

Review Article

Genomic Medicine

ALAN E. GUTTMACHER, M.D.,
AND FRANCIS S. COLLINS, M.D., PH.D., *Editors*

GENOMIC MEDICINE — A PRIMER

ALAN E. GUTTMACHER, M.D.,
AND FRANCIS S. COLLINS, M.D., PH.D.

HUMANS have known for millennia that heredity affects health.¹ However, Mendel's seminal contribution to the elucidation of the mechanisms by which heredity affects phenotype occurred less than 150 years ago, and Garrod began applying this knowledge to human health only at the start of the past century. For most of the 20th century, many medical practitioners viewed genetics as an esoteric academic specialty; that view is now dangerously outdated.

THE ADVENT OF GENOMIC MEDICINE

The recent completion of the draft sequence of the human genome^{2,3} and related developments have increased interest in genetics, but confusion remains among health professionals and the public about the role of genetic information in medical practice. Inaccurate beliefs about genetics persist, including the view that in the past it had no effect on the practice of medicine and that its influence today is pervasive. In fact, for decades knowledge of genetics has had a large role in the health care of a few patients and a small role in the health care of many. We have recently entered a transition period in which specific genetic knowledge is becoming critical to the delivery of effective health care for everyone.

If genetics has been misunderstood, genomics is even more mysterious — what, exactly, is the difference? Genetics is the study of single genes and their effects. "Genomics,"⁴ a term coined only 15 years ago, is the study not just of single genes, but of the functions and interactions of all the genes in the genome.

From the National Human Genome Research Institute, National Institutes of Health, Bethesda, Md. Address reprint requests to Dr. Guttmacher at Bldg. 31, Rm. 4B09, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-2152, or at guttmach@mail.nih.gov.

Genomics has a broader and more ambitious reach than does genetics. The science of genomics rests on direct experimental access to the entire genome and applies to common conditions, such as breast cancer⁵ and colorectal cancer,⁶ human immunodeficiency virus (HIV) infection,⁷ tuberculosis,⁸ Parkinson's disease,⁹ and Alzheimer's disease.¹⁰ These common disorders are also all due to the interactions of multiple genes and environmental factors. They are thus known as multifactorial disorders. Genetic variations in these disorders may have a protective or a pathologic role in the expression of diseases.

The role of genomics in health care is in part highlighted by the decreasing effect of certain environmental factors, such as infectious agents, on the burden of disease. Genomics also contributes to the understanding of such important infectious diseases as the acquired immunodeficiency syndrome (AIDS)⁷ and tuberculosis.⁸

The following two case vignettes illustrate how knowledge of genomics may lead to better management of common medical conditions.

Thirty-four-year-old Kathleen becomes pregnant and sees a new physician for her first prenatal visit. Her medical history is remarkable for an episode of deep venous thrombosis five years earlier while she was taking oral contraceptives; her mother had had deep venous thrombosis when pregnant with Kathleen. Her physician suspects that Kathleen has a hereditary thrombophilia and obtains blood tests to screen for a genetic predisposition to thrombosis. Kathleen proves to be among the approximately 4 percent of Americans who are heterozygous for a mutation in factor V known as factor V Leiden that increases the risk of thrombotic events. On the basis of this knowledge and her history of possibly estrogen-related thromboembolism, she is treated with prophylactic subcutaneous heparin for the balance of her pregnancy. She remains asymptomatic and delivers a healthy, term infant.

Four-year-old John has acute lymphoblastic leukemia and tolerates induction and consolidation chemotherapy well, with minimal side effects. As a key part of his maintenance-treatment protocol, he begins to receive oral mercaptopurine daily, but because a genetic test shows that John is homozygous for a mutation in the gene that encodes thiopurine S-methyltransferase, an enzyme that inactivates mercaptopurine, he receives a greatly reduced dose. Only a few years ago, about 1 in 300 patients had serious, sometimes lethal, hematopoietic adverse effects during mercaptopurine

GLOSSARY

The following terms are used in the text or figures of this article or others in the Genomic Medicine Series. (For a "talking glossary" of many genetics terms, see <http://www.genome.gov/glossary.cfm>.)

Allele — An alternative form of a gene.

Alternative splicing — A regulatory mechanism by which variations in the incorporation of a gene's exons, or coding regions, into messenger RNA lead to the production of more than one related protein, or isoform.

Autosomes — All of the chromosomes except for the sex chromosomes and the mitochondrial chromosome.

Centromere — The constricted region near the center of a chromosome that has a critical role in cell division.

Codon — A three-base sequence of DNA or RNA that specifies a single amino acid.

Conservative mutation — A change in a DNA or RNA sequence that leads to the replacement of one amino acid with a biochemically similar one.

Epigenetic — A term describing nonmutational phenomena, such as methylation and histone modification, that modify the expression of a gene.

Exon — A region of a gene that codes for a protein.

Frame-shift mutation — The addition or deletion of a number of DNA bases that is not a multiple of three, thus causing a shift in the reading frame of the gene. This shift leads to a change in the reading frame of all parts of the gene that are downstream from the mutation, often leading to a premature stop codon and ultimately, to a truncated protein.

Gain-of-function mutation — A mutation that produces a protein that takes on a new or enhanced function.

Genomics — The study of the functions and interactions of all the genes in the genome, including their interactions with environmental factors.

Genotype — A person's genetic makeup, as reflected by his or her DNA sequence.

Haplotype — A group of nearby alleles that are inherited together.

Heterozygous — Having two different alleles at a specific autosomal (or X chromosome in a female) gene locus.

Homozygous — Having two identical alleles at a specific autosomal (or X chromosome in a female) gene locus.

Intron — A region of a gene that does not code for a protein.

Linkage disequilibrium — The nonrandom association in a population of alleles at nearby loci.

Loss-of-function mutation — A mutation that decreases the production or function of a protein (or does both).

Missense mutation — Substitution of a single DNA base that results in a codon that specifies an alternative amino acid.

Monogenic — Caused by a mutation in a single gene.

Motif — A DNA-sequence pattern within a gene that, because of its similarity to sequences in other known genes, suggests a possible function of the gene, its protein product, or both.

Multifactorial — Caused by the interaction of multiple genetic and environmental factors.

Nonconservative mutation — A change in the DNA or RNA sequence that leads to the replacement of one amino acid with a very dissimilar one.

Nonsense mutation — Substitution of a single DNA base that results in a stop codon, thus leading to the truncation of a protein.

Penetrance — The likelihood that a person carrying a particular mutant gene will have an altered phenotype.

Phenotype — The clinical presentation or expression of a specific gene or genes, environmental factors, or both.

Point mutation — The substitution of a single DNA base in the normal DNA sequence.

Regulatory mutation — A mutation in a region of the genome that does not encode a protein but affects the expression of a gene.

Repeat sequence — A stretch of DNA bases that occurs in the genome in multiple identical or closely related copies.

Silent mutation — Substitution of a single DNA base that produces no change in the amino acid sequence of the encoded protein.

Single-nucleotide polymorphism (SNP) — A common variant in the genome sequence; the human genome contains about 10 million SNPs.

Stop codon — A codon that leads to the termination of a protein rather than to the addition of an amino acid. The three stop codons are TGA, TAA, and TAG.

therapy. Although John is in this at-risk minority, a simple genetic test, which is now routine for patients beginning mercaptopurine therapy, alerts his physicians to this genetic predisposition. They reduce his dose of mercaptopurine and carefully monitor his blood levels, ensuring that the drug levels remain therapeutic, rather than toxic. John subsequently has an uneventful several-year maintenance period and achieves complete remission.

THE HUMAN GENOME

These are two examples of genomic medicine, the application of our rapidly expanding knowledge of the human genome (Fig. 1) to medical practice. Much is known, but much remains mysterious. We know that less than 2 percent of the human genome codes for proteins, while over 50 percent represents repeat sequences of several types, whose function is less well understood. These stretches of repetitive sequences, sometimes wrongly dismissed as “junk DNA,” constitute an informative historical record of evolutionary biology, provide a rich source of information for population genetics and medical genetics, and by introducing changes into coding regions, are active agents for change within the genome.²

A draft sequence of the human genome was announced in 2000, and by September 2002 over 90 percent was in final form — that is, there were no gaps

and the data were greater than 99.99 percent accurate. The entire 3.1 gigabases of the DNA sequence should essentially be complete (except for the centromeres and rare unclonable segments) by the spring of 2003. Even with this knowledge in hand, however, we still do not know precisely how many genes the genome contains. Current data indicate that the human genome includes approximately 30,000 to 35,000 genes² — a number that is substantially smaller than was previously thought. Only about half these genes have recognizable motifs, or DNA-sequence patterns, that suggest possible functions. Mutations known to cause disease have been identified in approximately 1000 genes. However, it is likely that nearly all human genes are capable of causing disease if they are altered substantially. Whereas it was once dogma that one gene makes one protein, it now appears that, through the mechanism of alternative splicing¹² (Fig. 2), more than 100,000 proteins can be derived from these 30,000 to 35,000 genes.²

In addition to alternative splicing, a number of “epigenetic” phenomena, such as methylation and histone modification, can alter the effect of a gene. Furthermore, a complex array of molecular signals allows specific genes to be “turned on” (expressed) or “turned off” in specific tissues and at specific times.

Genes are distributed unevenly across the human genome (Fig. 1). Certain chromosomes, particularly

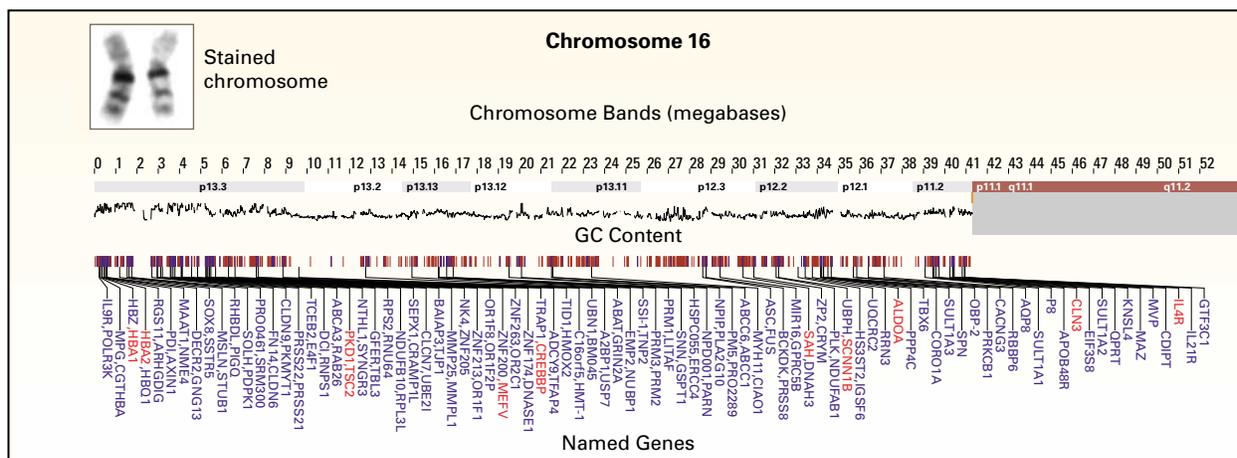


Figure 1. A Depiction of Chromosome 16 Based on the Determination of Its Actual Sequence by the Human Genome Project. The inset shows a Giemsa-stained chromosome. The figure shows a scale in megabases (1 megabase equals 1 million base pairs); the approximate positions of Giemsa-stained chromosome bands at a resolution of 800 bands; the proportion of bases in a 20,000-base window that are either guanine or cytosine (the GC content); the location of predicted genes; and the locations of named genes that have been located in the draft sequence (known disease genes in the Online Mendelian Inheritance in Man [OMIM] data base¹¹ are in red, and other genes in blue). Only about 30 percent of human genes have been named thus far. There is considerable variation in gene density across the chromosome, particularly in the dark band 16q21 on the right, which has an extremely low gene density and GC content. The rectangular dark-gray block present on both pages represents the centromere and adjacent repetitive sequence on 16q. Modified from the work of the International Human Genome Sequencing Consortium.²

17, 19, and 22, are relatively gene dense as compared with others, such as 4, 8, 13, 18, and Y.³ Moreover, gene density varies within each chromosome, being highest in areas rich in the bases cytosine and guanine, rather than adenine and thymine.^{2,3} Chromosomes 13, 18, and 21, the three autosomes with the fewest genes, are also the three for which the occurrence of trisomy (i.e., three copies of a chromosome) is compatible with viability.

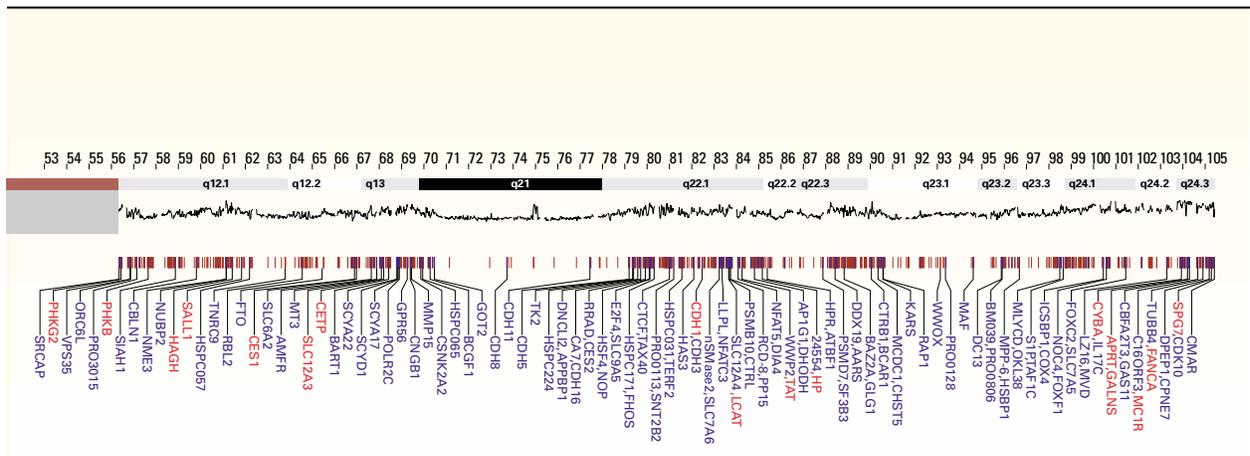
Not all genes reside on nuclear chromosomes; several dozen involved with energy metabolism are on the mitochondrial chromosome.¹³ Since ova are rich in mitochondria and sperm are not, mitochondrial DNA is usually inherited from the mother. Therefore, mitochondrial genes — and diseases due to DNA-sequence variants in them — are transmitted in a matrilineal pattern that is distinctly different from the pattern of inheritance of nuclear genes.

MONOGENIC CONDITIONS

Over the course of the 20th century, a combination of theoretical insights, basic-science research, and clinical observation elucidated the inheritance of single-gene, or monogenic, disorders (also known as mendelian disorders, since they are transmitted in a manner consonant with Mendel's laws of inheritance). Modes of inheritance have been established for thousands of conditions caused by mutations in single genes; these

have been catalogued in a textbook¹¹ and, more recently, in an online compendium¹⁴ known as Mendelian Inheritance in Man (OMIM). For nearly 100 years, autosomal dominant, autosomal recessive, and X-linked modes of inheritance have been understood and known to cause human disease. In the past few decades, other mechanisms of monogenic inheritance have been described. These include mitochondrial inheritance,¹³ imprinting (a mechanism by which the effects of certain genes depend on whether they are inherited through the mother or through the father),¹⁵ uniparental disomy (the occasional situation in which both members of 1 pair of a person's 23 pairs of chromosomes derive from one parent),¹⁶ and expanding trinucleotide repeats (a phenomenon in which a sequence of three base pairs that is normally repeated a number of times in a row in the genome becomes repeated by more than the normal number of times, sometimes causing disease).¹⁷

Most single-gene conditions are uncommon. Even the commonest, such as hereditary hemochromatosis (approximate incidence, 1 in 300 persons), cystic fibrosis (approximate incidence, 1 in 3000), alpha₁-antitrypsin deficiency (approximate incidence, 1 in 1700), or neurofibromatosis (approximate incidence, 1 in 3000), affect no more than 1 in several hundred people in the United States. However, the total effect of monogenic conditions is substantial, from both the



individual patient's and public health perspectives, and increased understanding of genetics has already begun to improve the health of some patients with such conditions. The delineation of the mechanisms by which genetic factors cause monogenic disorders has provided important information about basic pathophysiological processes that underlie related disorders that occur with far greater frequency than do these genetic disorders. For instance, insights regarding familial hypercholesterolemia, a genetic disorder that affects only 1 of every 500 people in the United States, were instrumental to understanding the pathophysiology of atherosclerosis, which affects a large fraction of the population, and the development of the statin drugs, which are among the most frequently prescribed medications.¹⁸

TYPES OF MUTATION

There are a number of ways to categorize mutations. One is according to the causative mechanism, whereas another is according to their functional effect. When classified according to the mechanism, point muta-

tions — that is, a change in a single DNA base in the sequence — are the most common. There are many types of point mutations. One type is a missense mutation (Fig. 3), a substitution that leads to an alternative amino acid, because of the way in which it changes the three-base sequence, or codon, that codes for an amino acid. Nonsense mutations (Fig. 3) are a more dramatically deleterious type of point mutation that change the codon to a “stop” codon, a codon that causes the termination of the protein instead of producing an amino acid. Another type of mutation is the frame-shift mutation (Fig. 3), which changes the reading frame of the gene downstream from it, often leading to a premature stop codon.

In terms of functional effect, rather than mechanism, many variants in the human-genome sequence have no phenotypic effect. Among these are silent mutations (Fig. 3), which replace one base with another, so that the resultant codon still codes for the same amino acid. Also, mutations may not change the phenotype if the altered codon substitutes one amino acid for another that produces little change in the

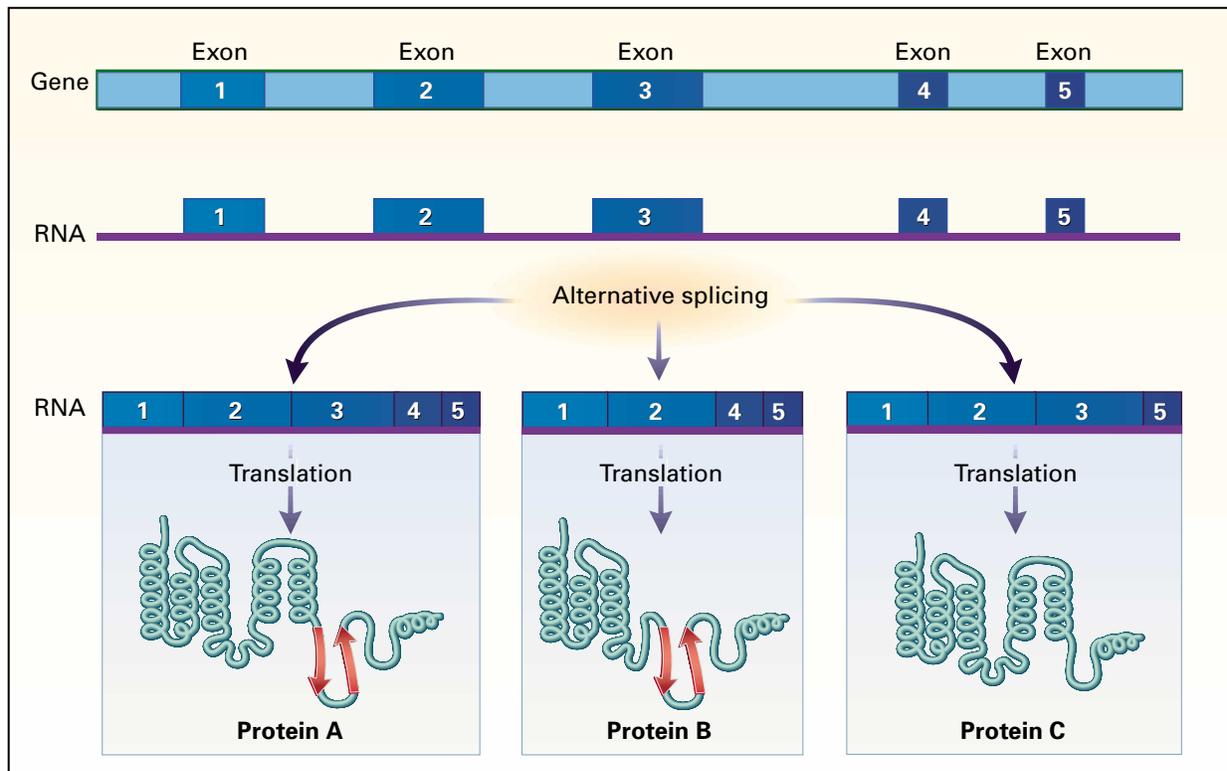


Figure 2. Alternative Splicing. A single gene can produce multiple related proteins, or isoforms, by means of alternative splicing.

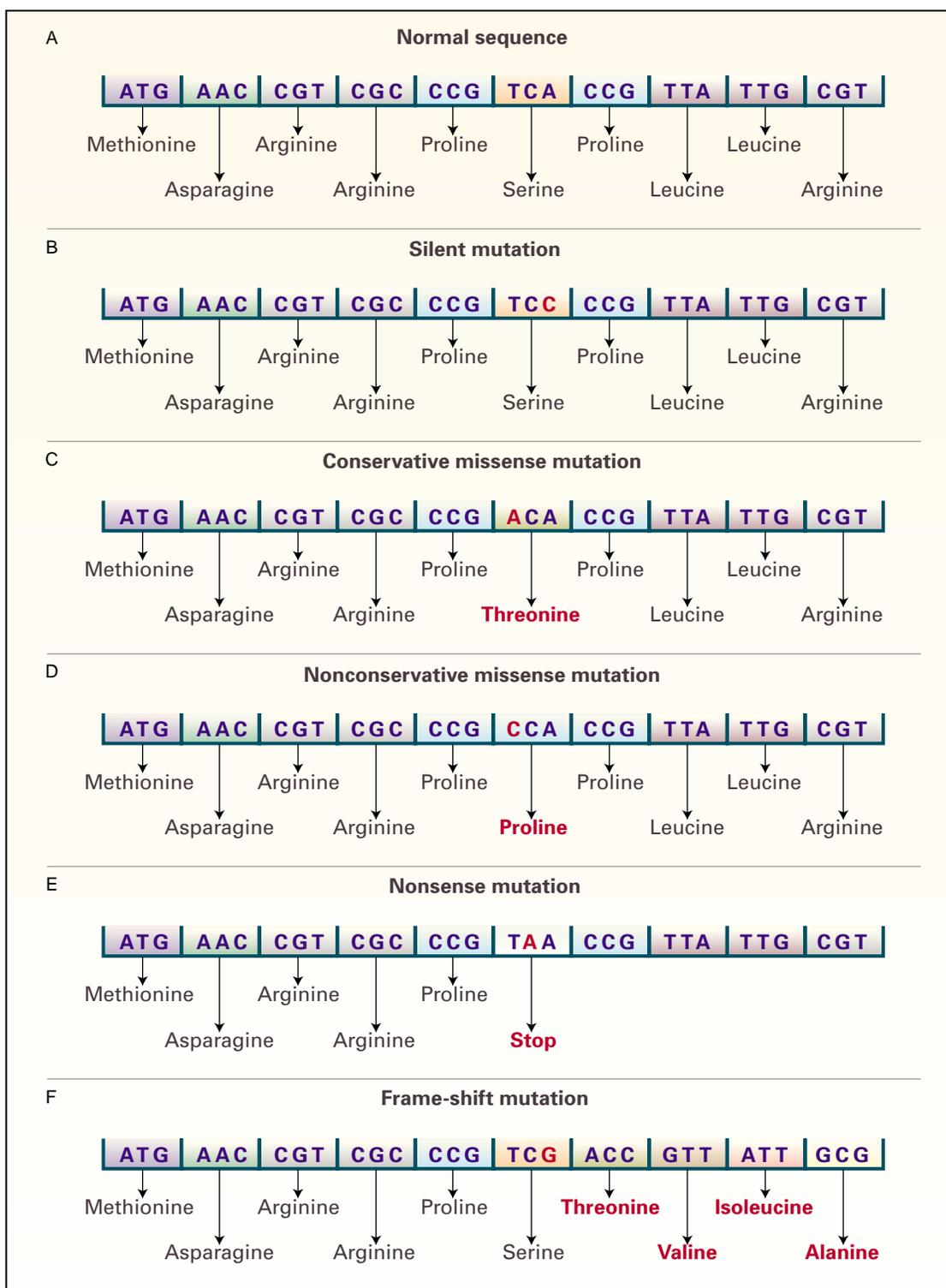


Figure 3. Examples of Point Mutations.

Panel A shows the normal sequence of DNA from one exon and the protein product it encodes. Panel B shows a silent mutation, Panel C a conservative missense mutation (serine and threonine have very similar structures), Panel D a nonconservative missense mutation (serine and proline have very different structures), Panel E a nonsense mutation, and Panel F a frame-shift mutation. In Panel F, the insertion of a single G throws off the reading frame, so that all amino acids downstream are changed radically.

function of the protein or proteins that the gene encodes. These are referred to as “conservative mutations” (Fig. 3). Nonconservative mutations (Fig. 3) replace an amino acid with a very different one and are more likely to affect the phenotype.

Although mutations can cause disease by a variety of means, the most common is loss of function. Loss-of-function mutations alter the phenotype by decreasing the quantity or the functional activity of a protein. For instance, mutations in the glucose-6-phosphate dehydrogenase (*G6PD*) gene on the X chromosome decrease the functional activity of this enzyme, leading to acute hemolytic anemia if a male (who would, of course, have only one copy of the X chromosome) with the mutation is exposed to certain drugs, including sulfonamides, primaquine, and nitrofurantoin. Since genes do not exist just to handle pharmacologic agents, variants that cause a more severe deficiency of glucose-6-phosphate dehydrogenase also lead to hemolytic anemia when affected males ingest fava beans (favism), since this enzyme is also important in the degradation of a component of the beans.¹⁹

Some mutations cause disease through a gain of function, whereby the protein takes on some new, toxic function. Expanding exonic CAG trinucleotide repeats that cause disorders including Huntington’s disease and spinocerebellar ataxia appear to lead to neuropathologic abnormalities by producing proteins that function abnormally because of expanded polyglutamine tracts (CAG codes for the amino acid glutamine).²⁰ Such gain-of-function mutations are often dominantly inherited, since a single copy of the mutant gene can alter function.

One might assume that mutations in the approximately 98.5 percent of the genome that does not code for proteins do not affect the phenotype. Indeed, most do not. But others are regulatory mutations that may ultimately prove as important in the etiologic process of common diseases as the coding region variants. Regulatory mutations act by altering the expression of a gene. For instance, a regulatory mutation might lead to the loss of expression of a gene, to unexpected expression in a tissue in which it is usually silent, or to a change in the time at which it is expressed. Examples of regulatory mutations associated with disease are those in the flanking region of the *FMR1* gene (causing fragile X syndrome),²¹ the insulin gene flanking region (increasing the risk of type 1 diabetes mellitus),²² a regulatory site of the type I collagen gene (increasing the risk of osteoporosis),²³ and an intronic regulatory site of the calpain-10 gene (increasing the risk of type 2 diabetes mellitus).²⁴

Mutations can also decrease the risk of a disease. One example of this is a 32-bp deletion (a frame shift) in a chemokine receptor gene, *CCR5*. Persons who are homozygous for this deletion prove almost com-

pletely resistant to infection with HIV type 1, and those who are heterozygous for the deletion have slower progression from infection to AIDS. These effects arise because *CCR5* is an important part of the mechanism by which HIV enters the cell.²⁵

GENES IN COMMON DISEASE

The study of genomics will most likely make its greatest contribution to health by revealing mechanisms of common, complex diseases, such as hypertension, diabetes, and asthma. So far, most genes involved in common diseases have been identified by virtue of their high penetrance — that is, the mutations lead to disease in a fairly large proportion of people who have them. Examples include mutations in *BRCA1* and *BRCA2* (increasing the risk of breast and ovarian cancer),²⁶ *HNPCC* (increasing the risk of hereditary nonpolyposis colorectal cancer),⁶ *MODY 1*, *MODY 2*, and *MODY 3* (increasing the risk of diabetes),²⁷ and the gene for α -synuclein (causing Parkinson’s disease).⁹ One can think of these as near-mendelian subgroups of disease within a larger group of affected persons. If a person has such a mutation, the likelihood of disease is great. However, each of these highly penetrant mutations associated with common disease has a prevalence in the general population of only one in several hundred to several thousand people.

From a public health perspective, genes with mutations that are less highly penetrant but much more prevalent have a greater effect on the population than genes that are highly penetrant but uncommon. Such mutations have been reported in genes such as *APC* (which increases the risk of colorectal cancer)²⁸ and factor V Leiden (which increases the risk of thrombosis).²⁹

An example of the relative contributions of rare, highly penetrant mutations as opposed to common, less penetrant ones is seen in Alzheimer’s disease. Rare mutations in presenilin 1, presenilin 2, or the β -amyloid precursor protein gene are highly penetrant causes of early-onset Alzheimer’s disease; indeed, Alzheimer’s disease develops by the age of 60 years in most people who are heterozygous for a mutation in one of these genes.³⁰ However, because so few people carry a mutation in any of these genes, these mutations play a part in fewer than 1 percent of cases of Alzheimer’s disease.³¹ In contrast, the apolipoprotein E $\epsilon 4$ allele also increases the risk of late-onset Alzheimer’s disease (and atherosclerosis³²), but more subtly. One representative Finnish study found that Alzheimer’s disease develops during the mid-70s in approximately 8 percent of persons who are heterozygous for the $\epsilon 4$ allele and 21 percent of those who are homozygous for it, as compared with 3 percent of those with no $\epsilon 4$ allele.³³ Nonetheless, because approximately 26 percent of the U.S. population is heterozygous and 2 percent

is homozygous³⁴ for the apolipoprotein E ϵ 4 allele, this genetic factor has a role in many more cases of Alzheimer's disease than do the mutations in the genes for presenilin 1, presenilin 2, and β -amyloid precursor protein combined.

VARIATION IN THE HUMAN GENOME

One characteristic of the human genome with medical and social relevance is that, on average, two unrelated persons share over 99.9 percent of their DNA sequences.³ However, given the more than 3 billion base pairs that constitute the human genome, this also means that the DNA sequences of two unrelated humans vary at millions of bases. Since a person's genotype represents the blending of parental genotypes, we are each thus heterozygous at about 3 million bases. Many efforts are currently under way, in both the academic and commercial sectors, to catalogue these variants, commonly referred to as "single-nucleotide polymorphisms" (SNPs), and to correlate these specific genotypic variations with specific phenotypic variations relevant to health.

Some SNP–phenotype correlations occur as a direct result of the influence of the SNP on health. More commonly, however, the SNP is merely a marker of biologic diversity that happens to correlate with health because of its proximity to the genetic factor that is actually the cause. In this sense, the term "proximity" is only a rough measure of physical closeness; instead, it connotes that, as genetic material has passed through 5000 generations from our common African ancestral pool, recombination between the SNP and the actual genetic factor has occurred only rarely. In genetic terms, the SNP and the actual genetic factor are said to be in linkage disequilibrium (Fig. 4).

An extension of the current efforts to catalogue SNPs and correlate them to phenotype are efforts to map and use haplotypes. Whereas a SNP represents a single-nucleotide variant, a haplotype represents a considerably longer sequence of nucleotides (averaging about 25,000), as well as any variants, that tend to be inherited together. SNPs and haplotypes will be the key to the association studies (i.e., studies of affected persons and control subjects) necessary to identify the genetic factors in complex, common diseases, just as family studies have been important to the identification of the genes involved in monogenic conditions.^{35,36} Also, at least until whole-genome sequencing of individual patients becomes feasible clinically, the identification of SNPs and haplotypes will prove instrumental in efforts to use genomic medicine to individualize health care.

CONCLUSIONS

Except for monozygotic twins, each person's genome is unique. All physicians will soon need to un-

derstand the concept of genetic variability, its interactions with the environment, and its implications for patient care. With the sequencing of the human genome only months from its finish, the practice of medicine has now entered an era in which the individual patient's genome will help determine the optimal ap-

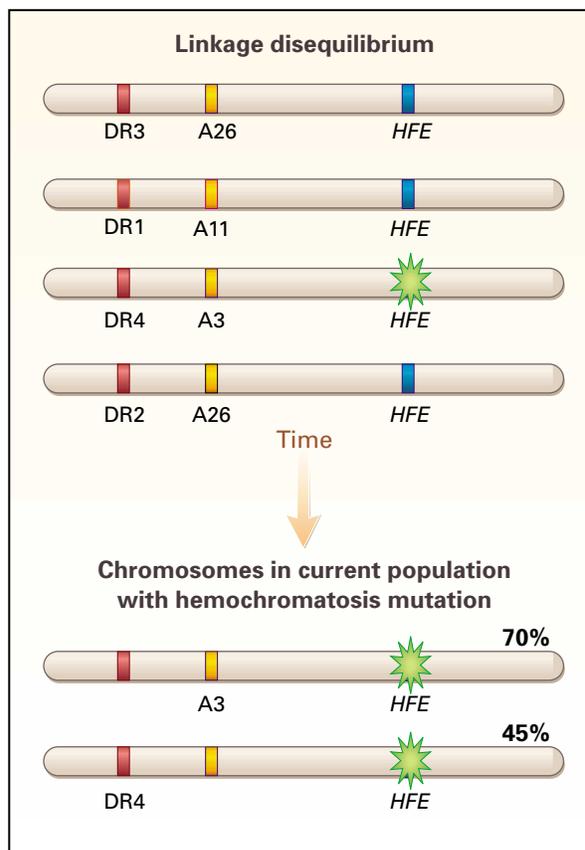


Figure 4. Example of Linkage Disequilibrium.

As shown in the top panel, several versions of chromosome 6 existed in a specific population a number of generations ago. A mutation (indicated by the green star) in the hemochromatosis gene (*HFE*) originates in an ancestral chromosome that also carried the HLA-A3 and DR4 alleles. For several subsequent generations, nearly all chromosomes carrying the *HFE* mutation also had the HLA-A3 and DR4 alleles. As shown in the bottom panel, over time, recombination between the *HFE* and HLA-A alleles occurred rarely, so that in the current population, the *HFE* mutation is associated with the HLA-A3 allele 70 percent of the time, even though the HLA-A3 allele occurs on only 15 percent of normal chromosomes in this population. Because the HLA-DR locus is farther away from the *HFE* locus than is the HLA-A locus, recombination between the *HFE* and HLA-DR loci has occurred more frequently than between the *HFE* and the HLA-A loci, so that the *HFE* mutation is now associated with the HLA-DR4 allele 45 percent of the time, even though the HLA-DR4 allele occurs on only 25 percent of normal chromosomes in this population. The *HFE* mutation is said to be in strong linkage disequilibrium with HLA-A3 and somewhat weaker linkage disequilibrium with HLA-DR4.

proach to care, whether it is preventive, diagnostic, or therapeutic. Genomics, which has quickly emerged as the central basic science of biomedical research, is poised to take center stage in clinical medicine as well.

Editor's note: As they are published, the articles in the Genomic Medicine Series will be available without charge at <http://www.nejm.org>.

REFERENCES

1. Adams FL. The genuine works of Hippocrates. Vol. 2. New York: William Wood, 1886:338.
2. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921. [Errata, *Nature* 2001; 411:720, 412:565.]
3. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304-51. [Erratum, *Science* 2001;292:1838.]
4. McKusick VA, Ruddle FH. A new discipline, a new name, a new journal. *Genomics* 1987;1:1-2.
5. Armstrong K, Eisen A, Weber B. Assessing the risk of breast cancer. *N Engl J Med* 2000;342:564-71.
6. Lynch HT. Hereditary nonpolyposis colorectal cancer (HNPCC). *Cytogenet Cell Genet* 1999;86:130-5.
7. Michael NL. Host genetic influences on HIV-1 pathogenesis. *Curr Opin Immunol* 1999;11:466-74.
8. Small PM, Fujiwara PI. Management of tuberculosis in the United States. *N Engl J Med* 2001;345:189-200.
9. Mouradian MM. Recent advances in the genetics and pathogenesis of Parkinson disease. *Neurology* 2002;58:179-85.
10. St George-Hyslop PH. Molecular genetics of Alzheimer disease. *Semin Neurol* 1999;19:371-83.
11. McKusick VA. Mendelian inheritance in man: catalogs of human genes and genetic disorders. 12th ed. Baltimore: Johns Hopkins University Press, 1998.
12. Graveley BR. Alternative splicing: increased diversity in the proteomic world. *Trends Genet* 2001;17:100-7.
13. Wallace DC. Mitochondrial diseases in man and mouse. *Science* 1999; 283:1482-8.
14. Online Mendelian Inheritance in Man, OMIM (TM). Baltimore: McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University, 2000. (Accessed October 15, 2002, at <http://www.ncbi.nlm.nih.gov/omim/>.)
15. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001;2:21-32.
16. Robinson WP. Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* 2000;22:452-9.
17. Sinden RR. Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am J Hum Genet* 1999;64:346-53.
18. Goldstein JL, Hobbs HH, Brown MS. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. The metabolic & molecular bases of inherited disease. 8th ed. Vol. 2. New York: McGraw-Hill, 2001:2863-913.
19. Luzzatto L, Mehta A, Vulliamy T. Glucose 6-phosphate dehydrogenase deficiency. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. The metabolic & molecular bases of inherited disease. 8th ed. Vol. 3. New York: McGraw-Hill, 2001:4517-53.
20. Goldberg YP, Nicholson DW, Rasper DM, et al. Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nat Genet* 1996;13:442-9.
21. Jin P, Warren ST. Understanding the molecular basis of fragile X syndrome. *Hum Mol Genet* 2000;9:901-8.
22. Todd JA. From genome to aetiology in a multifactorial disease, type 1 diabetes. *Bioessays* 1999;21:164-74.
23. Grant SFA, Reid DM, Blake G, Herd R, Fogelman I, Ralston SH. Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I α 1 gene. *Nat Genet* 1996;14:203-5.
24. Horikawa Y, Oda N, Cox NJ, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000;26:163-75. [Erratum, *Nat Genet* 2000;26:502.]
25. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. *Science* 1996;273:1856-62. [Erratum, 1996;274:1069.]
26. Nathanson KN, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med* 2001;7:552-6. [Erratum, *Nat Med* 2001;7:749.]
27. Froguel P, Velho G. Molecular genetics of maturity-onset diabetes of the young. *Trends Endocrinol Metab* 1999;10:142-6.
28. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. *Hum Mol Genet* 2001;10:721-33.
29. Major DA, Sane DC, Herrington DM. Cardiovascular implications of the factor V Leiden mutation. *Am Heart J* 2000;140:189-95.
30. Campion D, Dumanchin C, Hannequin D, et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet* 1999;65:664-70.
31. Roses AD. Alzheimer diseases: a model of gene mutations and susceptibility polymorphisms for complex psychiatric diseases. *Am J Med Genet* 1998;81:49-57.
32. Curtiss LK, Boisvert WA. Apolipoprotein E and atherosclerosis. *Curr Opin Lipidol* 2000;11:243-51.
33. Kuusisto J, Koivisto K, Kervinen K, et al. Association of apolipoprotein E phenotypes with late onset Alzheimer's disease: population based study. *BMJ* 1994;309:636-8.
34. Roses AD. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu Rev Med* 1996;47:387-400.
35. Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature* 2001;411:199-204.
36. Goldstein DB, Weale ME. Population genetics: linkage disequilibrium holds the key. *Curr Biol* 2001;11:R576-R579.

Copyright © 2002 Massachusetts Medical Society.

APPLY FOR JOBS ELECTRONICALLY AT THE NEW NEJM CAREER CENTER

Physicians registered at the new NEJM Career Center can now apply for jobs electronically using their own cover letters and CVs. You can now keep track of your job-application history with a personal account that is created when you register with the Career Center and apply for jobs seen online at our Web site. Visit www.nejmjobs.org for more information.
