The Human Genome Project: Implications for the Treatment of Musculoskeletal Disease

Daniel Jaffurs, BS, and Christopher H. Evans, PhD, DSc

Abstract

The ultimate goal of the Human Genome Project is the determination of the molecular sequence of the entire human chromosomal complement. Realization of this goal will include characterization of all the genes that cause or predispose to disease, which will most certainly lead to the development of powerful new tools for diagnosis, prevention, and treatment in all medical fields, including orthopaedics. The authors review the fundamentals of human genetics and gene mapping, summarize the progress of the Human Genome Project thus far, and discuss the implications of this research as it relates to the treatment of musculoskeletal diseases.

J Am Acad Orthop Surg 1998;6:1-14

Molecular genetics, once the sole domain of the laboratory scientist, is bringing about a revolution in the understanding of complex biologic phenomena. In addition to elucidating the intricacies of living processes, more is being discovered about the genetic basis of human disease. It is this new knowledge gained from insights into molecular biology that is clearly beginning to have an impact on medical science. An individual's genetic makeup is now understood to have a role not only in the occurrence of overt affliction but also in the predisposition to disease. As many disorders are considered to be multifactorial, involving both genetic predisposition and environmental influence, a complete genomic sequence is essential to their understanding. The ultimate goal of the Human Genome Project is the determination of the molecular sequence of the entire human chromosomal complement. This detailed knowledge of the human genome will eventually lead to the discovery of all of the genes that cause disease. The characterization of these genes will prove invaluable in deciphering pathophysiologic processes at the cellular and molecular levels. These new discoveries will most certainly lead to the development of powerful new tools for diagnosis, prevention, and treatment in all medical fields, including orthopaedics.

Gene Structure and Expression

Deoxyribonucleic acid (DNA) provides the basis for all fundamental biologic processes and is the foundation of our human identity. It carries within its structure the heritable information that determines the structure of proteins. The translation of this information, via a messenger ribonucleic acid (mRNA) intermediate, to proteins is known as gene expression.

Deoxyribonucleic acid is a double-stranded helix made up of the four bases adenine (A), thymine (T), guanine (G), and cytosine (C) (Fig. 1). Each of these bases is covalently bound via a phosphate moiety to a pentose sugar (deoxyribose). These structures, called nucleotides, are linked together to form a sugar-phosphate backbone from which the bases project. Each strand of the DNA helix is one such long molecule. The bases themselves are attracted to each other by weak, noncovalent hydrogen bonds. For chemical reasons related to the molecular structure of the bases, A will normally bind only to T and G only to C. In intact double-helical DNA, this hydrogen bonding between the

Mr. Jaffurs is a doctoral candidate, Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh. Dr. Evans is Henry J. Mankin Professor of Orthopaedic Surgery, Professor of Molecular Genetics and Biochemistry, and Director, Ferguson Orthopaedic Research Laboratory, University of Pittsburgh School of Medicine.

Reprint requests: Dr. Evans, C-313 Presbyterian University Hospital, 200 Lothrop Street, Pittsburgh, PA 15213.

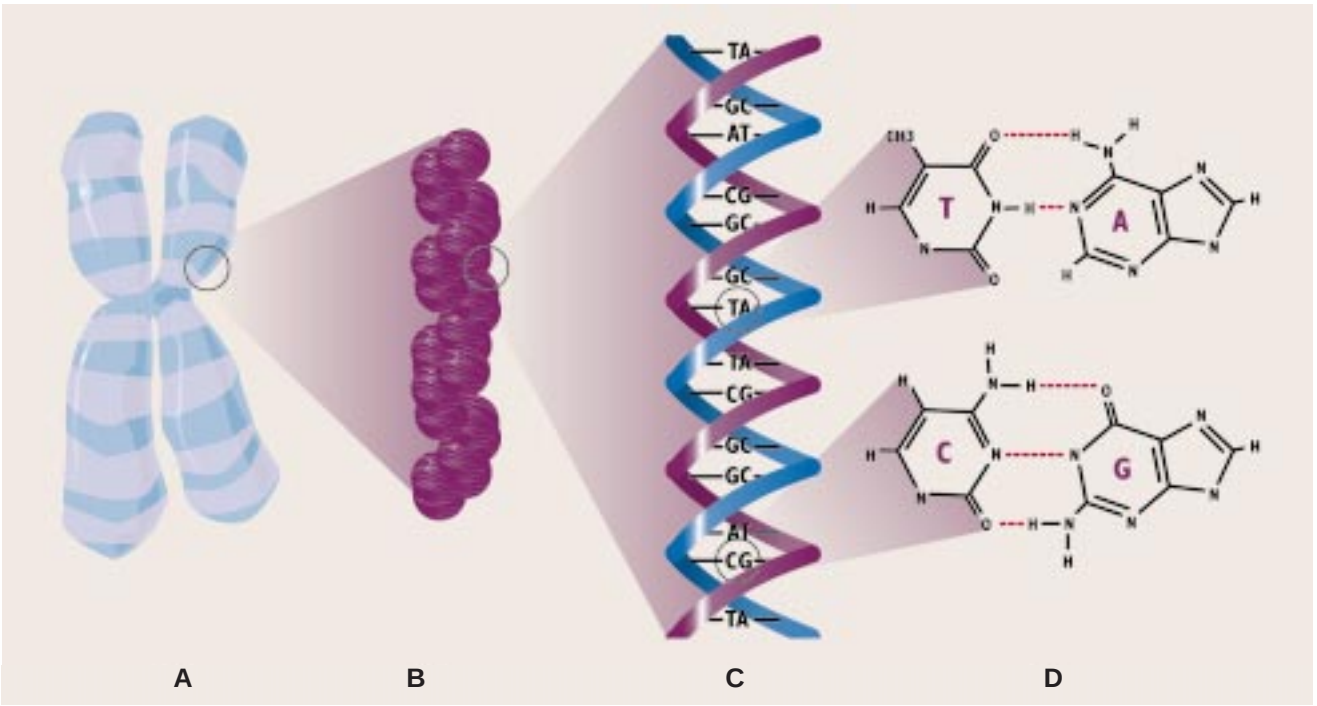

Fig. 1 DNA and chromosomal structure. **A,** In metaphase, pairs of condensed, homologous chromosomes align in a characteristic fashion. **B,** Each chromosome consists of long DNA molecules with their associated proteins. **C,** DNA is a double-stranded, helical molecule in which purine (A,G) and pyrimidine (T,C) bases project from a sugar-phosphate backbone. **D,** Base-pairing between adenine and thymine and between guanine and cytosine maintains the stability of the double helix (dotted lines represent hydrogen bonding).

bases occurs between the nucleotides on opposite strands of the helix.

The bases are arrayed linearly on long double-stranded DNA molecules, or chromosomes, in discrete units called genes. It is the linear sequence of the bases that acts as a template for the synthesis of mRNA. The RNA molecule is a single-stranded copy of one of the DNA strands. The process that makes a copy of the DNA strand to produce the mRNA is known as transcription. The mRNA is then translated by ribosomes to yield the final protein product (Fig. 2).

The ribosomes, which are organelles made of ribosomal RNA (rRNA) and protein, bind the beginning, or 5' end, of the mRNA. (As a matter of convention, the ends of the RNA and DNA molecules are distinguished by reference to one of the carbon atoms, 5' or 3', in the corresponding ribose or deoxyribose sugar.) Ribosomes "read" the sequence of the bases in the mRNA molecule and then synthesize the particular protein encoded by each specific mRNA. The bases are read in consecutive groups of three, each group signifying a particular amino acid or a signal to begin or stop the process. Each group of three bases is called a codon.

The base structure of the first codon, or "start" codon, is AUG (uracil [U] is substituted for T in the mRNA). When this codon is recognized, the ribosome begins the translational process by adding the appropriate amino acid residue via a transfer RNA (tRNA) molecule. The construction of the polypeptide chain continues until the first triplet of UAA, UAG, or UGA is reached. These "stop codons" are the universal signal for termination of protein translation. The relationship between codons and their cognate amino acids is known as the genetic code (Fig. 3).

The entire human genetic complement, or genome, consists of 50,000 to 100,000 genes, which reside on 23 pairs of chromosomes. Each autosome, or nonsex chromosome, is present in two copies, or homologues, in all somatic cells. Male somatic cells contain one X and one Y chromosome; female somatic cells contain a pair of X chromosomes. Germ-line cells, containing two copies of each chromosome, give rise to sperm and eggs, which contain only one copy of each homologue.

When one thinks of a chromosome, the image that is usually called to mind is that of a highly condensed molecule typically seen in the metaphase portion of mitosis. In this form, individual chromosomes that have undergone replication are easy to visualize with certain stains that show differential banding patterns (Fig. 1, A). These bands, consisting of long DNA strands wrapped around histone proteins (Fig. 1, B), contain on the order of 5 to 10 million base pairs.

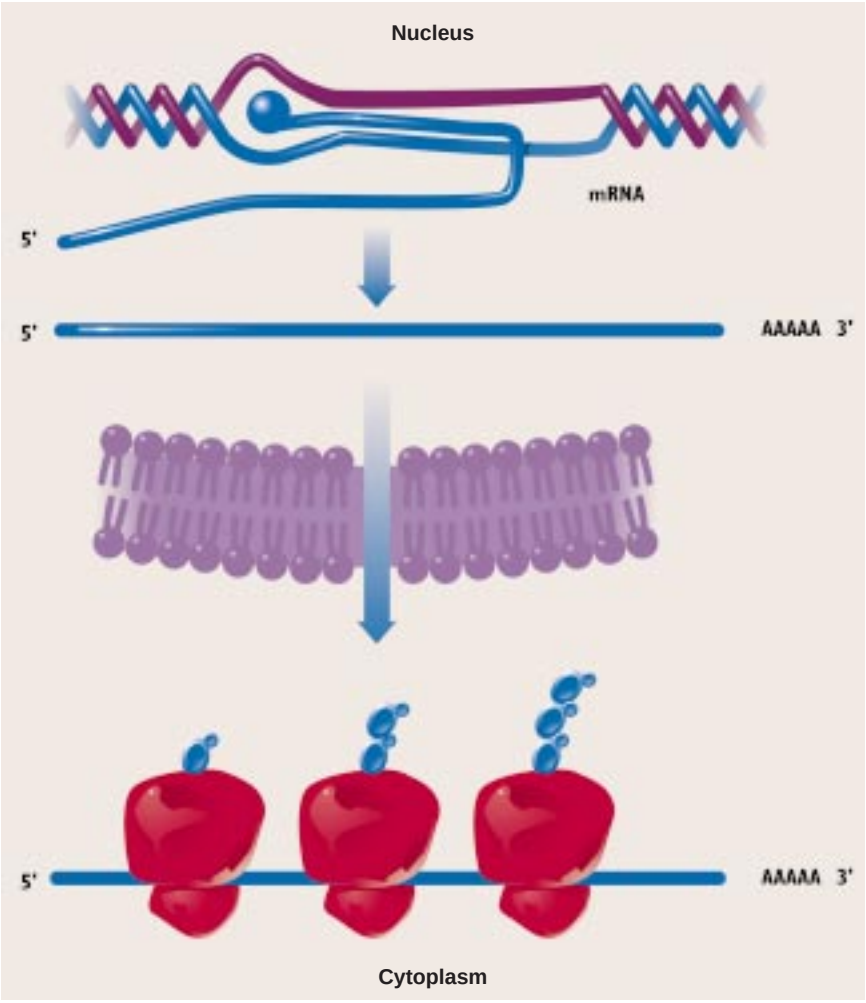

Fig. 2 Gene expression is the process whereby the genetic information contained in the DNA is manifested by the synthesis of a specific protein. The process occurs via an mRNA intermediate, which is synthesized by the enzyme RNA polymerase II. This enzyme uses one strand of the DNA helix as a template and produces a complementary RNA molecule in a process known as transcription. Partial unwinding of the DNA is necessary for transcription to occur. The RNA molecule thus formed undergoes a number of processing steps, including the addition of a polyadenine "tail." The mRNA leaves the nucleus via pores in the nuclear envelope and enters the cytoplasm. The genetic information contained in the mRNA then serves as a template for protein synthesis. The purine and pyrimidine bases in the mRNA are read three at a time as codons. Ribosomes (shown as a large circle on top of a small circle) bind to the 5' end of the mRNA and move toward the 3' end, adding amino acids to the nascent protein chain as they do so. This process is called translation.

Although the total amount of DNA in a human cell is in excess of 3 billion base pairs, only about 5% to 10% of those base pairs code for functional genes. The remaining base pairs, mostly repetitive sequences, were once thought to act solely as a chromosomal scaffold. It has recently been discovered that these "nonsense" sequences might very well be important for regulation of how much protein product is made from a given gene (i.e., gene expression). This "nonsense DNA" has also proved useful in genetic mapping.

It is important to note that while every cell in the human body contains a full complement of chromosomes (except for sex cells), not all of the genes in a given cell are expressed. Aside from the proposed role of noncoding nucleotide sequences in regulation, there are a host of well-studied mechanisms by which tissue-specific gene expression occurs.

Gene Function

Certain DNA sequences that exist outside a gene can either activate or repress gene expression. Promoter elements, which are nucleotide sequences recognized by cellular proteins called transcription factors, are usually present at the beginning of the gene. These transcription factors, or activators, physically bind to the promoter elements in order to stimulate transcription of the gene by RNA polymerase. Other sequences, called enhancers, also serve as binding sites for regulatory proteins but can be located up to 20,000 base pairs away from the gene. Both promoters and enhancers can be activated either internally in a specific fashion or externally by means of extracellular signals, such as steroid and peptide hormones. It

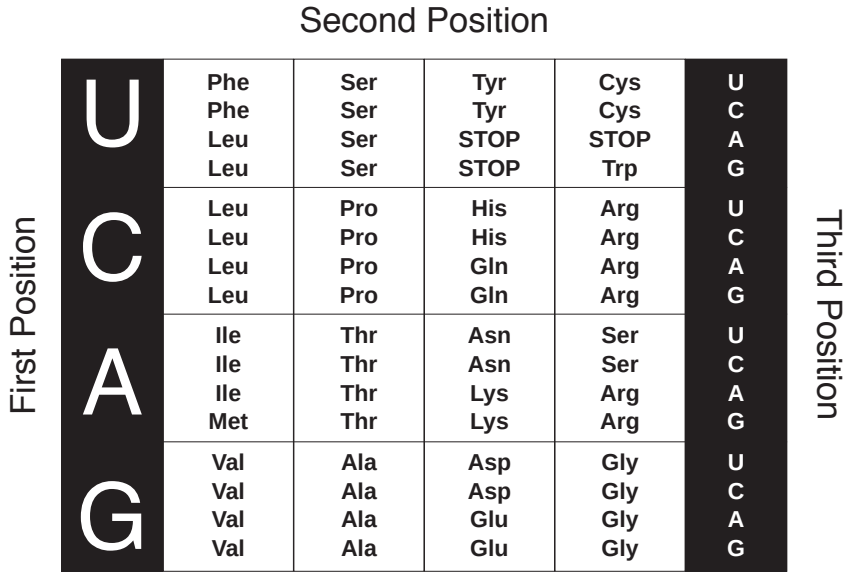

Fig. 3 The genetic code. The sequence of purine and pyrimidine bases in the mRNA is used to determine the sequence of amino acids in the protein synthesized from the mRNA. Bases are read as codons of three bases, each triplet signifying an amino acid. Diagram of the genetic code shows which codon codes for which amino acid (e.g., ACG encodes threonine; GCU encodes alanine). Three codons (UAA, UAG, and UGA) do not encode amino acids; instead, they instruct the ribosomes to terminate protein synthesis. AUG is universally used as the translation start signal and also encodes methionine; this means that all proteins initially begin with a methionine residue, although this may later be removed during processing.

is the combination of chromosomal structure, promoters, enhancers, and different cellular factors present in distinctive cell types that performs in the as yet indecipherable symphony of tissue-specific gene expression.

Mutations, or changes, in the nucleotide base sequence of a DNA molecule can occur in either coding or noncoding regions of a given chromosome. If the mutation occurs in a gene or its promoter elements, there may very well be an alteration in the ability of the gene to produce its protein product. The mutation may cause a change in a particular amino acid in the protein, which could cause that protein to become dysfunctional. This is the case with various neoplasms that occur when a tumor-suppressor gene is inactivated and with various skeletal dysplasias in which an important osteogenic or growth factor is altered. It is also possible that the mutation might lead to a gain of function; that is, the resulting change in the sequence of the amino acids may render the protein unable to respond to regulatory signals, or the protein might be expressed in a temporally aberrant manner. Finally, the change may give rise to a translation stop signal and thereby a prematurely terminated protein.

Diseases that are single-gene disorders resulting from a mutation in a particular gene can fall into one of three categories. In autosomal recessive disorders, such as cystic fibrosis, the affected person is homozygous, (i.e., the individual genes, or alleles, on both pairs of chromosomes carry the mutation). Heterozygotes have one normal, or "wild-type," allele on one chromosome but carry the mutation on the other; that is, individuals possessing this genotype are carriers of the mutant gene, but the gene is phenotypically silent because the wild-type gene makes a sufficient amount of the protein for normal function. Carriers of autosomal recessive disorders are usually discovered only after the birth of an affected child.

Autosomal dominant disorders, such as Huntington's disease and most forms of osteogenesis imperfecta, require only one copy of the mutant allele to be present. The presence of the mutant protein disrupts normal function.

In disorders that are X-linked recessive, such as hemophilia, the genetic alteration acts in a dominant fashion when transmitted to male offspring because males possess only one copy of the X chromosome. Females primarily act as carriers and, in most cases, express the disease phenotype only when both copies of the mutant gene are present.

The Human Genome Project

In the early days of genetic research, most gene mapping and DNA sequencing of organisms was performed by individual laboratories studying separate parts of the genomes of different species. With the understanding that genes were responsible for a wide variety of human disorders, particularly cancer, great interest developed in mapping the human genome. Progress in human gene mapping has historically been difficult due to a number of factors, including small family size and a relatively long generation time. Progress in a number of techniques in cytogenetic and molecular analysis, together with expanded family studies, has

brought the field of genome science to the forefront.

The Human Genome Project, cosponsored by the US Department of Energy, the US Office of Health and Environmental Research, the US Office of Energy Research, and many other research organizations around the world, officially began in 1988. The long-range goals of the program, as outlined in the initial 5-year plan,[1] were (1) to construct a high-resolution genetic map of the human genome; (2) to produce a variety of physical maps of all human chromosomes and of the DNA of selected model organisms; (3) to determine the complete sequence of human DNA and of the DNA of selected model organisms; (4) to develop capabilities for collecting, storing, distributing, and analyzing the data produced; and (5) to create the appropriate technologies necessary to achieve these objectives.

All of the short-term goals to be met in the first 5 years of the project have been accomplished ahead of schedule. These include the expansion of a high-resolution human genetic map[2] and complete physical maps of the mouse,[3] the nematode *Caenorhabditis elegans*, and various prokaryotes. The genomes of several organisms—most recently, a strain of yeast—have been sequenced in their entirety.[4-6] The surprising discovery that a *C elegans* gene is homologous to a human gene involved in early-onset Alzheimer's disease illustrates the potential insights into human disease that can be gained from the study of nonhuman model organisms. Sequences that are conserved between species are likely to point to genes in which the encoded proteins are particularly important. The yeast genome, which contains 6,000 genes, has already provided insights into human genes that may be involved in a variety of medical problems.

While the development of novel technologies that would increase the accuracy and efficiency of physical mapping and DNA sequencing has not yet been realized, refined application of existing technology, coupled with increases in computing power, has had a considerable impact on the Human Genome Project. This has prompted a change in direction from finalizing a complete physical map of the human genome to pushing ahead with a full-scale sequencing effort,[7,8] meaning that very possibly there will be an entire human genetic sequence by the end of this century.

Genetics in Medicine

Application of the tools of molecular biology to genomic science has generated a wealth of knowledge about disease processes. The number of known genetic defects involved in musculoskeletal diseases has grown tremendously in the past decade (Table 1). Our enhanced awareness of the role genetics plays in neoplastic disease has also led to an increase in the number of genes known to be involved in carcinogenesis (Table 2). In this new era of knowledge concerning the genetic components of disease, practicing clinicians from various medical specialties have already

Table 1
Genes Associated With Human Musculoskeletal Diseases

Disease	Gene
Achondrogenesis type IB	Diastrophic dysplasia sulfate-transporter gene
Atelosteogenesis type II	Diastrophic dysplasia sulfate-transporter gene
Cartilage-hair hypoplasia	Linked to chromosome 9
Chondrodysplasia punctata	Linked to Xp22.3
Craniosynostosis Adelaide type	Linked to 4p16
Crouzon syndrome	FGF receptor 2
Diastrophic dysplasia	Diastrophic dysplasia sulfate-transporter gene
Duchenne muscular dystrophy	Dystrophin
Familial osteoarthritis	COL2A1
Gaucher disease	Glucocerebrosidase
Hypochondroplasia	FGF receptor 3
Jackson-Weiss syndrome	FGF receptor 2
Kniest dysplasia	COL2A1
Marfan syndrome	Fibrillin
Multiple epiphyseal dysplasia	Cartilage oligomeric matrix protein
Osteogenesis imperfecta	COL1A1, COL1A2
Pfeiffer syndrome	FGF receptor 1
Pseudoachondroplasia	Cartilage oligomeric matrix protein
Schmid metaphyseal dysplasia	COL10A1
Spondyloepimetaphyseal dysplasia (Strudwick type)	COL2A1
Spondyloepiphyseal dysplasia	Linked to Xp22.12-p22.31 and COL2A1
Sporadic osteoarthritis	COL2A1
Stickler dysplasia	COL2A1, COL11A2

Table 2
Genes Associated With Neoplastic Diseases

Gene	Tumor Site/Type	Disease
p53	Breast, colon, bone	Li-Fraumeni syndrome
BRCA1	Breast	Early-onset breast cancer
APC	Colon	Familial adenomatous polyps
VHL	Kidney	von Hippel-Lindau disease
NF1	Neurofibroma	Neurofibromatosis type I
WT-1	Nephroblastoma	Wilms tumor
Rb	Retinoblastoma	Retinoblastoma
NF-2	Schwannomas and meningiomas	Neurofibromatosis type II
RET	Thyroid, pheochromocytoma	Multiple endocrine neoplasia type II

begun to appreciate the necessity of understanding the molecular aspects of disease.[9-11] As genome research continues, all of medicine will continue to feel the impact that genetic mapping and sequencing will have on the way diseases are diagnosed and treated.

For example, Duchenne muscular dystrophy (DMD), transmitted through the X chromosome, is a relatively common, severe, and untreatable disease involving persistent and eventually fatal muscular degeneration.[12] Patients with DMD usually die before their second decade due to either respiratory or cardiac failure. Accurate diagnosis of DMD in affected children previously involved muscle biopsies, analysis of serum enzyme levels, and electromyography. Before the gene was discovered, it was very difficult to predict with any accuracy the probability of having affected offspring.

The initial genetic research began with cytogenetic studies of affected females who were heterozygous for the DMD trait.[13] Because of the recessive nature of DMD, these carrier females should not have been affected. It was found that these individuals had undergone a translocation between their normal X chromosome and another chromosome, thereby incurring a mutation in the wild-type DMD gene. Analyses of the chromosomal banding pattern showed that the breakpoint of the translocation was in all cases isolated to a particular segment of the X chromosome. In addition, a large deletion was discovered in the same region of the X chromosome in a male DMD patient.[14] This information, combined with the results of linkage analysis, led to the eventual cloning and characterization of the dystrophin gene.[15]

Once the gene had been characterized, DNA-based diagnostics were developed for the detection of asymptomatic heterozygotes in DMD families and for prenatal screening in cases in which one of the parents was a carrier. Although it is not absolutely necessary to have the actual gene in hand to develop a genetic test, the accuracy of diagnosis increases substantially when one can assay directly for the mutation. It is important to mention that while a genetic test based on a known mutation is extremely accurate and is a very powerful method of screening for genetic diseases, it cannot provide useful information concerning new mutations that may occur in utero.

After identification of the causative gene, work could begin on replacement of the defective dystrophin gene with a properly working version.[16] This type of intervention, popularly known as gene therapy, will be discussed later.

Positional Cloning and the Genetic Map

The mapping, or localization, of a particular gene along a chromosome is a formidable task. Attempting to find a particular gene, which may encompass only a few thousand base pairs in a background of several billion, is the genetic equivalent of trying to find a needle in a haystack. Of the methods used to find and clone genes, the most straightforward is by utilizing information from the protein product of the gene. The protein is isolated, and its amino acids are sequenced. That sequence is then used to determine the most likely nucleotide sequence. Synthetic DNA oligonucleotides consist of single strands of nucleotides containing the "most likely" sequence as determined by the genetic code (Fig. 3). The small DNA strands produced are then used to probe libraries of human DNA to find the gene. However, this is possible only when the actual biochemical defect is known and the responsible protein is identified and purified. Unfortunately, as this is a relatively rare occurrence, it is the knowledge of the position of the gene that offers the best opportunity to identify and eventually clone the gene.[17]

A genetic map is made up of a collection of unique markers ordered along a specific chromosome. The classic marker used extensively in genetic mapping is the restric-

tion fragment length polymorphism (RFLP). An RFLP analysis utilizes fragments of DNA that are generated after digestion with an appropriate restriction endonuclease (an enzyme that recognizes and cleaves DNA at specific sequence sites). Polymorphisms (variations in the resultant pattern of DNA fragments) observed among members of a species can often be associated with specific diseases or physical characteristics. These patterns can be detected by transferring the DNA to a nitrocellulose membrane and then incubating the membrane with a short radiolabeled sequence of DNA that is complementary, and therefore binds in a specific fashion, to a diagnostic region in the RFLP. Because only complementary sequences will hybridize to each other via hydrogen bonding, a specific region of DNA can be identified, analyzed, and compared (Fig. 4).

More recently, short repetitive sequences of DNA have been used in the mapping process. These short tandem repeats (also called microsatellite markers) consist of a nucleotide sequence measuring one to four bases in length that is repeated several times. The use of CA repeats (DNA sequences containing many alternating cytosine and adenine nucleotides) has shown immense utility in mapping genes when analyzed by means of the process known as polymerase chain reaction.

In short, polymerase chain reaction is an in vitro process that involves the synthesis and massive amplification of specific DNA sequences.[18] Short oligonucleotide "primers" that are complementary to sequences flanking the gene or area of interest are synthesized in the laboratory. Then DNA extracted from cells is mixed with the primers; the four nucleotides A, T, C, and G; and a thermostable DNA polymerase. The mixture is heated to a temperature that denatures the DNA (i.e., causes the double helix to break apart) and is then cooled, so that the primers can base-pair to their specific complementary sequences on the DNA template. After this, the polymerase can synthesize two new DNA strands initiating from the ends of the primers. This cycle of denaturation, annealing, and synthesis is repeated many times and results in an exponential amplification of the target sequences. After 32 cycles, more than 1 million copies have been made. The resulting products can be size-fractionated on a polyacrylamide gel and then analyzed.

Whether the marker is a restriction site or a unique DNA sequence, the only requirement for that marker to be informative is that it be polymorphic. The diploid genome contains two copies, or homologues, of each chromosome. Because the gene or region of DNA being mapped resides on one of the homologues, the marker must be specific for the sequence variation of that particular chromosome. The region of DNA being analyzed must contain the mutation, but it is not a requirement that the marker be located within or be a part of the actual gene. Indeed, it is rarely the case that the marker is the gene itself. It need only be close enough

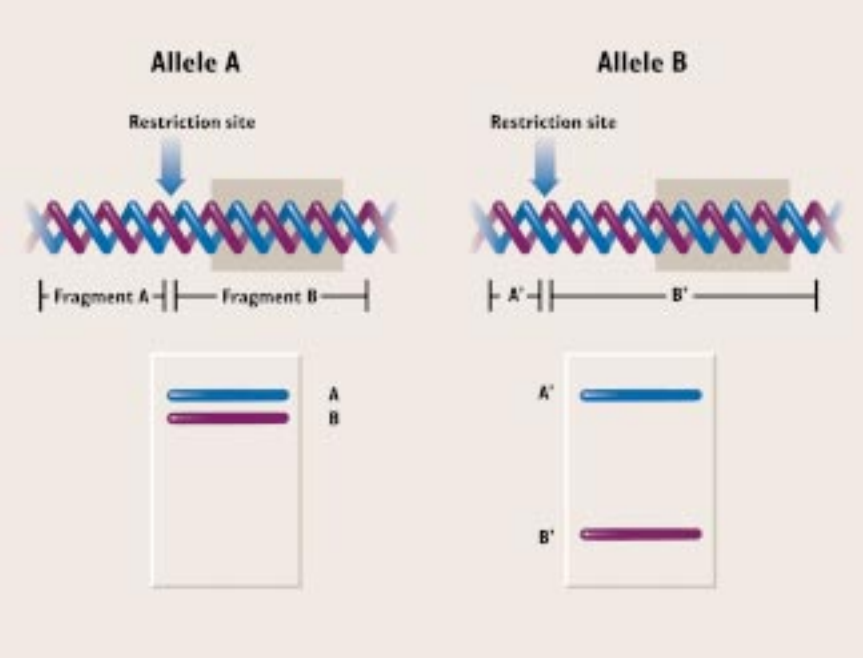

Fig. 4 Restriction fragment length polymorphism. Restriction enzymes cleave DNA in a highly specific manner at sites containing particular base sequences, known as restriction sites. In the example shown, there are two alternative forms of a gene (alleles). The difference between alleles A and B is that they both contain a restriction site for a particular restriction enzyme, but the site is at a different point in the gene. This means that digestion of the two alleles by the restriction enzyme will generate fragments of different sizes (A, B, A', B'). Electrophoretic techniques can be used to separate these fragments according to size. The fragments can then be visualized by hybridization to a labeled nucleotide probe with the complementary sequence. Inspection of these bands permits the investigator to determine whether allele A or allele B is present. In heterozygous individuals, all four restriction fragments would be present.

on the DNA strand that the marker is inherited with (i.e., is linked to) the disease locus during meiotic recombination. As the ovum matures during prophase I of meiosis, homologous chromosomes pair with each other, and some recombination (exchange of genetic information) occurs. If the marker is not physically close to the defective gene, it may be left behind during the recombination process. In this case, the marker is no longer linked to the defective gene and is not useful for diagnostic purposes.

For single-gene (monogenetic) disorders inherited in a mendelian fashion, one can compare the inheritance of a mutant gene with the inheritance of an informative marker within a particular family. Coinheritance of a marker with the disease phenotype suggests that the marker is physically close to the mutant gene. This technique has been used to find over 40 genes associated with various diseases,[19] not including those associated with the various forms of neoplastic disease.

Genetic maps are extremely useful for identifying the general location of genes associated with disease. Even if the gene itself has not been precisely located, a closely linked marker can provide information for genetic counseling and may serve as the basis of a DNA-based diagnostic test.

For all its utility, genetic mapping possesses several inherent drawbacks. The generation or construction of these maps is very labor-intensive, in large part because the limit of resolution is only about 1 centimorgan (cM). (The centimorgan, named for the geneticist T. H. Morgan, is a measurement of genetic distance based on the observed recombination frequency. For example, two markers are considered to be 1 cM apart [roughly equivalent to 1 million base pairs] if they recombine only once in every 100 meioses.) Nevertheless, 1 cM is still a sizable fragment within which to locate a specific gene. Also, genetic distance, as opposed to physical distance, relies on the observation of recombination between markers to elucidate the approximate position of a particular gene. Because some regions of the chromosome are prone to higher than what would be considered normal rates of recombination, genetic mapping is not a straightforward proposition. In the case of Huntington disease, the original marker that was used to trace the disease inheritance through some lineages was physically close to the Huntington gene, although abnormally high rates of recombination made it appear much farther away. It took more than 10 years for the gene to be cloned because of problems with mapping.

There are an estimated 4,000 disease-causing genes. This does not take into account complicated multifactorial and polygenic disorders such as obesity, diabetes, and arthritis. While a genetic map is useful for counseling and for giving rough approximations of gene locations, a much finer map is needed for precisely locating genes of interest and providing accurate information for genetic counseling and for diagnosis and prognosis of conditions with genetic components.

Physical Mapping of Human Chromosomes

Identifying and cloning a disease-associated gene is an arduous process. For polygenic diseases (i.e., those involving more than one genetic locus), highly resolved maps and more complex analyses are required. Aside from its utility in counseling and approximating the position of a particular gene, mapping through observation of recombination is of limited resolution. In contrast, a physical map is a specific ordering of DNA markers along a chromosome according to the actual physical distance between them. Because the physical distance does not vary as it does with linkage analysis, resolution is limited only by the number of markers one can order along a given chromosome. The ultimate physical map would, of course, be the complete genomic sequence of a given organism.

It is becoming increasingly apparent that efforts in both genetic and physical mapping of human chromosomes, along with gene discovery efforts, have yielded enough information to begin full-scale DNA sequencing of the human genome. Once complete sequence data have been developed, new genes will be identified on the basis of similarity to known genes (e.g., identification of promoter regions and similarities of functional domains). As the electronic databases become filled with new genetic sequences and even the entire genomes of living organisms, scientific and medical researchers are looking for better ways to convert this information into an understanding of gene expression and the mechanisms of disease. At an ever increasing rate, this information is being taken from the laboratory to the clinic. One can only begin to imagine the diagnostic and therapeutic potential that complete knowledge of the genome will bring.

Genetics and Orthopaedic Surgery

Diagnostic Tools

The discovery of new disease-causing genes and the increasing ease of genetic analysis have al-

ready had a considerable impact on clinical medicine, particularly in the area of molecular diagnostics. Within the field of musculoskeletal disease, 20 or more diseases have already been linked to specific genes or chromosomes. Linkage studies can provide valuable diagnostic information if the gene has been genetically mapped but still remains unidentified. However, because of the need for genetic markers that are somewhat close to the disease locus (so as to avoid the possibility of recombination), devising a more accurate diagnostic test to identify the defective coding region requires the actual molecular sequence of the disease gene.

The characterization of individual mutations, while sometimes technically difficult in the case of a large gene, is essential for several reasons. Different mutations in the same gene may give rise to different phenotypes. This is certainly the case for the gene encoding the α_1 chain of type II collagen (COL2A1). The clinical presentations include Kniest dysplasia,[20] Stickler dysplasia,[21] spondyloepimetaphyseal dysplasia (Strudwick type),[22,23] sporadic osteoarthritis,[24] and familial osteoarthritis.[25] In the case of Stickler dysplasia, it has been noted that the same phenotype can occur with mutations in two different genes (COL2A1 and COL11A2).[26]

The clinical outcomes resulting from these mutations may also differ. Becker muscular dystrophy and DMD both result from mutations in the gene encoding the protein dystrophin. As discussed earlier, DMD is a devastating illness that usually results in death before the age of 20. The mutation is usually one or more large deletions of the dystrophin gene, which interfere with translation. Becker muscular dystrophy is a milder variant in which the mutation gives rise to a shorter but somewhat functional protein. There are also differing variants within each, based on the severity of the mutation.

Osteogenesis imperfecta is another inherited disorder that comprises a wide array of clinical variations.[27] The clinical phenotypes, ranging from a mild increase in fracture frequency to a lethal perinatal form,[28] vary according to the chain of type I procollagen that is affected and the type and location of the mutation. A complete analysis of the gene will make it possible to predict the phenotype that will result from each type of molecular defect.

After the appropriate disease-causing gene has been characterized, various methods for detecting mutations can be applied. In the case of prenatal diagnosis, samples of fetal amniotic fluid or tissues are obtained and tested for either biochemical defects or genetic abnormalities. Fetal cells can be obtained from cord blood, amniotic fluid, or chorionic villi. Advances in fluorescent activated cell sorting may soon make it possible to obtain fetal cells directly from the peripheral blood of the mother. Fetal erythroblasts isolated from the maternal circulation have been used to diagnose the presence of mutations causing sickle cell anemia and thalassemia.[29]

Tests that detect the biochemical component of a metabolic disorder have a somewhat limited application. The sample of amniotic fluid or chorionic villi often does not contain enough material for testing. Culturing of fetal cells can be accomplished, although it is not facile. In those diseases in which a gene is expressed only in certain tissues, chorionic villi or amniotic fluid may not express the gene at all. Furthermore, relatively few of the protein products for the various inherited disorders are known.

Of the many molecular tools that have revolutionized medical diagnostics, the most important has been polymerase chain reaction. Once the nucleotide sequence of a gene is known, polymerase chain reaction can be used to rapidly amplify the DNA and thereby detect mutations in a very small sample.

A resounding success in both basic and applied sciences, polymerase chain reaction seems to offer overwhelming utility in diagnostics. Because the amplified DNA can be sequenced, the resultant detection of any changes in the nucleotide sequence is unequivocal. If the frequency of a mutation in a given population is known, one can assay directly for that mutation. For example, if the defect is a deletion, the amplified DNA can be resolved according to size with gel electrophoresis. In the case of a mutation involving a single base pair, primers can be designed that anneal to either the normal gene or a mutant copy of the gene. The reaction mixture can then be probed to determine whether one allele or the other (or both in the case of a heterozygote) is present. A more straightforward approach is to use primers for polymerase chain reaction that flank regions of a gene shown to develop mutations. Once these sections have been amplified, they can be probed with a labeled oligonucleotide specific to either the normal or the mutant form.

Prognostic Indicators

It is becoming readily apparent that the outcomes of virtually all diseases are influenced by the genetic makeup of the individual. The discovery and characterization of genes and the mutations that render them dysfunctional will certainly be beneficial for understanding the pathophysiology of the disease

process. Once we have a reasonable understanding of the full human sequence, knowledge of a patient's specific genetic identity will be useful in determining susceptibility and outcome.

It has been reported that as many as 10% of all fractures occurring in the United States result in delayed or impaired union.[30] This situation may arise from difficulties with operative interventions, such as inadequate mobilization or fixation of the fracture, or may be a consequence of damage to surrounding soft tissues. Excessive early motion of a fracture postoperatively may also interfere with healing. Furthermore, certain regions of the skeleton, such as the neck of the femur, the neck of the talus, and the carpal scaphoid, are known to have an increased risk for difficulties during healing. These problems could be related to the distribution and adequacy of the local blood supply or to the control of mechanical strain in a particular location.

Similarly, graft failure occurs in a substantial number of patients who have undergone anterior cruciate ligament reconstructions.[31] As with fracture healing, the possible explanations include both operative and postoperative factors, such as improper notchplasty, incorrect tunnel placement, inadequate tensioning, inappropriate rehabilitation, and repeated trauma.

However, the continued development of new techniques for exploring cellular and molecular mechanisms may lead us to another explanation. Mechanical and iatrogenic factors aside, it is highly likely that some individuals are genetically predisposed to a less than satisfactory result. Continued advances in genetic analysis will surely enhance our understanding of the basic events that regulate the repair of skeletal tissues. Subtle genetic differences between individuals may explain why some patients heal better than others and why some are predisposed to orthopaedic complications. Diagnostic advances and new knowledge that can lead to a more complete comprehension of inherited disease will also lead to improved and novel therapeutic approaches.

Interventions and Therapy

The information gained from the cloning and subsequent characterization of a disease-causing mutation within a gene will not only yield a more complete understanding of the pathophysiology of the disease; new interventions utilizing various genetic manipulations will undoubtedly evolve. Recombinant DNA technology has already had an impact on clinical practice in the treatment of certain disorders. Bacteria that contain the gene for human insulin economically produce vast amounts of the hormone for use in the treatment of diabetes. Human growth hormone, once extracted from cadaver tissue that in some instances gave rise to neurodegenerative disease, is now also safely made by microorganisms. Studies in animals have demonstrated the utility of the addition of bone morphogenic proteins in accelerating repair of skeletal defects.

Not only will analysis of the structure and function of proteins be valuable for better drug design and disease management, the gene itself may also provide the means of treatment. Gene therapy, although still in its infant stage,[32] will undoubtedly change the way orthopaedists treat their patients. The many ongoing and approved clinical trials of gene therapy are evidence of the interest in this type of treatment for a variety of inherited and acquired diseases. The theory behind gene transfer is a simple one: replacing or augmenting the mutant gene with the wild-type gene compensates for the defect. This approach has been used successfully to correct genetic defects in vitro and will eventually become commonplace in the clinic. Its potential applications to the musculoskeletal system were recently reviewed by Evans and Robbins.[33]

Genetic modification of cells, tissues, and even whole organisms can be achieved with several different methods (Tables 3 and 4). Of the variety of viral vectors currently being utilized, the two most commonly used are the adenovirus and the retrovirus. Many of the clinical trials taking place in gene therapy involve one of these delivery systems. Liposomes have been explored clinically as carriers of exogenous DNA, as have DNA-protein conjugates. With these methods, it may even be possible to target the gene to a specific tissue through interactions with its cellular surface receptors. The "gene gun," which delivers DNA-coated microscopic particles, or "bullets," has also shown promise as a gene delivery system. While none of these methods has yet provided a cure, further research into the biologic nature of various methods of gene transfer will undoubtedly provide the clues for the needed refinements.

Gene therapy can itself can be divided into two categories: germline therapy and somatic-cell therapy. Germ-line therapy causes a permanent genetic change, which is then transmitted to the offspring of a given individual. There is a considerable amount of debate concerning the ethical consequences of germ-line therapies, and there are no current or forthcoming trials involving this approach. Somatic-cell therapy, which may be considered similar to an organ or tissue transplant, does not modify germ

Table 3
Viral Methods of Gene Transfer

Vector	Advantages	Disadvantages
Retrovirus	Very efficient and stable transfer of genes; well developed as a gene transfer system; widely used in human trials; easy to produce.	Does not transduce nondividing cells; variability of length of gene expression; requires extensive safety testing for microorganismic contamination, replication-competent viruses, and toxicity.
Adenovirus	Highly efficient infection of dividing and nondividing cells; infects a broad range of host cells; easy to produce; currently used in clinical trials.	DNA does not integrate and is therefore lost from daughter cells after division; toxicity; immunogenic due to persistent production of viral proteins.
Adeno-associated virus	Nonpathogenic; wild-type virus can be produced at high titer and has site-specific chromosomal integration; infects a broad range of host cells; efficient infection of dividing and nondividing cells.	Only small (<4.3-kilobase) inserts can be packaged; production is not straightforward; recombinant virus loses site-specificity of integration and currently cannot be made at high titers.
Herpes simplex virus	Infects an extremely broad range of host cells; very efficient infection of dividing and nondividing cells; very large (>35-kilobase) inserts can be packaged; easily produced at very high titer.	Gene expression is transient; currently toxic; still being developed as a gene-therapy vector.

Table 4
Nonviral Methods of Gene Transfer

Vector	Advantages	Disadvantages
Naked DNA	No limit to the size of the gene insert; not immunogenic; very easy to produce; uptake by nondividing cells.	Very low efficiency of transduction; variability of length of gene expression.
DNA-ligand complexes	A wide range of inserts can be accommodated; targeted delivery can be achieved.	Requires knowledge of receptors specific to the target tissue.
Liposomes	More efficient than other nonviral methods; capacity for binding large amounts of DNA; available commercially; has been used in clinical trials.	Gene expression is transient; toxicity at higher concentrations; efficiency of transduction is cell-type-dependent; does not work well in vivo.
Gene gun	Efficient in vitro transfer; not limited as to size of DNA.	Gene expression is transient; in vivo access is difficult.

cells, and the genetic alteration is restricted to the treated patient.

Somatic-cell genetic therapies are currently being undertaken for a wide variety of diseases. Most of these are life-threatening, such as cancer, hemophilia, and familial hypercholesterolemia. In one recently inaugurated clinical trial, gene therapy is being used to treat arthritis; this is the first such trial for the treatment of a nonlethal disease.[34] It is very likely that this kind of treatment for other nonlethal disorders will soon follow. For example, some of the immunogenic problems that occur with bone and tissue allografts[35] may be precluded by transducing the transplanted tissue with a DNA construct that expresses an immunoregulatory cytokine. The cytokine could possibly suppress the host immune response locally or cause induction of antigenic tolerance. For patients who experience difficulties with delayed union and nonunion, the delivery of genes coding for factors that promote bone growth may alter an otherwise unfavorable outcome.[36] It can even be envisioned that, once the genes governing the growth and regeneration of various tissues are discovered, it will be possible to repair damaged cartilage, ligament, tendon, and other components of the musculoskeletal system.

Ethical Considerations

The Ethical, Legal, and Social Implication Working Group was established at the inception of the Human Genome Project in order to discuss and deal with issues that would arise as new genetic knowledge became available. The area that has occasioned the most concern, and has arguably caused the most debate, centers on genetic screening. Most of the information derived from genetic tests will be predictive at best. Since certain genes may only increase the likelihood of the development of a disorder, should the general population be tested? Or should testing be limited to individuals at high risk for diseases for which some type of therapeutic intervention may be possible? Suppose an individual is tested and found to have a mutation in a tumor-suppressor gene. The patient might benefit from preventive measures, such as avoiding exposure to environmental hazards that may lead to an accumulation of further mutations and possible carcinogenesis. However, there is no guarantee that the disease would progress. It is the statistical nature of genetic testing that must not only be effectively communicated to the patient but also must be safeguarded from possible misapplication. There have already been cases of genetic discrimination, in which people have been denied health insurance on the basis of this type of data. It is easy to envision how this might also affect employability.

There is some general consensus to limit genetic testing, especially prenatal testing, to that which is medically relevant.[37] As more genes are discovered and new therapies evolve, this type of testing will undoubtedly expand. As with all types of medical testing, patient autonomy and privacy should be respected and preserved.

Conclusion

The molecular tools used in the Human Genome Project have made it possible to identify some of the genes that are directly responsible for many inherited and acquired human diseases. As the project nears completion, all human genes will be found. This information will undoubtedly revolutionize medical practice and biologic research as we advance our understanding of how genes affect both health and disease. Animal models for human disease research will be more easily developed, facilitating the understanding of gene function in healthy and affected individuals, and dependable diagnostics will be developed for most inherited diseases.

It is very likely that, at the beginning of the next century, multiple rapid-screen genotype tests will be developed for clinical use. Similar to the way blood is now taken from a newborn for routine tests, the DNA from circulating leukocytes will be extracted, and certain sequences will be amplified and then sequenced by an automatic process. Before the baby leaves the hospital, portions of its genetic identity will have been analyzed by a computer program designed to look for genetic patterns that would either predispose to or cause particular illnesses. On the basis of this analysis, various preventive or therapeutic interventions will be adopted, and a monitoring schedule will be implemented.

The technology for this type of test is already available, and its applications extend well beyond the realm of prenatal and postnatal screening. This new and exciting diagnostic methodology utilizes a small card that contains an array of immobilized target-gene sequences. The DNA from a patient is labeled with a fluorescent tag and hybridized to the target sequences.[38,39] The resulting differences in fluorescence, which are dependent on the degree of hybridization, allow analysis of the presence or absence of mutations. If mRNA is used instead of DNA, one can obtain a profile of the genes being expressed in a particular tissue. This will obviously have a great impact

in the area of cancer biology; tumors will be profiled and staged according to their gene expression. One may also surmise that all diseases are due to or are accompanied by changes in gene expression. Monitoring of expression will not only provide information necessary for disease diagnosis and prognosis but also will aid in therapeutic intervention.

As more of the human genome is discovered, this technology and others will change the face of clinical medicine. Physicians who do not grasp the profound nature of the impact of the Human Genome Project and incorporate this information into their daily practice may find themselves becoming obsolete.

References

1. *Understanding Our Genetic Inheritance: The U.S. Human Genome Project—The First Five Years: Fiscal Years 1991-1995*. Washington, DC: US Dept of Health and Human Services publication NIH 90-1590, 1995.
2. Murray JC, Buetow KH, Weber JL, et al: A comprehensive human linkage map with centimorgan density. *Science* 1994;265:2049-2054.
3. Copeland NG, Jenkins NA, Gilbert DJ, et al: A genetic linkage map of the mouse: Current applications and future prospects. *Science* 1993;262:57-66.
4. Fraser CM, Gocayne JD, White O, et al: The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995; 270:397-403.
5. Fleischmann RD, Adams MD, White O, et al: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;269:496-512.
6. Butler D: Interest ferments in yeast genome sequence. *Nature* 1996;380:660-661.
7. Gibbs RA: Pressing ahead with human genome sequencing. *Nat Genet* 1995;11: 121-125.
8. Olson MV: A time to sequence. *Science* 1995;270:394-396.
9. Hall CC, Herring JA, Hall TJ: Molecular oncology and the surgeon. *Am Surg* 1995;61:156-160.
10. Smith MP: Exploring molecular biology: An older surgeon looks at a new universe. *Arch Surg* 1995;130:811-816.
11. Rossiter BJF, Caskey CT: Impact of the Human Genome Project on medical practice. *Ann Surg Oncol* 1995;2:14-25.
12. Witkowski JA: Dystrophin-related muscular dystrophies. *J Child Neurol* 1989;4:251-271.
13. Ray PN, Belfall B, Duff C, et al: Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. *Nature* 1985;318:672-675.
14. Kunkel LM, Monaco AP, Middlesworth W, Ochs HD, Latt SA: Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc Natl Acad Sci USA* 1985;82:4778-4782.
15. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* 1987;50:509-517.
16. Cox GA, Cole NM, Matsumura K, et al: Overexpression of dystrophin in transgenic mdx mice eliminates dystrophic symptoms without toxicity. *Nature* 1993;364:725-729.
17. Collins FS: Positional cloning moves from perditional to traditional. *Nat Genet* 1995;9:347-350.
18. Mullis KB, Faloona FA: Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* 1987;155:335-350.
19. Gottesman MM, Collins FS: The role of the Human Genome Project in disease prevention. *Prev Med* 1994;23:591-594.
20. Bogaert R, Wilkin D, Wilcox WR, et al: Expression, in cartilage, of a 7-amino-acid deletion in type II collagen from two unrelated individuals with Kniest dysplasia. *Am J Hum Genet* 1994;55: 1128-1136.
21. Winterpacht A, Hilbert M, Schwarze U, Mundlos S, Spranger J, Zabel BU: Kniest and Stickler dysplasia phenotypes caused by collagen type II gene (COL2A1) defect. *Nat Genet* 1993;3: 323-326.
22. Tiller GE, Polumbo PA, Weis MA, et al: Dominant mutations in the type II collagen gene, COL2A1, produce spondyloepimetaphyseal dysplasia, Strudwick type. *Nat Genet* 1995;11:87-89.
23. Bleasel JF, Bisagni-Faure A, Holderbaum D, et al: Type II procollagen gene (COL2A1) mutation in exon 11 associated with spondyloepiphyseal dysplasia, tall stature and precocious osteoarthritis. *J Rheumatol* 1995;22:255-261.
24. Bleasel JF, Holderbaum D, Haqqi TM, Moskowitz RW: Clinical correlations of osteoarthritis associated with single base mutations in the type II procollagen gene. *J Rheumatol [Suppl]* 1995;43:34-36.
25. Ritvaniemi P, Körkkö J, Bonaventure J, et al: Identification of COL2A1 gene mutations in patients with chondrodysplasias and familial osteoarthritis. *Arthritis Rheum* 1995;38:999-1004.
26. Vikkula M, Mariman ECM, Lui VCH, et al: Autosomal dominant and recessive osteochondrodysplasias associated with the *COL11A2* locus. *Cell* 1995;80:431-437.
27. Prockop DJ, Baldwin CT, Constantinou CD: Mutations in type I procollagen genes that cause osteogenesis imperfecta. *Adv Hum Genet* 1990;19: 105-132.
28. Cole WG, Dalgleish R: Perinatal lethal osteogenesis imperfecta. *J Med Genet* 1995;32:284-289.
29. Cheung MC, Goldberg JD, Kan YW: Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood. *Nat Genet* 1996;14:264-268.
30. Einhorn TA: Enhancement of fracture-healing. *J Bone Joint Surg Am* 1995;77: 940-956.
31. Vergis A, Gillquist J: Graft failure in intra-articular anterior cruciate ligament reconstructions: A review of the literature. *Arthroscopy* 1995;11:312-321.
32. Friedmann T: Human gene therapy: An immature genie, but certainly out of the bottle. *Nat Med* 1996;2:144-147.
33. Evans CH, Robbins PD: Possible orthopaedic applications of gene therapy. *J Bone Joint Surg Am* 1995;77: 1103-1114.
34. Evans CH, Robbins PD, Ghivizzani SC, et al: Clinical trial to assess the safety, feasibility, and efficacy of transferring a potentially anti-arthritic cytokine gene to human joints with rheumatoid arthritis. *Hum Gene Ther* 1996;7:1261-1280.

35. Hofmann GO, Kirschner MH, Wangemann T, Falk C, Mempel W, Hammer C: Infections and immunological hazards of allogeneic bone transplantation. *Arch Orthop Trauma Surg* 1995;114:159-166.
36. Gerich TG, Kang R, Fu FH, Robbins PD, Evans CH: Gene transfer to the rabbit patellar tendon: potential for genetic enhancement of tendon and ligament healing. *Gene Ther* 1996;3: 1089-1093.
37. Knoppers BM, Chadwick R: The Human Genome Project: Under an international ethical microscope. *Science* 1994;265:2035-2036.
38. Williamson B: Towards non-invasive prenatal diagnosis. *Nat Genet* 14:239-240.
39. Cheung MC, Goldberg JD, Kan YW: Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood. *Nat Genet* 14:264-268.