer,¹⁸³ breast cancer tumor contamination of bone marrow,¹⁸⁴ viral con-

tamination of vaccines,¹⁸⁵ and monitoring of minimal residual disease in acute leukemia.^{186,187}

Real-Time PCR.—In RT-PCR, the steady-state level of a specific mRNA is indirectly estimated from the level of the corresponding amplified product. The amplified product is usually quantified after the reaction is completed (ie, “end point” quantification).¹⁸⁰ Alternatively, the presence and quantity of PCR products may be monitored in “real time” as the reaction progresses.¹⁸⁸ Monitoring the abundance of PCR products in multiple cycles of PCR provides dynamic range values that are more informative than end point measurements.¹⁸⁹ Of note, target amplification during the initial cycles of PCR is exponential because the presence of excess primers in the reaction minimizes renaturation of the complementary strands of the target DNA. As the concentration of an amplified product increases, the possibility of renaturation of its complementary strands increases. As a result, the amplification rate decreases and plateaus subsequently. The transition point from the exponential to the plateau phase of PCR may vary markedly for different samples. Therefore, it is more accurate to quantify the specific mRNA products of PCR during the exponential phase of target amplification. Real-time PCR accomplishes this task by quantifying reaction products in every cycle of PCR.^{178,188,190,191}

Real-time PCR depends on the detection and quantification of fluorescent molecules that are bound to the PCR product. Specific binding of fluorescent molecules to the PCR product is achieved by several methods, including TaqMan¹⁹² and SYBR green.¹⁹³ The TaqMan probe is a dual-labeled oligonucleotide probe that contains a fluorescent dye at its 5' base and another molecule (a quencher) at its 3' base. The quencher absorbs (quenches) fluorescence when it is situated close to the fluorescent molecule. This phenomenon is called *fluorescence resonance energy transfer* (FRET).¹⁹⁴ Therefore, the probe itself does not fluoresce. However, the fluorescent molecule is separated from the quencher molecule during primer extension through the exonuclease activity of *Taq* polymerase (the specific polymerase used in this particular PCR), whereby fluorescence is unmasked.^{195,196} Accordingly, the degree of fluorescence increases with an increase in the level of PCR product. Other FRET-dependent methods include molecular beacons¹⁹⁷ and scorpions.¹⁹⁸ In contrast, the SYBR green method is not FRET dependent. SYBR green is a dye that emits fluorescence only when it binds to double-stranded DNA.¹⁹³ Therefore, the fluorescence signal is enhanced with continued synthesis of the PCR product.

A discussion on the quantification of the signals obtained from real-time PCR is beyond the scope of this review, and the reader is referred to recent overviews on this subject.^{180,189} Potential applications of real-time PCR

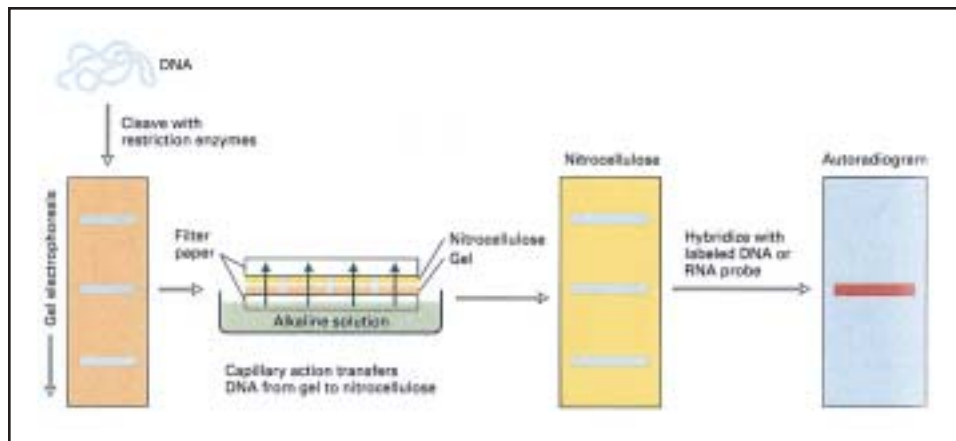


Figure 7. Southern blotting. See text for further details. Reproduced with permission from Lodish et al.^{121(p240)}

include disease diagnosis and monitoring,¹⁹⁸⁻²⁰¹ biosafety and genetic stability testing,²⁰² radiation exposure assessment,²⁰³ mutation detection,²⁰⁴ quantification of cytokine profiles in immune cells or inflamed tissue,¹⁸⁹ and in vivo imaging of cellular processes.²⁰⁵

Electrophoresis and Blotting

DNA, RNA, or proteins may be separated by size with use of agarose or polyacrylamide gels through electrophoresis. DNA and RNA are uniformly, negatively charged, and under most circumstances their gel separation is a function of their size and topology (linear fragments move in a different manner from supercoiled fragments of the same length). In contrast, proteins must first be denatured (unfolded) and negatively coated with an ionic detergent (sodium dodecyl sulfate) to allow gel separation solely by size, avoiding disturbances from their inherent positive and negative charge. Isoelectric focusing separates molecules by their net (positive or negative) charge.²⁰⁶

In Southern blotting (named after Dr Edward Southern, who described the technique), DNA is first cut into pieces using restriction endonucleases (enzymes that digest DNA at specific sites that are marked by a 4- to 8-member specific nucleotide sequence), and then the restriction fragments are subjected to agarose gel electrophoresis (Figure 7).²⁰⁷ The double-stranded DNA fragments on the gel are then turned into single strands by denaturation and transferred to a DNA-binding membrane, such as nitrocellulose or nylon, to make a permanent copy of single-stranded DNA. A DNA sequence of interest is then visualized by a radiolabeled reporter probe after a hybridization step followed by autoradiography. In addition to target identification and quantification, Southern blotting may allow detection of mutations that result in alterations of

restriction fragment lengths (ie, detect restriction fragment length polymorphisms [RFLPs]). Larger DNA molecules are separated by pulsed-field gel electrophoresis (intermittent application of voltage gradients to disorient large DNA molecules and optimize separation by size only).^{208,209}

Northern blotting (named to reflect similarity to Southern blotting) is used to identify specific RNAs.¹⁷³ Western blotting is used to identify specific proteins that have been separated by size with use of polyacrylamide gel electrophoresis.^{210,211} The size-fractionated proteins are transferred to a nitrocellulose membrane, and proteins of interest are identified and quantified by visualization with reporter-linked antibodies. In 2-dimensional gel electrophoresis, polyacrylamide gel electrophoresis is combined with separation of proteins by their isoelectric point (the pH where their overall charge is neutral).²¹²

Cutting, Pasting, and Altering the Topology of DNA

Different enzymes specialize in digesting and twisting (and untwisting) the DNA molecule. Endonucleases cleave DNA (break the sugar-phosphate ester bond between 2 deoxynucleotides) from the middle, while exonucleases sequentially remove nucleotides from the ends of DNA strands.^{213,214} DNA polymerases often possess exonuclease activity as part of their proofreading capability in addition to their synthetic function.²¹⁵ For many laboratory experiments, it is useful to remove the exonuclease activity of *Escherichia coli* DNA polymerase I by brief proteolytic digestion, leading to a synthetically active DNA polymerase called the *Klenow fragment*.^{216,217} Restriction endonucleases act at specific DNA sites, which are marked by precise 4- to 8-nucleotide sequences. Such restriction digestion of DNA is the key to recombinant technology and

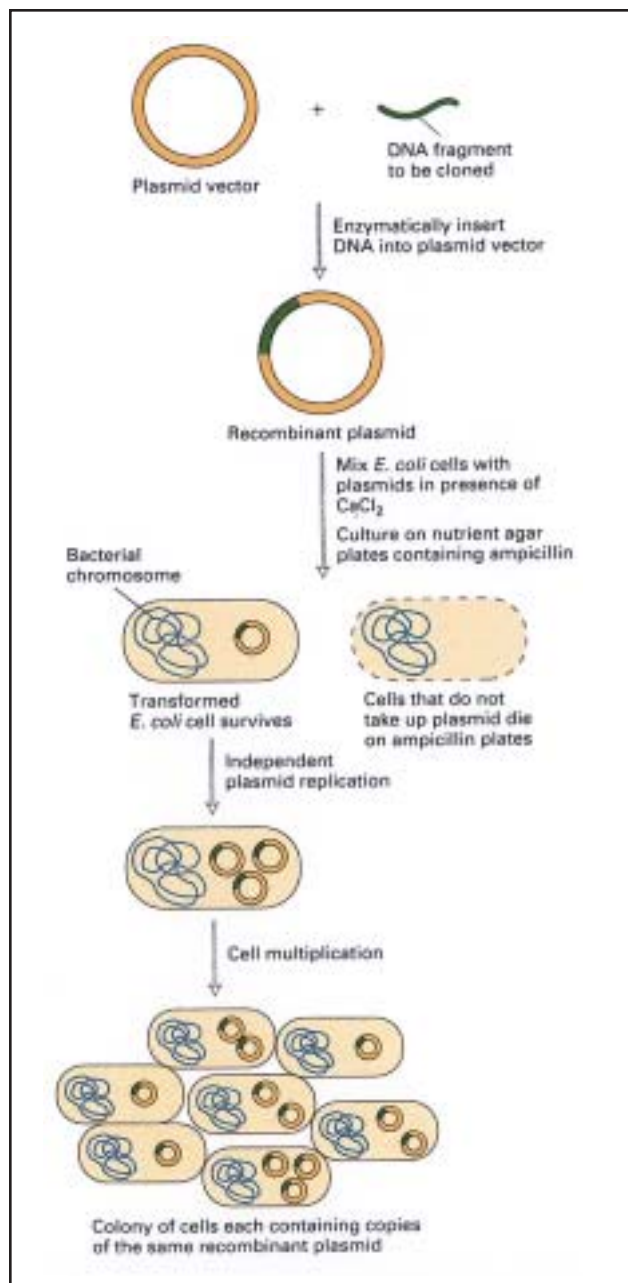


Figure 8. General procedure for cloning a DNA fragment in a plasmid vector. Reproduced with permission from Lodish et al.^{121(p210)}

chromosome mapping. Ligases attach 2 separate DNA fragments (reestablish the sugar-phosphate bond) and are important in recombinant technology.²¹⁸

DNA is naturally highly compacted but may take a relaxed (α -helical) or even a supercoiled state (a coiling and double coiling of the α -helical thread). A positive supercoil tends to tighten further and supercoil the neigh-

boring double helix, and a negative supercoil tends to unwind the neighboring double helix. Supercoiling and relaxing of DNA are often mediated by topoisomerases.^{219,220} Topoisomerases alter the topology (degree of coiling) of DNA by creating transient, either single (topoisomerase I) or double (topoisomerase II) strand breaks and allowing DNA to wind or unwind through those breaks. DNA gyrase is a type of topoisomerase that relieves supercoiling.²²¹

Clones, Libraries, and Polymorphisms

Clones.—Any DNA can be amplified (mass and strand number increased) in the laboratory by using either recombinant technology or PCR. Recombinant cloning starts with preparation of restriction fragments of DNA (eg, from human cells) that are then fused (using DNA ligase) to vector DNA (a carrier DNA molecule capable of self-replication) that has been digested with the same restriction enzymes so that they have similar split ends that allow DNA ligation.²²² This “recombinant” DNA molecule is transported and amplified by the vector (a carrier molecule capable of self-replication) (Figure 8). Vectors in current use include plasmids (circular double-stranded DNA molecules that replicate extrachromosomally in bacteria), bacteriophages (bacterial virus with a larger than plasmid double-stranded DNA molecule), cosmids (large plasmids capable of carrying large DNA inserts), and bacterial artificial chromosomes (BACs) or yeast artificial chromosomes (YACs) (artificially constructed chromosome-like structures that carry very large DNA fragments to be amplified [cloned]).^{29,223-226} Specifically, plasmids and bacteriophages are used for the insertion of small foreign DNA (<20 kb); BACs and YACs are used for larger DNA molecules (300-1000 kb). Because YACs can carry large DNA fragments, they have been particularly important in the mapping of the human genome.²²⁷

Libraries.—Chromosomal DNA isolated directly from a cell is called *genomic DNA*. DNA made from reverse transcription of mRNA is called *complementary DNA*. A set of DNA clones from a particular source is called a *DNA library* (genomic or cDNA library depending on the source of the clones) (Figure 9). To prepare a human DNA library, genomic DNA or cDNA is first digested with restriction enzymes, and the restriction fragments (inserts) are recombined with a similarly digested vector DNA. A population of recombinant DNA molecules with different human DNA inserts is then cloned in a bacterial host. A collection of thousands of vector-linked human DNA inserts represents the human DNA library.²²⁸

Complementary DNA libraries have the advantage of carrying only exons (protein-encoding regions of genes) and may be modified by flanking the inserts with transcription and translation signals (expression vectors), thus allow-

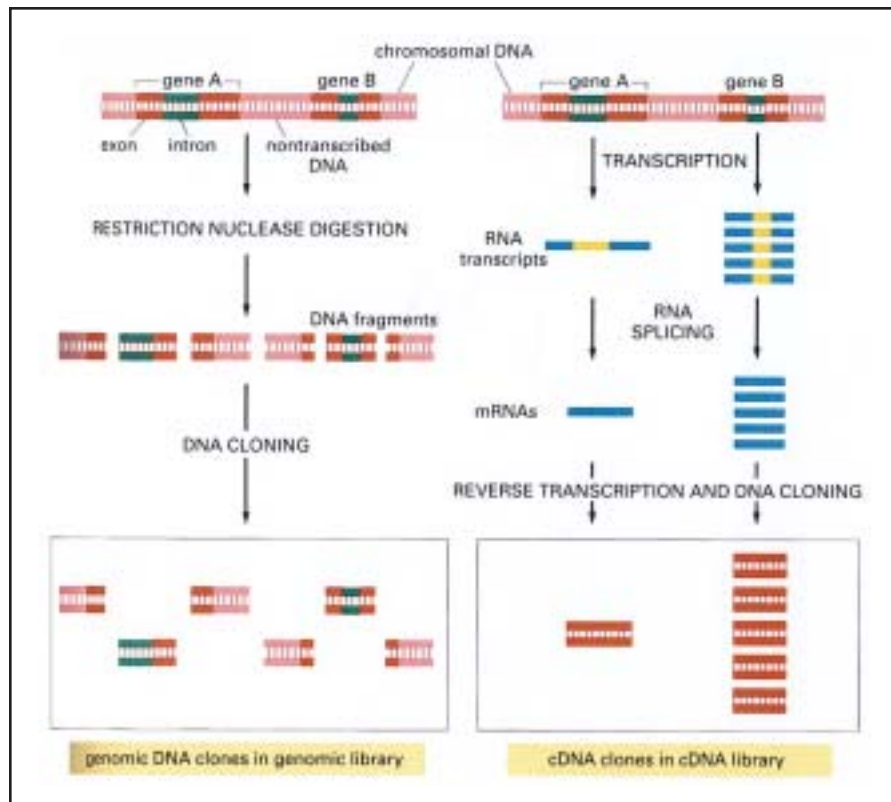


Figure 9. Making of a genomic vs complementary DNA (cDNA) library. See text for further details. Copyright 1998 from Alberts et al.^{30(p331)} Reproduced by permission of Garland Publishing, Inc, part of The Taylor & Francis Group.

ing screening for coded proteins.²²⁹ Complementary DNA clones representing named and unnamed genes may be partially sequenced to generate expressed sequence tags (ESTs). Although incomplete, ESTs provide a snapshot of the expressed sequences in a given cell type and also can be used to identify gene location in genomic DNA sequences.²³⁰ Further sequence information on the identified gene can then be obtained by chromosome walking (described subsequently). A section of a DNA sequence that starts with an initiation codon, which is a recognition site to start protein synthesis (translation), and ends in a stop codon (a codon that signals termination of translation), is called an *open reading frame*, suggesting the potential to code for a protein.

Polymorphisms.—The human genome is extensively affected by sequence variation.¹ When the rate of variation at a specific point in the DNA yields a variant sequence that is found in more than 1% of a given population, it is referred to as a *polymorphism*. When the incidence of a variant sequence is less than 1%, it is referred to as a *mutation*. However, such definitions are imprecise because what may be considered a mutation in a given population

may qualify as a polymorphism in another population (eg, sickle cell mutation).

Genetic polymorphisms may be revealed by restriction enzyme digestion that results in variably sized restriction fragments, a phenomenon known as *restriction fragment length polymorphism* (Figure 10). Usually occurring in intergenic DNA regions, RFLPs generally have no phenotypic effect. Certain regions of DNA are highly polymorphic because of the presence of variable numbers of tandem repeats (VNTRs).²³¹ The latter are short DNA sequences that are repeated in sequence multiple times. Variations among individuals in the number and size of such repeat sequences may be revealed by using RFLPs induced by restriction enzymes that cut near the boundaries that encompass the repeat sequences. The degree of polymorphism in VNTRs is sufficiently high to allow DNA fingerprinting.²³²

Single nucleotide polymorphisms occur as often as every 200 bp in the human genome.²³³ Single nucleotide polymorphisms may be detected by microarrays and have the potential to be used in studies of genetic susceptibility to diseases.²³⁴ Of note, however, only about 1% of single

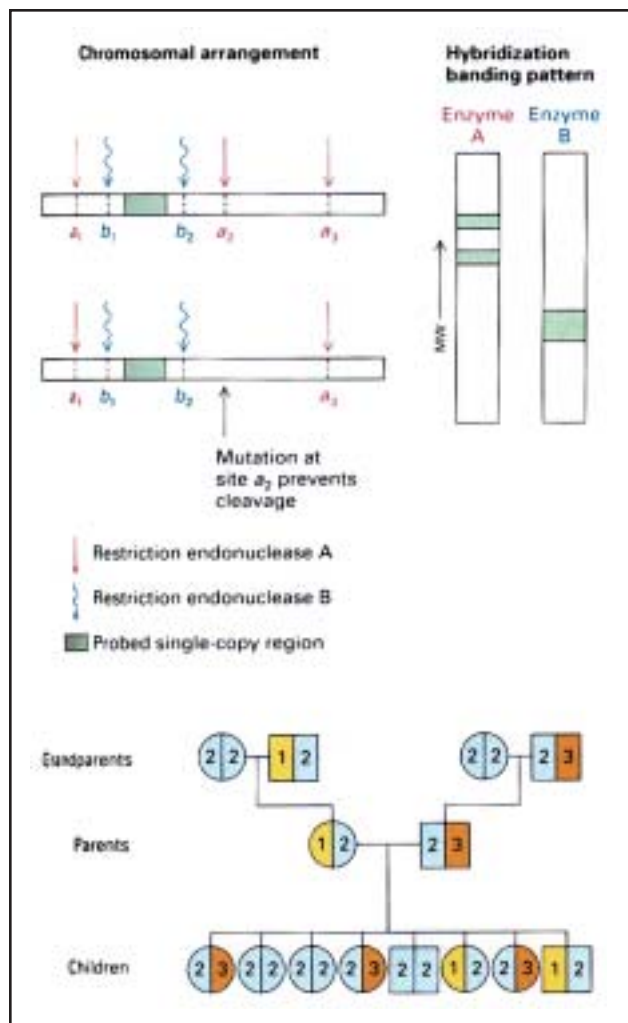


Figure 10. Analysis of restriction fragment length polymorphisms. See text for further details. Reproduced with permission from Lodish et al.^{121(p271)}

nucleotide polymorphisms result in an alteration in protein coding at the amino acid level.

DNA Sequencing

DNA sequencing entails spelling out the nucleotide chain of a DNA region of interest.²³⁵⁻²³⁷ Sufficient quantities of the target DNA region of interest are obtained from either a bacterial-cloned fragment or a PCR-amplified sequence. A short DNA primer is used to start the synthesis of a complementary strand to the original template to be sequenced. With the use of specific chemical analogues of nucleotides (dideoxynucleotides), which are capable of terminating chain extension at specific nucleotide bases, the synthesis of the complementary strand is halted (nested) at specific nucleotide bases, resulting in the synthesis of DNA

fragments with varying lengths that differ by 1 nucleotide (Figure 11). The set of nested fragments that are radiolabeled or tagged with fluorescent dyes can then be separated by gel electrophoresis, resulting in a pattern that allows the determination of the specific nucleotide at the end of each successive fragment. Currently, an automated computerized laser detection system is used for rapid accrual of sequence information.

Gene Mapping

Human genetic mapping entails (1) the localization or mapping of genes to each of the 23 chromosome pairs and (2) the determination of the order and spacing of the gene on the particular chromosome.²³⁸ This is done in 1 of 2 ways: physical mapping (based on estimates of physical distance measured in bp) or genetic linkage mapping (based on the frequency of meiotic crossing over that is observed between 2 loci of a chromosome).²³⁹ Linkage distance is measured in centiMorgans (cM), named for Thomas H. Morgan, the father of *Drosophila* (fruit fly) genetics, who defined the concept of genetic linkage.²⁴⁰ Locating a disease gene is the first step toward cloning the gene itself.

Physical Mapping.—Physical mapping of genes may be achieved by different methods with differing levels of resolution that range from an entire chromosome to a single base pair (the base sequence). Methods of low-resolution physical mapping include karyotyping, fluorescence in situ hybridization (FISH), and somatic cell hybridization.

Human chromosomes may be identified by size and by distinctive banding patterns seen under light microscopy after staining with special DNA-binding dyes (Giemsa).²⁴¹ This method of chromosome visualization is called *karyotyping* and provides a metaphase picture of the 22 homologous pairs of autosomal chromosomes and the 1 additional pair of sex chromosomes lined up by size. Such a chromosome picture allows identification of gross abnormalities (chromosomal deletions, duplications, translocations) that may be associated with a particular disease phenotype (eg, chronic myeloid leukemia is associated with a gross translocation between chromosomes 9 and 22).²⁴² As a result, the disease gene is physically mapped to the vicinity of the abnormal karyotypic lesion.

In FISH, a priori knowledge of a DNA sequence of the specific gene or of a marker near the gene is required. Fluorescent-tagged probes, by cloning of the specific sequence, are then prepared and used to localize the gene of interest to a particular chromosome by slide exposure of metaphase or interphase chromosomes with the labeled probe.²⁴³

In somatic cell hybridization, human and nonhuman cells are fused under specific experimental conditions.²⁴⁴

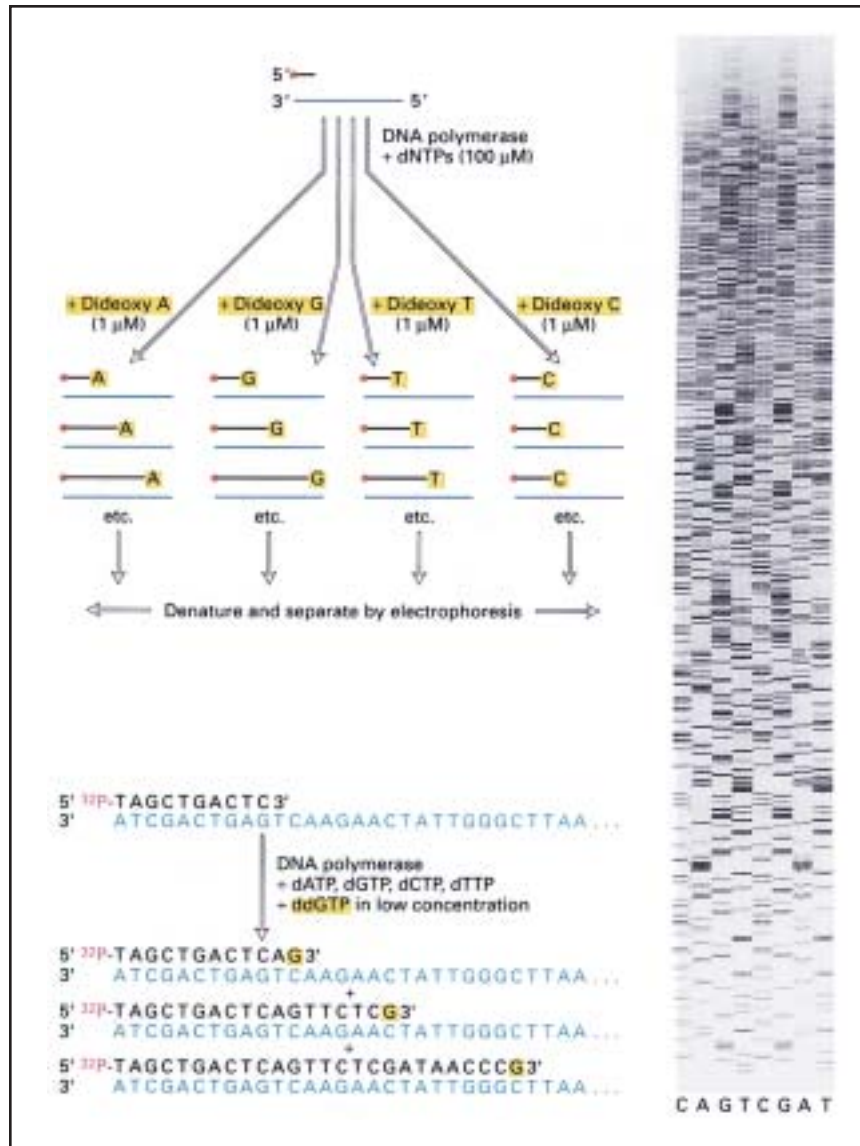


Figure 11. Sanger (dideoxy) method for sequencing DNA fragments. See text for further details. Reproduced with permission from Lodish et al.^{121(p234)}

Subsequently, the hybrid cells are allowed to replicate under specific conditions whereby they start losing a variable number of the human chromosomes.^{245,246} Eventually, hybrid cells that have retained 1 or 2 specific human chromosomes are selected for further testing. Karyotyping allows distinction of human from nonhuman (usually mouse) chromosomes and identification of the specific human chromosome that is retained in a hybrid cell. Additional experiments are then performed to look for the presence or absence of a specific DNA sequence or protein product in a given hybrid cell, and this would allow the assignment of the gene for a particular protein product to a specific chro-

mosome. Somatic cell hybrids may provide additional resolution when the hybrid is constructed from cells containing either structurally abnormal or radiation-damaged human chromosomes (so-called radiation hybrids).²⁴⁷ Similarly, somatic cell hybridization and FISH techniques may be combined for improved resolution.²⁴⁸

Other physical maps include cDNA maps that localize the chromosomal positions of expressed genes,^{249,250} RFLP maps of restriction cleavage sites,²⁵¹ and contig maps of overlapping sequences of YACs, BACs, or cosmids (Figure 12) (described in more detail subsequently).²²⁷ The resolution grade of karyotypic identification of a genetic

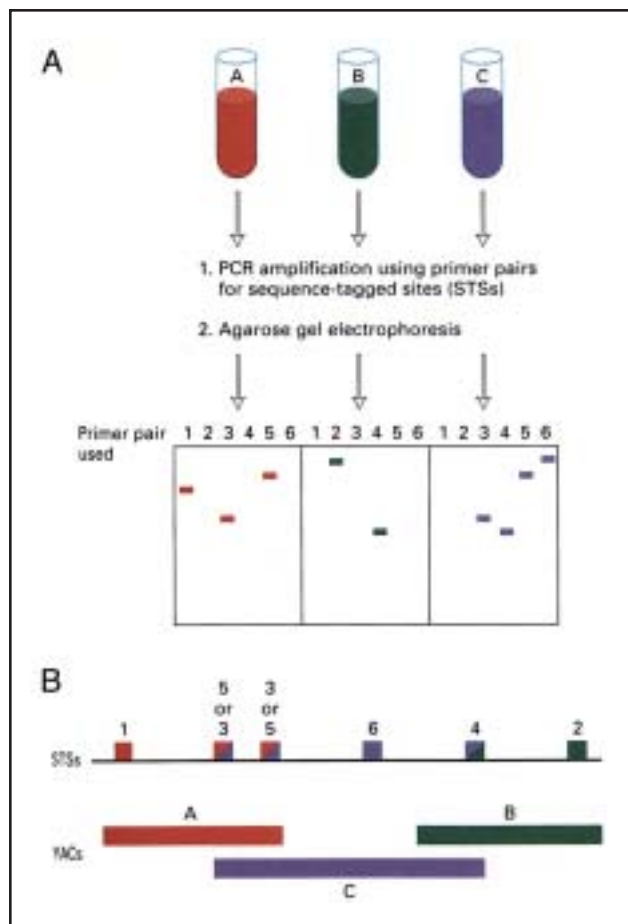


Figure 12. Ordering of contiguous overlapping yeast artificial chromosome (YAC) clones. A, Aliquots of DNA prepared from each YAC clone are polymerase chain reaction (PCR)-amplified by using primer pairs 1 through 6 corresponding to the ends of various sequence-tagged sites (STSs). B, Based on the gel patterns in this example, the 3 YAC clones can be ordered as shown. Reproduced with permission from Lodish et al.^{121(p277)}

lesion is estimated at 10 million bases (Mb), that of FISH at about 2 Mb, that of restriction map at 0.5 to 1 Mb, and contig maps at an even lower scale.

Genetic Linkage Mapping.—Genes that are located on different chromosomes segregate independently during meiosis (the law of independent assortment). In other words, 2 genes on 2 different chromosomes have a 50% chance of being inherited together in a daughter cell. In contrast, genes on the same chromosome have a tendency to be inherited together (ie, are genetically linked).²³ However, the chance of segregating together is not 100% because of the “phenomenon” (process) of recombination (crossing over).²⁵²⁻²⁵⁶ During recombination, the naturally occurring exchange of chromosome regions between ho-

mologous pairs of chromosomes happens during meiosis.²⁵⁷ The chance for a recombination event to occur depends on the distance between the 2 genes on the same chromosome.²⁵⁸ The closer the genes are to each other, the less chance for a recombination event to occur and vice versa.

Genetic linkage is defined as the tendency for genes on the same chromosome to stick together during meiosis (Figure 13). The strength of this linkage is used as a unit of measurement (Morgan) of the distance between the 2 linked genes.²⁵⁹ One centiMorgan denotes a recombination event of 1% (ie, the genes stick together 99% of the time during meiosis). A genetic distance of 1 cM therefore denotes 2 genes that are relatively close to each other. Accordingly, a genetic linkage map may be perceived as a map of a chromosome with genes or DNA markers that are aligned in a linear fashion with intervening distances that reflect the frequency of recombination between the adjacent genes or markers.²⁵² The total genetic length of the human genome is estimated at 3000 cM. As such, 1 cM corresponds to about 1 million bp.

Of note, chromosomal regions (DNA sequences) that are evaluated for linkage analysis may not be associated necessarily with a phenotypically apparent trait. For example, hair color is a phenotypically apparent trait, and variation in the DNA sequence responsible for hair color is reflected by variation in hair color. In contrast, DNA sequence variation of a phenotypically silent genetic marker is revealed by molecular techniques that exploit the presence or absence of specific restriction enzyme cleavage sites (RFLPs) and variations in the number of tandem repeats at specific chromosomal regions (VNTRs).²⁶⁰ Regardless of whether the study markers are phenotypically apparent (genes) or not (RFLPs, VNTRs), linkage analysis requires that heterozygosity exists at the locus of both the gene and the genetic marker (ie, alleles of the specific gene or DNA sequence must be heterozygous).²⁶¹

Alleles are variants of the same gene occupying the same locus on chromosomes. Therefore, a homologous pair of chromosomes may either have the same allele on the 2 chromosomes forming the pair (ie, are homozygous for the specific allele) or have 2 different alleles at a specific locus (ie, are heterozygous). Genetic mapping methods are based on the observed segregation of alleles on a homologous pair of chromosomes. For example, if 2 separate traits are inherited together 99% of the time, then the alleles responsible for the traits are probably on the same chromosome at a distance that is 1 cM between each other.

Precise Localization and Sequencing of the Gene of Interest.—In most diseases, a causal genetic aberration is not yet known. As described previously, one can use linkage analysis to associate a putative disease gene with a

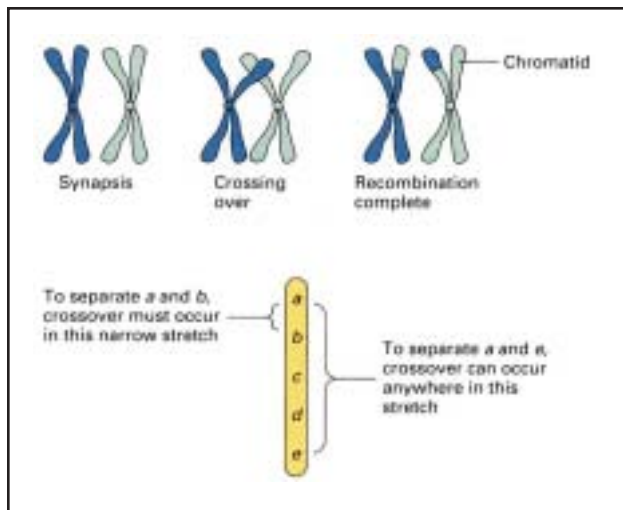


Figure 13. Recombination during meiosis. See text for further details. Reproduced with permission from Lodish et al.^{121(p270)}

chromosomal locale that is detectable either molecularly or phenotypically.²⁶⁰ Once the gross location of the putative disease gene is determined by linkage analysis, one can apply positional cloning for more precise identification and sequencing of the gene of interest.^{262,263}

Positional cloning starts with the selection, by using the DNA sequence of the linked marker as a probe, of vector clones from a genomic library that carry DNA sequences that overlap with those from the linked genetic marker. As mentioned previously, a genomic library is constructed by partially digesting the entire genome with restriction enzymes, followed by cloning of the resultant fragments into appropriate vectors.²⁶⁴ Because the DNA products of restriction digestion are relatively large, the vectors usually used are YAC and BAC vectors. Using the DNA sequence of the linked marker, one can pick YAC or BAC clones with sequences that overlap with those of the linked marker.²⁶⁵ Alternatively, sequence tagged sites, which are unique, relatively short DNA sequences with known chromosomal location, may be used to identify clones with overlapping sequences. Subsequently, the overlapping sequences themselves may be used as probes to select additional YAC or BAC clones with sequences that are partially overlapping with the probes used.^{226,266}

The stepwise process of identifying YAC or BAC clones based on overlapping DNA sequences is called *chromosome walking*.²⁶⁷ This process of chromosome walking continues until the disease gene is identified. The linear alignment of YACs or BACs with overlapping sequences results in a continuous DNA segment called a *contig map*. Such contig maps have been essential for the

analysis of human and nonhuman genomes as well as for specific chromosomal regions.^{225,227,268,269}

Probe-identified clones can also be individually mapped for sites complementary to ESTs. An EST that cosegregates with the gene of interest is determined and then sequenced. Alternatively, the clone containing the linkage markers may have already been sequenced, and resident genes are sought by looking for ORFs. Finally, the candidate gene is identified by mutational analysis that compares ORFs from a patient-derived DNA library to those obtained from nonaffected individuals.

CONCLUSION

This review on the basic principles of molecular biology is not intended to be comprehensive but to serve as an introduction to this subject. Additional information is easily accessible from numerous outstanding textbooks of molecular biology.^{30,121,270}

REFERENCES

- Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.
- Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
- Meldrum D. Automation for genomics, part two: sequencers, microarrays, and future trends. *Genome Res*. 2000;10:1288-1303.
- Gershon D. Microarray technology: an array of opportunities. *Nature*. 2002;416:885-891.
- Gulino A. Biotechnology and molecular diagnostics. *Forum (Genova)*. 1999;9(3, suppl 3):37-46.
- Jordan B, Charest A, Dowd JF, et al. Genome complexity reduction for SNP genotyping analysis. *Proc Natl Acad Sci U S A*. 2002;99:2942-2947.
- Topol EJ, McCarthy J, Gabriel S, et al. Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation*. 2001;104:2641-2644.
- Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503-511.
- Marshall T, Williams KM. Proteomics and its impact upon biomedical science. *Br J Biomed Sci*. 2002;59:47-64.
- Service RF. Structural genomics offers high-speed look at proteins. *Science*. 2000;287:1954-1956.
- Chance MR, Bresnick AR, Burley SK, et al. Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci*. 2002;11:723-738.
- Altman RB, Klein TE. Challenges for biomedical informatics and pharmacogenomics. *Annu Rev Pharmacol Toxicol*. 2002;42:113-133.
- Weng Z, DeLisi C. Protein therapeutics: promises and challenges for the 21st century. *Trends Biotechnol*. 2002;20:29-35.
- Bayat A. Science, medicine, and the future: bioinformatics. *BMJ*. 2002;324:1018-1022.
- Roos DS. Computational biology: bioinformatics—trying to swim in a sea of data. *Science*. 2001;291:1260-1261.
- Collins FS, McKusick VA. Implications of the Human Genome Project for medical science. *JAMA*. 2001;285:540-544.
- Subramanian G, Adams MD, Venter JC, Broder S. Implications of the human genome for understanding human biology and medicine. *JAMA*. 2001;286:2296-2307.

18. Butler D. Genomics: are you ready for the revolution? *Nature*. 2001;409:758-760.
19. Magee J, Gordon JI, Whelan A. Bringing the human genome and the revolution in bioinformatics to the medical school classroom: a case report from Washington University School of Medicine. *Acad Med*. 2001;76:852-855.
20. Fears R, Weatherall D, Poste G. The impact of genetics on medical education and training. *Br Med Bull*. 1999;55:460-470.
21. Rothstein MA, Epps PG. Ethical and legal implications of pharmacogenomics. *Nat Rev Genet*. 2001;2:228-231.
22. Mendel G. Experiments in plant hybridization. *Verh Naturforsch Vereines Brunn*. 1865;4:3-47.
23. Morgan TH. Sex-limited inheritance in *Drosophila*. *Science*. 1910;32:120-122.
24. Demerec M. What is a gene? *J Hered*. 1933;24:368-378.
25. Avery O, MacLeod C, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal type. *J Exp Med*. 1944;79:137-158.
26. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*. 1952;36:39-56.
27. Beadle GW, Tatum E. Genetic control of biochemical reactions in neurospora. *Proc Natl Acad Sci U S A*. 1941;27:499-506.
28. Jhiang SM, Garey JR, Riggs AF. Exon-intron organization in genes of earthworm and vertebrate globins. *Science*. 1988;240:334-336.
29. Cohen JA. Chemistry and structure of nucleic acids of bacteriophages: many forms of nucleic acids of bacteriophages show the ways that information is stored and reproduced. *Science*. 1967;158:343-351.
30. Alberts B, Bray D, Johnson A, et al. *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell*. New York, NY: Garland Publishing; 1998.
31. Dhingra MM, Sarma RH. Why do nucleic acids have 3'5' phosphodiester bonds? *Nature*. 1978;272:798-801.
32. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-738.
33. Meselson M, Stahl FW. The replication of DNA in *E. coli*. *Proc Natl Acad Sci U S A*. 1958;44:671-682.
34. Saha S, Sparks AB, Rago C, et al. Using the transcriptome to annotate the genome. *Nat Biotechnol*. 2002;20:508-512.
35. Avise JC. Evolving genomic metaphors: a new look at the language of DNA. *Science*. 2001;294:86-87.
36. Schmid CW, Deininger PL. Sequence organization of the human genome. *Cell*. 1975;6:345-358.
37. Gu Z, Wang H, Nekrutenko A, Li WH. Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence. *Gene*. 2000;259:81-88.
38. Dott PJ, Chuang CR, Saunders GF. Inverted repetitive sequences in the human genome. *Biochemistry*. 1976;15:4120-4125.
39. Willard HF. The genomics of long tandem arrays of satellite DNA in the human genome. *Genome*. 1989;31:737-744.
40. Moyzis RK, Torney DC, Meyne J, et al. The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics*. 1989;4:273-289.
41. Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev*. 1996;6:743-748.
42. Moustacchi E, Williamson DH. Physiological variations in satellite components of yeast DNA detected by density gradient centrifugation. *Biochem Biophys Res Commun*. 1966;23:56-61.
43. Moore DE, Abuelo JG. Human repetitive DNA, satellite DNA and chromosome interarm fibres. *Nature*. 1971;234:467-468.
44. Yunis JJ, Roldan L, Yasmin WG, Lee JC. Staining of satellite DNA in metaphase chromosomes. *Nature*. 1971;231:532-533.
45. Moxon ER, Wills C. DNA microsatellites: agents of evolution? *Sci Am*. 1999;280:94-99.
46. Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann Hum Genet*. 1987;51(pt 4):269-288.
47. Prak ET, Kazazian HH Jr. Mobile elements and the human genome [published correction appears in *Nat Rev Genet*. 2001;2:78]. *Nat Rev Genet*. 2000;1:134-144.
48. Jagadeeswaran P, Forget BG, Weissman SM. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? *Cell*. 1981;26(2, pt 2):141-142.
49. Patience C, Wilkinson DA, Weiss RA. Our retroviral heritage. *Trends Genet*. 1997;13:116-120.
50. Johns DR. The other human genome: mitochondrial DNA and disease. *Nat Med*. 1996;2:1065-1068.
51. Nass MM. Mitochondrial DNA: advances, problems, and goals. *Science*. 1968;165:25-35.
52. Cann RL, Brown WM, Wilson AC. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics*. 1984;106:479-499.
53. Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1979;76:1967-1971.
54. Cann RL, Wilson AC. Length mutations in human mitochondrial DNA. *Genetics*. 1983;104:699-711.
55. Mattson MP. Mother's legacy: mitochondrial DNA mutations and Alzheimer's disease. *Trends Neurosci*. 1997;20:373-375.
56. Avise JC. Mitochondrial DNA and the evolutionary genetics of higher animals. *Philos Trans R Soc Lond B Biol Sci*. 1986;312:325-342.
57. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. African populations and the evolution of human mitochondrial DNA. *Science*. 1991;253:1503-1507.
58. Ferris SD, Wilson AC, Brown WM. Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1981;78:2432-2436.
59. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974;184:868-871.
60. Elgin SC. Chromatin structure, DNA structure. *Nature*. 1982;300:402-403.
61. Brown SW. Heterochromatin. *Science*. 1966;151:417-425.
62. Yasmin WG, Yunis JJ. Isolation of mammalian heterochromatin and euchromatin. *Methods Cell Biol*. 1974;8:151-177.
63. Wallrath LL. Unfolding the mysteries of heterochromatin. *Curr Opin Genet Dev*. 1998;8:147-153.
64. Quintana-Murci L, Jamain S, Fellous M. Origin and evolution of mammalian sex chromosomes [in French]. *C R Acad Sci III*. 2001;324:1-11.
65. Nirenberg MW, Leder P. RNA codewords and protein synthesis. *Science*. 1964;145:1399-1407.
66. Efstratiadis A, Kafatos FC, Maniatis T. The primary structure of rabbit beta-globin mRNA as determined from cloned DNA. *Cell*. 1977;10:571-585.
67. Lewin B. Units of transcription and translation: the relationship between heterogeneous nuclear RNA and messenger RNA. *Cell*. 1975;4:11-20.
68. Busby S, Ebright RH. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell*. 1994;79:743-746.
69. Saltzman AG, Weinmann R. Promoter specificity and modulation of RNA polymerase II transcription. *FASEB J*. 1989;3:1723-1733.
70. Jeang KT, Khoury G. The mechanistic role of enhancer elements in eukaryotic transcription. *Bioessays*. 1988;8:104-107.
71. Reinberg D, Orphanides G, Ebright R, et al. The RNA polymerase II general transcription factors: past, present, and future. *Cold Spring Harb Symp Quant Biol*. 1998;63:83-103.
72. Muller CW. Transcription factors: global and detailed views. *Curr Opin Struct Biol*. 2001;11:26-32.
73. Lopez-Rodriguez C, Aramburu J, Rakeman AS, et al. NF-AT5: the NF-AT family of transcription factors expands in a new direction. *Cold Spring Harb Symp Quant Biol*. 1999;64:517-526.

74. Jones NC, Rigby PW, Ziff EB. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes Dev.* 1988;2:267-281.
75. Sassone-Corsi P, Verma IM. Modulation of c-fos gene transcription by negative and positive cellular factors. *Nature.* 1987;326:507-510.
76. Mathis DJ, Chambon P. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. *Nature.* 1981;290:310-315.
77. Rowlands T, Baumann P, Jackson SP. The TATA-binding protein: a general transcription factor in eukaryotes and archaeobacteria. *Science.* 1994;264:1326-1329.
78. Nimer S, Fraser J, Richards J, Lynch M, Gasson J. The repeated sequence CATT(A/T) is required for granulocyte-macrophage colony-stimulating factor promoter activity. *Mol Cell Biol.* 1990;10:6084-6088.
79. White RJ, Jackson SP. The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. *Trends Genet.* 1992;8:284-288.
80. Pabo CO, Sauer RT. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem.* 1992;61:1053-1095.
81. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. *Genes Dev.* 1996;10:2657-2683.
82. Khorasanizadeh S, Rastinejad F. Transcription factors: the right combination for the DNA lock. *Curr Biol.* 1999;9:R456-R458.
83. Proudfoot N. Connecting transcription to messenger RNA processing. *Trends Biochem Sci.* 2000;25:290-293.
84. Fong N, Bentley DL. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* 2001;15:1783-1795.
85. Winicov I, Perry RP. Synthesis, methylation, and capping of nuclear RNA by a subcellular system. *Biochemistry.* 1976;15:5039-5046.
86. Minvielle-Sebastia L, Keller W. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr Opin Cell Biol.* 1999;11:352-357.
87. Reeder RH, Lang WH. Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends Biochem Sci.* 1997;22:473-477.
88. Darnell JE, Philipson L, Wall R, Adesnik M. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science.* 1971;174:507-510.
89. Crick F. Split genes and RNA splicing. *Science.* 1979;204:264-271.
90. Palacios IM. RNA processing: splicing and the cytoplasmic localization of mRNA. *Curr Biol.* 2002;12:R50-R52.
91. Yang VW, Lerner MR, Steitz JA, Flint SJ. A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proc Natl Acad Sci U S A.* 1981;78:1371-1375.
92. Ohshima Y, Itoh M, Okada N, Miyata T. Novel models for RNA splicing that involve a small nuclear RNA. *Proc Natl Acad Sci U S A.* 1981;78:4471-4474.
93. Choi YD, Grabowski PJ, Sharp PA, Dreyfuss G. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science.* 1986;231:1534-1539.
94. Kiss T. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 2001;20:3617-3622.
95. Hernandez N. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem.* 2001;276:26733-26736.
96. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell.* 1982;31:147-157.
97. Valadkhan S, Manley JL. Splicing-related catalysis by protein-free snRNAs. *Nature.* 2001;413:701-707.
98. Newman A. Molecular biology: RNA enzymes for RNA splicing. *Nature.* 2001;413:695-696.
99. Rogers J, Wall R. A mechanism for RNA splicing. *Proc Natl Acad Sci U S A.* 1980;77:1877-1979.
100. Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science.* 1984;225:898-903.
101. Keller W. The RNA lariat: a new ring to the splicing of mRNA precursors. *Cell.* 1984;39(3, pt 2):423-425.
102. Chetverin AB. The puzzle of RNA recombination. *FEBS Lett.* 1999;460:1-5.
103. Choi E, Kuehl M, Wall R. RNA splicing generates a variant light chain from an aberrantly rearranged kappa gene. *Nature.* 1980;286:776-779.
104. Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. *Nature.* 1999;399:75-80.
105. Lamond AI. RNA splicing: running rings around RNA. *Nature.* 1999;397:655-656.
106. Graveley BR, Hertel KJ, Maniatis T. SR proteins are 'locators' of the RNA splicing machinery [letter]. *Curr Biol.* 1999;9:R6-R7.
107. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell.* 1992;68:365-375.
108. Wu JI, Reed RB, Grabowski PJ, Artzt K. Function of quaking in myelination: regulation of alternative splicing. *Proc Natl Acad Sci U S A.* 2002;99:4233-4238.
109. Cooper TA. Highlights of alternative splicing regulation session: yes, no, maybe—a history of paradigm shifts. *Sci STKE.* 2001;2001:PE35.
110. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene.* 2001;276:73-81.
111. Adams JM, Cory S. Untranslated nucleotide sequence at the 5'-end of R17 bacteriophage RNA. *Nature.* 1970;227:570-574.
112. Forget BG, Marotta CA, Weissman SM, Cohen-Solal M. Nucleotide sequences of the 3'-terminal untranslated region of messenger RNA for human beta globin chain. *Proc Natl Acad Sci U S A.* 1975;72:3614-3618.
113. Chang JC, Temple GF, Poon R, Neumann KH, Kan YW. The nucleotide sequences of the untranslated 5' regions of human alpha- and beta-globin mRNAs. *Proc Natl Acad Sci U S A.* 1977;74:5145-5149.
114. Veyrune JL, Hesketh J, Blanchard JM. 3' Untranslated regions of c-myc and c-fos mRNAs: multifunctional elements regulating mRNA translation, degradation and subcellular localization. *Prog Mol Subcell Biol.* 1997;18:35-63.
115. Wormington M. Unmasking the role of the 3' UTR in the cytoplasmic polyadenylation and translational regulation of maternal mRNAs. *Bioessays.* 1994;16:533-535.
116. Sonenberg N. mRNA translation: influence of the 5' and 3' untranslated regions. *Curr Opin Genet Dev.* 1994;4:310-315.
117. Dickey LF, Wang YH, Shull GE, Wortman IA III, Theil EC. The importance of the 3'-untranslated region in the translational control of ferritin mRNA. *J Biol Chem.* 1988;263:3071-3074.
118. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. *Genome Biol.* 2002;3:0004.1-0004.10.
119. Lafontaine DL, Tollervey D. The function and synthesis of ribosomes. *Nat Rev Mol Cell Biol.* 2001;2:514-520.
120. Schimmel P, Ribas de Pouplana L. Transfer RNA: from minihelix to genetic code. *Cell.* 1995;81:983-986.
121. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. *Molecular Cell Biology.* 4th ed. New York, NY: WH Freeman and Co; 2000.
122. Brown GL, Lee S. Amino-acid-transfer ribonucleic acid: structure and function. *Br Med Bull.* 1965;21:236-243.
123. Saks ME, Sampson JR, Abelson JN. The transfer RNA identity problem: a search for rules. *Science.* 1994;263:191-197.
124. Noller HF, Yusupov MM, Yusupova GZ, Baucom A, Cate JH. Translocation of tRNA during protein synthesis. *FEBS Lett.* 2002;514:11-16.

125. Moore PB. Ribosomes: protein synthesis in slow motion. *Curr Biol*. 1997;7:R179-R181.
126. Pestova TV, Hellen CU. The structure and function of initiation factors in eukaryotic protein synthesis. *Cell Mol Life Sci*. 2000; 57:651-674.
127. Czworkowski J, Moore PB. The elongation phase of protein synthesis. *Prog Nucleic Acid Res Mol Biol*. 1996;54:293-332.
128. Tate WP, Brown CM. Translational termination: "stop" for protein synthesis or "pause" for regulation of gene expression. *Biochemistry*. 1992;31:2443-2450.
129. Tuite MF, Stansfield I. Termination of protein synthesis. *Mol Biol Rep*. 1994;19:171-181.
130. Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Annu Rev Biophys Biomol Struct*. 2001;30:457-475.
131. Wold F. In vivo chemical modification of proteins (post-translational modification). *Annu Rev Biochem*. 1981;50:783-814.
132. Yewdell JW. Not such a dismal science: the economics of protein synthesis, folding, degradation and antigen processing. *Trends Cell Biol*. 2001;11:294-297.
133. Han KK, Martinage A. Post-translational chemical modification(s) of proteins. *Int J Biochem*. 1992;24:19-28.
134. Saibil H. Molecular chaperones: containers and surfaces for folding, stabilising or unfolding proteins. *Curr Opin Struct Biol*. 2000; 10:251-258.
135. Johnson ES, Bartel B, Seufert W, Varshavsky A. Ubiquitin as a degradation signal. *EMBO J*. 1992;11:497-505.
136. Hochstrasser M. Ubiquitin-dependent protein degradation. *Annu Rev Genet*. 1996;30:405-439.
137. Laney JD, Hochstrasser M. Substrate targeting in the ubiquitin system. *Cell*. 1999;97:427-430.
138. Ciechanover A. The ubiquitin-proteasome proteolytic pathway. *Cell*. 1994;79:13-21.
139. Lo RS, Massague J. Ubiquitin-dependent degradation of TGF-beta-activated smad2. *Nat Cell Biol*. 1999;1:472-478.
140. Jesenberger V, Jentsch S. Deadly encounter: ubiquitin meets apoptosis. *Nat Rev Mol Cell Biol*. 2002;3:112-121.
141. Rosenthal N. Tools of the trade—recombinant DNA. *N Engl J Med*. 1994;331:315-317.
142. Morrow JF. Recombinant DNA techniques. *Methods Enzymol*. 1979;68:3-24.
143. Emery AE. Recombinant DNA technology. *Lancet*. 1981;2:1406-1409.
144. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*. 1986;51(pt 1):263-273.
145. Eisenstein BI. The polymerase chain reaction: a new method of using molecular genetics for medical diagnosis. *N Engl J Med*. 1990;322:178-183.
146. Saiki RK, Scharf S, Faloona F, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985;230:1350-1354.
147. Erlich HA. Principles and applications of the polymerase chain reaction. *Rev Immunogenet*. 1999;1:127-134.
148. Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. *Science*. 1991;252:1643-1651.
149. Baumforth KR, Nelson PN, Digby JE, O'Neil JD, Murray PG. Demystified...the polymerase chain reaction. *Mol Pathol*. 1999; 52:1-10.
150. Dieffenbach CW, Lowe TM, Dveksler GS. General concepts for PCR primer design. *PCR Methods Appl*. 1993;3:S30-S37.
151. Lowe T, Sharefkin J, Yang SQ, Dieffenbach CW. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res*. 1990;18:1757-1761.
152. Rychlik W. Selection of primers for polymerase chain reaction. *Mol Biotechnol*. 1995;3:129-134.
153. Brock TD, Freeze H. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol*. 1969;98:289-297.
154. Chien A, Edgar DB, Trela JM. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol*. 1976;127:1550-1557.
155. Sato S, Hutchinson CA III, Harris JI. A thermostable sequence-specific endonuclease from *Thermus aquaticus*. *Proc Natl Acad Sci U S A*. 1977;74:542-546.
156. Aurelius E, Johansson B, Skoldenberg B, Staland A, Forsgren M. Rapid diagnosis of herpes simplex encephalitis by nested polymerase chain reaction assay of cerebrospinal fluid. *Lancet*. 1991; 337:189-192.
157. Burgart LJ, Robinson RA, Heller MJ, Wilke WW, Iakubova OK, Chevillat JC. Multiplex polymerase chain reaction. *Mod Pathol*. 1992;5:320-323.
158. Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. *J Clin Lab Anal*. 2002;16:47-51.
159. Mahony JB. Multiplex polymerase chain reaction for the diagnosis of sexually transmitted diseases. *Clin Lab Med*. 1996;16:61-71.
160. Na BK, Kim JH, Shin GC, et al. Detection and typing of respiratory adenoviruses in a single-tube multiplex polymerase chain reaction. *J Med Virol*. 2002;66:512-517.
161. Bustin SA, Dorudi S. Molecular assessment of tumour stage and disease recurrence using PCR-based assays. *Mol Med Today*. 1998;4:389-396.
162. Biondi A, Rambaldi A. Polymerase chain reaction (PCR) approach for the evaluation of minimal residual disease in acute leukemia. *Stem Cells*. 1994;12:394-401.
163. Raj GV, Moreno JG, Gomella LG. Utilization of polymerase chain reaction technology in the detection of solid tumors. *Cancer*. 1998;82:1419-1442.
164. Jackson JB. The polymerase chain reaction in transfusion medicine. *Transfusion*. 1990;30:51-57.
165. Westwood SA, Werrett DJ. An evaluation of the polymerase chain reaction method for forensic applications. *Forensic Sci Int*. 1990; 45:201-215.
166. Decorte R, Cassiman JJ. Forensic medicine and the polymerase chain reaction technique. *J Med Genet*. 1993;30:625-633.
167. Hageberg E, Bell LS, Allen T, Boyde A, Jones SJ, Clegg JB. Analysis of ancient bone DNA: techniques and applications. *Philos Trans R Soc Lond B Biol Sci*. 1991;333:399-407.
168. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. Neandertal DNA sequences and the origin of modern humans. *Cell*. 1997;90:19-30.
169. Lo YM, Mehal WZ, Fleming KA. False-positive results and the polymerase chain reaction [letter]. *Lancet*. 1988;2:679.
170. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*. 1987;162:156-159.
171. Mies C. A simple, rapid method for isolating RNA from paraffin-embedded tissues for reverse transcription-polymerase chain reaction (RT-PCR). *J Histochem Cytochem*. 1994;42:811-813.
172. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*. 1977;74:5350-5354.
173. Fernyhough P. Quantification of mRNA levels using Northern blotting. *Methods Mol Biol*. 2001;169:53-63.
174. Tanay VA, Tancowny BP, Glencorse TA, Bateson AN. The quantitative analysis of multiple mRNA species using oligonucleotide probes in an S1 nuclease protection assay. *Mol Biotechnol*. 1997; 7:217-229.
175. Rappolee DA, Wang A, Mark D, Werb Z. Novel method for studying mRNA phenotypes in single or small numbers of cells. *J Cell Biochem*. 1989;39:1-11.
176. Becker-Andre M, Hahlbrock K. Absolute mRNA quantification using the polymerase chain reaction (PCR): a novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res*. 1989;17:9437-9446.

177. Hamoui S, Benedetto JP, Garret M, Bonnet J. Quantitation of mRNA species by RT-PCR on total mRNA population. *PCR Methods Appl.* 1994;4:160-166.
178. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol.* 2000;25:169-193.
179. Brooks EM, Sheflin LG, Spaulding SW. Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR. *Biotechniques.* 1995;19:806-812, 814-815.
180. Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. *Biotechniques.* 1999;26:112-122, 124-125.
181. Gattei V, Degan M, De Iulius A, Rossi FM, Aldinucci D, Pinto A. Competitive reverse-transcriptase PCR: a useful alternative to Northern blotting for quantitative estimation of relative abundances of specific mRNAs in precious samples. *Biochem J.* 1997;325:565-567.
182. Gilliland G, Perrin S, Blanchard K, Bunn HF. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc Natl Acad Sci U S A.* 1990;87:2725-2729.
183. Deguchi T, Yang M, Kawada Y. Micrometastasis of prostate cancer to lymph nodes: detection by means of reverse transcription-polymerase chain reaction. *J Natl Cancer Inst.* 1997;89:1471-1473.
184. Ballestrero A, Coviello DA, Garuti A, et al. Reverse-transcriptase polymerase chain reaction of the maspin gene in the detection of bone marrow breast carcinoma cell contamination. *Cancer.* 2001;92:2030-2035.
185. Hauptli D, Bruckner L, Ottiger HP. Use of reverse transcriptase polymerase chain reaction for detection of vaccine contamination by avian leukosis virus. *J Virol Methods.* 1997;66:71-81.
186. Diverio D, Pandolfi PP, Rossi V, Biondi A, Pelicci PG, Lo Coco F. Monitoring of treatment outcome in acute promyelocytic leukemia by RT-PCR. *Leukemia.* 1994;8(suppl 2):S63-S65.
187. Muto A, Mori S, Matsushita H, et al. Serial quantification of minimal residual disease of t(8;21) acute myelogenous leukaemia with RT-competitive PCR assay. *Br J Haematol.* 1996;95:85-94.
188. Gibson UE, Heid CA, Williams PM. A novel method for real time quantitative RT-PCR. *Genome Res.* 1996;6:995-1001.
189. Giulietti A, Overbergh L, Valckx D, Decallonne B, Bouillon R, Mathieu C. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods.* 2001;25:386-401.
190. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y).* 1993;11:1026-1030.
191. Oldach D. "Real-time" polymerase chain reaction. *Gastroenterology.* 1999;116:763-764.
192. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res.* 1996;6:986-994.
193. Simpson DA, Feeney S, Boyle C, Stitt AW. Retinal VEGF mRNA measured by SYBR green I fluorescence: a versatile approach to quantitative PCR. *Mol Vis.* 2000;6:178-183.
194. Cardullo RA, Agrawal S, Flores C, Zamecnik PC, Wolf DE. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A.* 1988;85:8790-8794.
195. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A.* 1991;88:7276-7280.
196. King HC, Sinha AA. Gene expression profile analysis by DNA microarrays: promise and pitfalls. *JAMA.* 2001;286:2280-2288.
197. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol.* 1996;14:303-308.
198. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol.* 1999;17:804-807.
199. Aldape K, Ginzinger DG, Godfrey TE. Real-time quantitative polymerase chain reaction: a potential tool for genetic analysis in neuropathology. *Brain Pathol.* 2002;12:54-66.
200. Satou J, Funato T, Satoh N, et al. Quantitative PCR determination of human cytomegalovirus in blood cells. *J Clin Lab Anal.* 2001;15:122-126.
201. Kaiser K, Rabodonirina M, Picot S. Real time quantitative PCR and RT-PCR for analysis of *Pneumocystis carinii hominis*. *J Microbiol Methods.* 2001;45:113-118.
202. Lovatt A. Applications of quantitative PCR in the biosafety and genetic stability assessment of biotechnology products. *J Biotechnol.* 2002;82:279-300.
203. Blakely WF, Prasanna PG, Grace MB, Miller AC. Radiation exposure assessment using cytological and molecular biomarkers. *Radiat Prot Dosimetry.* 2001;97:17-23.
204. Lyon E. Mutation detection using fluorescent hybridization probes and melting curve analysis. *Expert Rev Mol Diagn.* 2001;1:92-101.
205. Tung CH, Mahmood U, Bredow S, Weissleder R. In vivo imaging of proteolytic enzyme activity using a novel molecular reporter. *Cancer Res.* 2000;60:4953-4958.
206. Latner AL. Isoelectric focusing in liquid and gels as applied to clinical chemistry. *Adv Clin Chem.* 1975;17:193-250.
207. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol.* 1975;98:503-517.
208. Cantor CR, Smith CL, Mathew MK. Pulsed-field gel electrophoresis of very large DNA molecules. *Annu Rev Biophys Biochem.* 1988;17:287-304.
209. Maule J. Pulsed-field gel electrophoresis. *Mol Biotechnol.* 1998;9:107-126.
210. Egger D, Bienz K. Protein (Western) blotting. *Mol Biotechnol.* 1994;1:289-305.
211. Fido RJ, Tatham AS, Shewry PR. Western blotting analysis. *Methods Mol Biol.* 1995;49:423-437.
212. Lilley KS, Razzaq A, Dupree P. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol.* 2002;6:46-50.
213. Heitman J. How the EcoRI endonuclease recognizes and cleaves DNA. *Bioessays.* 1992;14:445-454.
214. Clark D, Henikoff S. Ordered deletions using exonuclease III. *Methods Mol Biol.* 1996;57:139-147.
215. Cai H, Yu H, McEntee K, Kunkel TA, Goodman MF. Purification and properties of wild-type and exonuclease-deficient DNA polymerase II from *Escherichia coli*. *J Biol Chem.* 1995;270:15327-15335.
216. Joyce CM, Grindley ND. Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1983;80:1830-1834.
217. Steitz TA, Beese L, Freemont PS, Friedman JM, Sanderson MR. Structural studies of Klenow fragment: an enzyme with two active sites. *Cold Spring Harb Symp Quant Biol.* 1987;52:465-471.
218. Lehman IR. DNA ligase: structure, mechanism, and function. *Science.* 1974;186:790-797.
219. Giacomone G. DNA topoisomerases and topoisomerase inhibitors. *Pathol Biol (Paris).* 1994;42:346-352.
220. Osheroff N, Zechiedrich EL, Gale KC. Catalytic function of DNA topoisomerase II. *Bioessays.* 1991;13:269-273.
221. Cozzarelli NR. DNA gyrase and the supercoiling of DNA. *Science.* 1980;207:953-960.
222. Sinsheimer RL. Recombinant DNA. *Annu Rev Biochem.* 1977;46:415-438.
223. Jakobovits A. YAC vectors: humanizing the mouse genome. *Curr Biol.* 1994;4:761-763.
224. Minton NP, Oultram JD. Host: vector systems for gene cloning in *Clostridium*. *Microbiol Sci.* 1988;5:310-315.
225. Kim UJ, Shizuya H, Kang HL, et al. A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc Natl Acad Sci U S A.* 1996;93:6297-6301.

226. Moir DT, Dorman TE, Smyth AP, Smith DR. A human genome YAC library in a selectable high-copy-number vector. *Gene*. 1993;125:229-232.
227. Chumakov IM, Rigault P, Le Gall I, et al. A YAC contig map of the human genome. *Nature*. 1995;377(6547 suppl):175-297.
228. Kao FT, Hartz JA, Law ML, Davidson JN. Isolation and chromosomal localization of unique DNA sequences from a human genomic library. *Proc Natl Acad Sci U S A*. 1982;79:865-869.
229. Christensen AC. Bacteriophage lambda-based expression vectors. *Mol Biotechnol*. 2001;17:219-224.
230. Wolfsberg TG, Landsman D. Expressed sequence tags (ESTs). *Methods Biochem Anal*. 2001;43:283-301.
231. Nakamura Y, Leppert M, O'Connell P, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*. 1987;235:1616-1622.
232. Jeffreys AJ, Wilson V, Wong Z, et al. Highly variable minisatellites and DNA fingerprints. *Biochem Soc Symp*. 1987;53:165-180.
233. Wang Z, Moulton J. SNPs, protein structure, and disease. *Hum Mutat*. 2001;17:263-270.
234. Gura T. Genetics: can SNPs deliver on susceptibility genes? *Science*. 2001;293:593-595.
235. Graham CA, Hill AJ. Introduction to DNA sequencing. *Methods Mol Biol*. 2001;167:1-12.
236. Griffin HG, Griffin AM. DNA sequencing. *Methods Mol Biol*. 1993;23:1-8.
237. Marziali A, Akeson M. New DNA sequencing methods. *Annu Rev Biomed Eng*. 2001;3:195-223.
238. Smith MJ, Goodfellow PN. Gene mapping and genetic diseases. *Curr Opin Cell Biol*. 1989;1:460-465.
239. Human gene mapping, genetic linkage, and clinical applications. *Ann Intern Med*. 1980;93:469-479.
240. Morgan TH. Chromosomes and heredity. *Am Nat*. 1910;44:449-496.
241. Korenberg JR, Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell*. 1988;53:391-400.
242. Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. *J Natl Cancer Inst*. 1960;25:85.
243. Dewald GW, Wyatt WA, Juneau AL, et al. Highly sensitive fluorescence in situ hybridization method to detect double BCR/ABL fusion and monitor response to therapy in chronic myeloid leukemia. *Blood*. 1998;91:3357-3365.
244. Ruddle FH, Bootsma D, Stefani M, et al. Workshop on mapping by somatic cell hybridization. *Prog Clin Biol Res*. 1982;103(pt A):145-153.
245. Tischfield JA, Ruddle FH. Assignment of the gene for adenine phosphoribosyltransferase to human chromosome 16 by mouse-human somatic cell hybridization. *Proc Natl Acad Sci U S A*. 1974;71:45-49.
246. Deisseroth A, Nienhuis A, Turner P, et al. Localization of the human alpha-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell*. 1977;12:205-218.
247. Cox DR, Burmeister M, Price ER, Kim S, Myers RM. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*. 1990;250:245-250.
248. Watson G, See CG, Woo P. Use of somatic cell hybrids and fluorescence in situ hybridization to localize the functional serum amyloid A (SAA) genes to chromosome 11p15.4-p15.1 and the entire SAA superfamily to chromosome 11p15. *Genomics*. 1994;23:694-696.
249. Korenberg JR, Chen XN, Adams MD, Venter JC. Toward a cDNA map of the human genome. *Genomics*. 1995;29:364-370.
250. Raymond A, Friedli M, Henrichsen CN, et al. From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics*. 2001;78:46-54.
251. Vissing H, Grosveld F, Solomon E, et al. Progress towards construction of a total restriction fragment map of a human chromosome. *Nucleic Acids Res*. 1987;15:1363-1375.
252. Sturtevant AH. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Biol*. 1913;14:43-59.
253. Putrament A. Recombination and chromosome structure in eukaryotes. *Genet Res*. 1971;18:85-95.
254. Bowcock AM, Hebert JM, Wijisman E, Gadi I, Cavalli-Sforza LL, Boyd CD. High recombination between two physically close human basement membrane collagen genes at the distal end of chromosome 13q. *Proc Natl Acad Sci U S A*. 1988;85:2701-2705.
255. Smith KN, Nicolas A. Recombination at work for meiosis. *Curr Opin Genet Dev*. 1998;8:200-211.
256. Lynn A, Kashuk C, Petersen MB, et al. Patterns of meiotic recombination on the long arm of human chromosome 21. *Genome Res*. 2000;10:1319-1332.
257. Tease C. Cytological detection of crossing-over in BUdR substituted meiotic chromosomes using the fluorescent plus Giemsa technique. *Nature*. 1978;272:823-824.
258. Morton NE. Sequential tests for the detection of linkage. *Hum Opin Genet Dev*. 1955;7:277-318.
259. Robson EB. The human gene map. *Philos Trans R Soc Lond B Biol Sci*. 1988;319:229-237.
260. Pulst SM. Genetic linkage analysis. *Arch Neurol*. 1999;56:667-672.
261. White R, Lalouel JM, Leppert M, Lathrop M, Nakamura Y, O'Connell P. Linkage maps of human chromosomes. *Genome*. 1989;31:1066-1072.
262. Collins FS. Of needles and haystacks: finding human disease genes by positional cloning. *Clin Res*. 1991;39:615-623.
263. Xu L, Li Y. Positional cloning. *Methods Mol Biol*. 2000;136:285-296.
264. Ruddle FH. A new era in mammalian gene mapping: somatic cell genetics and recombinant DNA methodologies. *Nature*. 1981;294:115-120.
265. Parimoo S, Patanjali SR, Kolluri R, Xu H, Wei H, Weissman SM. cDNA selection and other approaches in positional cloning. *Anal Biochem*. 1995;228:1-17.
266. Korenberg JR, Chen XN, Sun Z, et al. Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. *Genome Res*. 1999;9:994-1001.
267. Spoerel NA, Kafatos FC. Isolation of full-length genes: walking the chromosome. *Methods Enzymol*. 1987;152:598-603.
268. Rubio JP, Triglia T, Kemp DJ, de Bruin D, Ravetch JV, Cowman AF. A YAC contig map of *Plasmodium falciparum* chromosome 4: characterization of a DNA amplification between two recently separated isolates. *Genomics*. 1995;26:192-198.
269. Lu-Kuo JM, Le Paslier D, Weissenbach J, Chumakov I, Cohen D, Ward DC. Construction of a YAC contig and a STS map spanning at least seven megabasepairs in chromosome 5q34-35. *Hum Mol Genet*. 1994;3:99-106.
270. Lewin B. *Genes VII*. Oxford, England: Oxford University Press; 2000.

Primer on Medical Genomics Part III will appear in the September issue.

Glossary of Terms Frequently Used in Medical Genomics

Allele—One of 2 or more alternative forms of a DNA sequence; eg, most persons have 2 functional (normal) alleles of the phenylalanine hydroxylase gene, whereas carriers of phenylketonuria have 1 functional allele and 1 mutant allele.

Alternative splicing—Process by which different messenger RNAs (mRNAs) are produced from the same primary transcript, through variations in the splicing pattern of the transcript.

Alu element—Short repetitive DNA sequence that is scattered throughout the genome and is identified by a restriction enzyme (Alu I endonuclease) recognition site near the middle of the sequence.

Autosomes—Chromosomes not involved with sex differentiation (1-22 in humans).

Bacteriophage—Virus that infects bacteria, commonly known as a phage.

Gene bank—Collection of recombinant DNA molecules containing inserts that together comprise the entire genome of an organism.

CentiMorgan (cM)—Distance between DNA loci as determined on a genetic map. One centiMorgan distance indicates 2 markers are inherited separately 1% of the time. For the physical map, 1 cM is approximately equal to 1 million bases.

Centromere—Heterochromatic-constricted portion of a chromosome where the chromatids are joined.

Chromosome—Self-replicating nucleic acid molecule containing a number of genes. In bacteria, the entire genome is contained within 1 double-stranded (ds), circular DNA chromosome. In eukaryotes, chromosomes are linear DNA duplexes, and most organisms have genomes divided between a number of such chromosomes, that number being characteristic for a particular species.

Chromosome walking—Method used to identify which clone in a gene bank contains a desired gene or sequence that cannot be selected for easily. The gene bank must contain the entire DNA sequence of the chromosome as a series of overlapping fragments. These fragments can be generated either by random shear or by partial digestion with a 4-base pair [bp] cutter such as Sau 3A. A series of colony hybridizations is then carried out, starting with same cloned gene that has already been identified and that is known to be on the same chromosome as the desired gene. This identified gene is used as a probe to pick out clones containing adjacent sequences; these are then used as probes themselves to identify clones carrying sequences adjacent to them, etc. At each round of hybridization one “walks” further along the chromosome from the identified gene.

Clone—Colony of cells, all descended from a single ancestral cell, therefore genetically homogeneous; also the act of preparing and propagating a genetically defined cell. This term is used in a number of senses. As a noun it may mean (1) a population of recombinant DNA molecules all carrying the same inserted sequence or (2) a population of cells or organisms of identical genotype. It is most frequently used to describe a colony of microorganisms that harbors a specific DNA fragment inserted into a vector molecule. As a verb, “to clone” means to use in vitro recombination techniques to insert a particular gene or other DNA sequence into a vector molecule.

Coding sequence—Portion of a gene that directly specifies the amino acid sequence of its protein product. Noncoding sequences of genes include control regions, such as promoters, operators, and terminators, as well as the intron sequences of certain eukaryotic genes.

Complementary—Nucleic acid strand is said to be complementary to another strand if it is able to form a perfectly hydrogen-bonded duplex with it, according to the Watson-Crick rules of base pairing, which include the antiparallel alignment of the strands 3' to 5' at each end of the duplex. Example: An mRNA molecule is complementary to one of the DNA strands of the gene that encodes it. The RNA sequence 5'-AUGGCAUUUCGGCCCCACUGA-3' is complementary to the DNA sequence 3'-TACCGTAAAGCCGGGGTGACT-5'.

Complementary DNA (cDNA)—DNA complement of an RNA sequence; it is synthesized by the enzyme RNA-primed DNA polymerase of reverse transcriptase. The single-stranded DNA product of this enzyme (the reverse transcript) may be converted into the double-stranded form by DNA-primed DNA polymerase (DNA polymerase) and inserted into a suitable vector to make a cDNA clone; cDNA cloning is commonly used to achieve expression of mammalian genes in bacteria or yeast. Since the cDNA clone is a copy of the mature mRNA molecule, it contains no introns to act as a barrier to expression.

Digestion—Treatment of a substrate molecule with an enzyme preparation in which covalent bonds are hydrolyzed.

DNA fingerprinting—Method to generate a pattern of DNA restriction fragments that are unique to an individual, especially for forensic purposes. DNA from blood, semen, or other tissue sample is isolated and cleaved with a restriction endonuclease, the products being separated by polyacrylamide gel electrophoresis, and fragments containing a specific nucleotide sequence are detected by hybridization with an appropriate DNA probe to produce the unique pattern.

DNA gyrase—A type II topoisomerase of *Escherichia coli*, which is essential for DNA replication. This enzyme can induce or relax supercoiling.

DNA polymerase—Any enzyme that synthesizes DNA by copying a template strand. DNA polymerases synthesize DNA in the 5' to 3' direction by successively adding nucleotides to the free 3' hydroxyl group of the growing strand. The template strand determines the order of addition of nucleotides via Watson-Crick base pairing. *E coli* has 3 DNA polymerases: polymerase I is involved in repair synthesis, polymerase III is responsible for DNA replication, and the function of polymerase II is unknown at present (see DNA polymerase I, nick translation, Okazaki fragment). Animal cells have α , β , γ , and δ polymerases, with α and δ apparently responsible for replication of nuclear DNA and γ for replication of mitochondria. All these function with a DNA strand as a template. Retroviruses possess a unique DNA polymerase (reverse transcriptase) that uses an RNA template.

DNA topoisomerase—Enzyme capable of altering the linking number of DNA, usually measured by following a change in the degree of supercoiling of dsDNA molecules. Various topoisomerases can increase or relax supercoiling, convert single-stranded rings to intertwined double-stranded rings, tie and untie knots in single-stranded and duplex rings, and catenate and decatenate duplex rings. Topoisomerase II of *E coli* = gyrase.

Endonuclease—One of a large group of enzymes that cleave nucleic acids at positions within the chain. Some act on both RNA and DNA (eg, S1 nuclease, EC 3.1.30.1, which is specific for single-stranded molecules). Ribonucleases such as pancreatic, T1, etc, are specific for RNA, deoxyribonucleases for DNA. Bacterial restriction endonucleases are crucial in recombinant DNA technology for their ability to cleave dsDNA at highly specific sites.

Enhancer—Sequence of DNA that increases the rate of transcription of adjacent sequences; it may be located at varying distances and oriented upstream from, downstream from, or within a structural gene (see promoter).

Euchromatin—In cytology, the lightly staining regions of chromosomes that contain less-condensed chromatin and that are the regions transcribed as RNA (see heterochromatin).

Exon—From “expressed,” a region of a eukaryotic gene that encodes protein, as opposed to an intron, ie, a region of the gene that does not.

Expressed sequence tags (ESTs)—Known DNA sequences from gene segments expressed as RNA that can be used to clone and characterize these genes.

Fluorescence in situ hybridization (FISH)—Nonisotopic method to label DNA probes for in situ hybridization. The ability to use multiple fluorochromes in the same reaction increases the utility of this procedure. The resolving power of FISH is further enhanced if interphase chromosomes are studied.

Gene bank—Collection of cloned DNA fragments from a single genome. Ideally, the bank should contain cloned representatives of all the DNA sequences in the genome. For large genomes, such banks are made by shotgun cloning.

Genetic map—Indirect measure of distance, constructed by determining how frequently 2 markers (DNA polymorphisms, physical traits, or syndromes) are inherited together. Distances in genetic maps are measured in cM (see physical map).

Genomic library—Collection of transformed cells, each of which contains DNA fragments, the entire population representing the total genome of an organism, eg, a rat library containing fragments that together comprise the entire rat genome. Appropriate screening methods select a single transformed cell containing a specific gene.

Heterochromatin—Condensed chromatin in certain chromosome regions that often exhibits individual variability and is not thought to contain actively transcribed genes.

Heterozygote—Individual with 2 different alleles (eg, gene, polymorphic marker) at a single locus.

Histones—Proteins found in the nuclei of all eukaryotic cells where they are complexed to DNA in chromatin and chromosomes. They are of relatively low molecular weight and are basic, having a very high arginine-lysine content; highly conserved and can be grouped into 5 major classes. Two copies of H2A, H2B, H3, and H4 bind to about 200 bps of DNA to form the repeating structure of chromatin, the nucleosome, with H1 binding to the linker sequence. Histones may act as nonspecific repressors of gene transcription.

Homologous chromosomes—Chromosomes that pair with each other at meiosis or chromosomes in different species that have retained most of the same genes during their evolution from a common ancestor.

Homozygote—Individual with 2 identical alleles (eg, gene, polymorphism) at a single locus.

Intron—Sequence within a eukaryotic gene that is not represented in the protein product sequence of that gene. Intron sequences are transcribed into RNA and must be excised, and the RNA molecule relegated before it can be translated, a process known as intron splicing. Some genes of higher eukaryotes contain a large number of introns that make up the bulk of the DNA sequence of the gene. Introns are also found in genes whose RNA transcripts are not translated, ie, eukaryotic ribosomal RNA (rRNA) and transfer RNA (tRNA) genes. In these cases the intron sequence does not appear in the functional RNA molecule.

Inverted repeat—Two regions of a nucleic acid molecule that have the same nucleotide sequence but in opposite orientation to one another (compare, direct repeat), eg, GCACTTG..... GTTCACG CGTGAAC.....CAAGTGC.

Karyotype—Individual's or cell's chromosomal constitution (number, size, and morphology), determined by examination of chromosomes with light microscopy and use of stains.

Klenow fragment—Product of limited proteolysis of *E coli* DNA polymerase I that retains the polymerase and 3' to 5' exonuclease (proofreading) activity but is missing the 5' to 3' nuclease activity.

Ligase—One of a class of enzymes that joins 2 substrate molecules in energy-dependent, usually adenosine triphosphate (ATP)-dependent reactions, eg, an amino acid-activating enzyme, a carboxylase (see synthetase). It seals the nicks in DNA molecules by forming a phosphodiester bond between adjacent nucleotides, which have a free 5'-phosphate and 3'-hydroxyl group. The ligase enzyme encoded by phage T4 is usually used in gene cloning experiments; it requires ATP as a cofactor. T4 ligase is used in vitro to join vector and insert DNA.

Linkage—Tendency to inherit together 2 or more nonallelic genes or DNA markers than is to be expected by independent assortment. Genes-DNA markers are linked because they are sufficiently close to each other on the same chromosome.

Messenger RNA (mRNA)—Single-stranded RNA molecule that specifies the amino acid sequence of 1 or more polypeptide chains; information is translated during protein synthesis when ribosomes bind to the mRNA. In prokaryotes mRNA is the primary transcript from DNA sequence, and protein synthesis starts while the mRNA is still being synthesized. Prokaryote mRNAs are usually very short lived (average half-life, 5 minutes). In contrast, in eukaryotes the primary transcripts (heterogeneous nuclear RNA [hnRNA]) are synthesized in the nucleus, and they are extensively processed to give the mRNA which is exported to the cytoplasm where protein synthesis takes place. This processing includes the addition of a 5'-5'-linked 7-methylguanylate “cap” at the 5' end and a sequence of adenylate groups at the 3' end (the poly-A tail), as well as the removal of any introns and the splicing together of exons. Only 10% of hnRNA leaves the nucleus. Eukaryote mRNAs are comparatively long lived, with half-life ranging from 30 minutes to 24 hours.

Microsatellites—Same as for minisatellites except that the polymorphism allele size is smaller, eg, <1 kilobase, and the basic core repeat unit involves a 2- to 4-nucleotide bp repeat motif; also known as SSRs (simple sequence repeats). One example is repeats of the motif CA, eg, CACACACACA, which is also written as AC (ACACACACAC, etc). Terminology is confusing because these repeats are identical.

Minisatellites—Repetitive DNA consisting of small numbers of repeating nucleotides in a tandem configuration, used for genetic linkage and DNA fingerprinting studies.

Northern blotting—Procedure analogous to Southern blotting in which RNA, not DNA, is transferred or blotted from a gel to a suitable binding matrix, such as a nitrocellulose sheet.

Nucleic acid—DNA or RNA molecule that can be single stranded or double stranded. When very small, it is an oligonucleotide.

Nucleotide—Repeating structural unit of chromatin that consists of a complex of 8 molecules of histones and a DNA double helix wrapped twice around them.

Okazaki fragment—Short fragments of newly synthesized DNA strands produced during DNA replication. All the known DNA polymerases can only synthesize DNA in 1 direction, the 5' to 3'

direction. However, as the strands separate, replication forks will be moving along 1 parental strand in the 3' to 5' direction and 5' to 3' on the other parental strand. On the former, the leading strand, DNA can be synthesized continuously in the 5' to 3' direction. On the other, the lagging strand, the DNA synthesis can occur only when a stretch of single-stranded DNA has been exposed and proceeds in the direction opposite to the movement of the replication fork (still 5' to 3'). Thus, it is discontinuous, and the series of fragments are then covalently linked by ligases to give a continuous strand. Such fragments were first observed by Okazaki using pulse-chase labeling with radioactive thymidine. In eukaryotes, Okazaki fragments are typically a few hundred nucleotides long, whereas in prokaryotes they may contain several thousands of nucleotides.

Oligonucleotides—Small single-stranded segments of DNA, typically 20 to 30 nucleotide bases, that are synthesized in vitro. Uses include DNA sequencing, DNA amplification, and DNA probes.

Open reading frame (ORF)—A reading frame uninterrupted by stop codons. The existence of ORFs is usually inferred from the DNA (rather than the RNA) sequence.

Polyacrylamide gel electrophoresis (PAGE)—Method for separating nucleic acid or protein molecules based on their molecular size. Molecules migrate through the inert gel matrix under the influence of an electric field. For protein PAGE, detergents such as sodium dodecyl sulfate are often added to ensure that all molecules have a uniform charge. Secondary structure often leads to anomalous migration of molecules. Thus, it is common to denature protein samples by boiling them before PAGE. For nucleic acids, denaturing agents such as formamide, urea, or methyl mercuric hydroxide are often incorporated into the gel itself, which may also be run at high temperature. PAGE is used to separate the products of DNA-sequencing reactions, and the gels used are highly denaturing because molecules differing in size by a single nucleotide must be resolved.

Polymerase—Enzyme that will form oligomeric molecules from monomers. A DNA polymerase will synthesize DNA from deoxynucleoside triphosphates by using a cDNA strand and a primer. An RNA polymerase will synthesize RNA from monoribonucleoside triphosphates and a cDNA strand.

Polymerase chain reaction (PCR)—Technique to amplify a specific region of dsDNA. An excess of 2 oligonucleotide primers complementary to 2 sequences that flank the region to be amplified are annealed to denatured DNA and subsequently elongated, usually by a heat-stable DNA polymerase from *Thermus aquaticus* (Taq polymerase). Each cycle involves heating to denature dsDNA and cooling to allow annealing of excess primer to template and elongation of the primers by the Taq polymerase; the number of target sequence fragments between flanking primers doubles with each cycle.

Primary structure—Amino acid sequence of a protein or the nucleotide sequence of a polynucleotide.

Primer—DNA polymerase, unlike RNA polymerase, is unable to initiate the de novo synthesis of a polynucleotide chain; it can only add nucleotides to a free 3'-hydroxyl group at the end of a preexisting chain. A short oligonucleotide, known as a primer, is needed to supply such a hydroxyl group for initiation of DNA synthesis; RNA primer is used for initiation of DNA replication in vivo.

Probe—As a noun, a probe is a specific DNA or RNA sequence that has been radioactively labeled to a high specific activity. Probes are used to detect complementary sequences by hybridiza-

tion techniques such as Southern or Northern blotting or colony hybridization. As a verb, "to probe" is the act of hybridization to detect a specific gene or transcript, eg, "We probed our bank with labeled rRNA to detect clones containing rDNA sequences."

Processing—Posttranscriptional modification of RNA or post-translational modification of a protein.

Promoter—DNA region, usually upstream to the coding sequence of a gene or operon, which binds RNA polymerase and directs the enzyme to the correct transcriptional start site.

Quaternary structure—Arrangement in space of polypeptide subunits that make up a multimeric protein.

Recombination—Crossing over (breakage and rejoining) between 2 loci, resulting in new combinations of genetic markers—traits at those loci, eg, imagine that 1 locus has 4 genetic markers linearly arranged, a-b-c-d, and the second locus is b-b-c-a. Recombination involving these 2 regions between the b-c markers would give new genetic combinations, ie, a-b-c-a and b-b-c-d.

Recombination frequency—Number of recombinants divided by the total number of progeny. This frequency is used as a guide in assessing the relative distances between loci on a genetic map.

Restriction endonucleases—Class of bacterial enzymes that cut DNA at specific sites. In bacteria their function is to destroy foreign DNA, such as that of bacteriophages (host DNA is specifically modified by methylation at these sites). They have very specific recognition and cutting sites. The recognition sites are short, 4 to 8 nucleotides, and are usually palindromic sequences. Because both strands have the same sequence running in opposite directions, the enzymes make double-stranded breaks, which, if the site of cleavage is off-center, generate fragments with short single-stranded tails; these can hybridize to the tails of other fragments and are called "sticky ends." They are generally named according to the bacterium from which they were isolated (the first letter of genus name and the first 2 letters of the specific name). The bacterial strain is identified, and multiple enzymes from the same strain are given Roman numerals. For example, the 2 enzymes isolated from the R strain of *E coli* are designated Eco RI and Eco RII.

Restriction fragment length polymorphism (RFLP)—DNA polymorphisms deriving from changes in nucleotides at restriction sites, yielding DNA segments of different lengths after restriction and gel electrophoresis.

Reverse transcriptase—RNA-directed DNA polymerase; enzyme first discovered in Retroviridae, which can construct dsDNA molecules from the single-stranded RNA templates of their genomes. Can be inhibited by the drug zidovudine. Reverse transcription now appears also to be involved in movement of certain mobile genetic elements such as the Ty plasmid in yeast, in the replication of other viruses such as hepatitis B, and possibly in the generation of mammalian pseudogenes.

Ribonucleases—Phosphodiesterases that degrade RNA, very widely distributed.

Ribonucleoprotein (RNP)—Complexes of RNA and protein involved in a wide range of cellular processes. Besides ribosomes (with which RNP was originally almost synonymous), in eukaryotic cells both initial RNA transcripts in the nucleolus (hnRNA) and cytoplasmic mRNAs exist as complexes with specific sets of proteins. Processing (splicing) of the former is carried out by small nuclear RNPs. Other examples are the signal recognition particle responsible for targeting proteins to endoplasmic reticulum and a complex involved in termination of transcription.

Ribosomal protein—Protein present within the ribosomal subunits. In prokaryotes there are 31 to 34 proteins in the large subunit and 21 in the small subunit. Eukaryotic subunits have 45 to 50 (large subunit) and 33 (small subunits) proteins.

Ribosome—Subcellular complex responsible for protein synthesis. It consists of 2 subunits that both contain 1 or more RNA molecules (rRNAs) and a large number of proteins. The small subunit is responsible for binding mRNA and is subsequently joined by the large subunit, which accepts the aminoacyl tRNA molecule and carries out the process of peptide bond formation.

RNA polymerase—Enzyme that transcribes DNA into RNA; it is able to initiate RNA synthesis on a DNA template in the absence of any primer molecule. Prokaryotes have a single RNA polymerase that synthesizes all classes of RNA molecules. Eukaryotes have 3 RNA polymerases with different transcriptional specificities: RNA polymerase A (or I) synthesizes the large rRNA precursor, B (or II) synthesizes mRNA, C (or III) synthesizes tRNA and 5s rRNA species.

Satellite chromosome—Chromosome that seems to be an addition to the normal genome.

Satellite DNA—Polynucleotides that are separable (based on their characteristic density) from the bulk of nuclear DNA and that have repetitive sequences.

Sodium dodecyl sulfate (SDS) gel electrophoresis—Technique to dissociate multimeric proteins and to separate all proteins according to their apparent molecular weights; proteins are treated with SDS, an anionic detergent, and are separated by electrophoresis on polyacrylamide gels that incorporate SDS, the detergent adsorbing to the protein in proportion to its mass and dominating its electric charge.

Secondary structure—Regular folding of a protein in repeated patterns, eg, α -helix, β -pleated sheet, β -turns; also the double helix structure of a polynucleotide.

Sequence tagged site (STS)—Method to provide unambiguous identification of DNA markers generated by the Human Genome Project. STSs comprise short single-copy DNA sequences that characterize mapping landmarks on the genome.

Single nucleotide polymorphisms—Single nucleotide substitutions detected by DNA sequencing.

Somatic cell hybrid—Hybrid formed from the fusion of different cells. These usually come from different species, eg, human and rodent hybrids are frequently used for human gene mapping.

Southern blotting—Technique that combines the resolving power of agarose gel electrophoresis with the sensitivity of nucleic acid hybridization. DNA fragments separated in agarose gel are denatured in situ and then blotted or transferred, usually by capillary action, from the gel to a nitrocellulose sheet, or other binding matrix, placed directly on top of the gel. Single-stranded DNA binds to the nitrocellulose and is then available for hybridization with labeled, 32P or biotinylated, single-stranded DNA or

RNA. Labeled nucleic acid is known as the probe and, in the case of DNA, is often prepared by nick translation. The hybrids are detected by autoradiography, in the case of 32P, or a color change, in the case of a biotinylated probe. This extremely sensitive and powerful technique, often described as “blotting,” is named after its inventor, E. M. Southern.

Tandem repeats—Small sections of repetitive DNA in the genome, arranged in head-to-tail formation.

Telomeres—Two ends of a chromosome.

Termination codon—Three codons, UAA known as ochre, UAG as amber, and UGA as opal, that do not code for an amino acid but act as signals for the termination of protein synthesis. They are not represented by any tRNA, and termination is catalyzed by protein release factors. There are 2 release factors in *E coli*: RF1 recognizes UAA and UAG, and RF2 recognizes UAA and UGA. Eukaryotes have a single GTP-requiring factor, eRF.

Topoisomerases—Enzymes that change the degree of supercoiling in DNA by cutting 1 or both strands. Type I topoisomerases cut only 1 strand of DNA; type I topoisomerase of *E coli* (omega protein) relaxes negatively supercoiled DNA and does not act on positively supercoiled DNA. Type II topoisomerases cut both strands of DNA; type II topoisomerase of *E coli* (DNA gyrase) increases the degree of negative supercoiling in DNA and requires ATP. It is inhibited by several antibiotics, including nalidixic acid and ovobiocin.

Transfer RNA (tRNA)—Low-molecular-weight RNAs that specifically bind amino acids by aminoacylation and possess a special nucleotide triplet, the anticodon, sometimes containing the base inosine, by which they recognize codons on mRNA. By this recognition the appropriate tRNAs are brought into alignment. See text for further classification.

Variable numbers of tandem repeats (VNTRs)—DNA polymorphisms deriving from genomics regions with stretches of repeating nucleotides, yielding differently sized DNA segments after restriction and gel electrophoresis. Largest category of VNTRs are minisatellites including CA repeats.

Vector—General term applied to a DNA molecule, derived from a plasmid or bacteriophage, into which fragments of DNA may be inserted or cloned. The vector should contain 1 or more unique restriction sites for this purpose and be capable of autonomous replication in a defined host or vehicle organism, such that the clone sequence is reproduced. The vector molecule should confer some well-defined phenotype on the host organism, which is either selectable, eg, drug resistance, or readily detected, eg, plaque formation.

Western blotting—Technique used to detect specific proteins, a mixture of proteins being separated by PAGE, blotted onto nitrocellulose, and then exposed to a radiolabeled immunoglobulin directed to the desired protein, which is revealed by autoradiography (see Northern blotting, Southern blotting).