Annu. Rev. Pharmacol. Toxicol. 2000. 40:97–132

SEQUENCING THE ENTIRE GENOMES OF FREE-LIVING ORGANISMS: The Foundation of Pharmacology in the New Millennium

Samuel Broder and J. Craig Venter
Celera Genomics, Rockville, Maryland 20850; e-mail: BroderSE@celera.com, jcventer@celera.com

Key Words human genomics, comparative genomics, shotgun sequencing, single nucleotide polymorphism (SNP), DNA variation, pharmacogenetics

■ **Abstract** The power and effectiveness of clinical pharmacology are about to be transformed with a speed that earlier in this decade could not have been foreseen even by the most astute visionaries. In the very near future, we will have at our disposal the reference DNA sequence for the entire human genome, estimated to contain approximately 3.5 billion bp. At the same time, the science of whole genome sequencing is fostering the computational science of bioinformatics needed to develop practical applications for pharmacology and toxicology. Indeed, it is likely that pharmacology, toxicology, bioinformatics, and genomics will merge into a new branch of medical science for studying and developing pharmaceuticals from molecule to bedside.

THE IMPENDING AVAILABILITY OF A COMPLETE REFERENCE SEQUENCE FOR THE HUMAN GENOME

Our DNA sequence and its variation provide a special record of human evolution and the migration of populations (1–14). We will learn how this sequence varies among populations and among individuals, including the role of such variation in the pathogenesis of important illnesses and responses to pharmaceuticals. We will localize and annotate every human gene and the regulatory elements that control the timing, tissue-site specificity, and extent of gene expression. For any given physiologic process, we will have a new paradigm for addressing its evolution, its development, its function, and its mechanism. This will revitalize medicine by identifying important new targets for prevention, diagnosis, and therapy. Clinical pharmacologists will be able to approach issues of rational candidate drug design and the reduction of serious side effects by using bioinformatics to analyze the relevant genes and gene variations (polymorphisms), including the promoters and the enhancers, involved. Knowledge regarding the alleles that gov-

0362–1642/00/0415–0097$14.00

ern the safety and efficacy of pharmaceutical agents (including comparative genomics) will make it possible to streamline the preclinical and clinical development of new drugs and customize interventions to the specific genotypes of patients. Medical progress will be driven more by knowledge of gene structure and function and less by empiricism and intuition.

It is important to recall that the first complete genome of any free-living organism (*Haemophilus influenzae*) was published by scientists at the Institute for Genomic Research roughly 5 years ago (15). During the past 5 years, the sequences of the entire genomes of 23 organisms have been published (Table 1; see also http://www.tigr.org/tdb/mdb/mdb.html) (15–37). What was, a few short years ago, thought to be impossible has become not merely possible but inevitable.

WHOLE GENOME SEQUENCING

What has led us to such a future? The development of advanced automation, robotics, and computer software for industrial-scale DNA sequencing has proceeded at a remarkable pace. With the successful sequencing of the *H. influenzae* genome in its entirety, it became clear that the DNA of entire complex organisms many megabases in size could be accurately and rapidly sequenced by using a ''shotgun'' sequencing strategy (14).

In this strategy, a single random DNA-fragment library is prepared following mechanical or sonic shearing of entire genomic DNA and inserted in suitable vector systems (e.g. plasmids). The ends of a large number of randomly selected fragments are sequenced from both insert ends until every part of the genome has been sequenced several times on average. For any given average sequence read length, the number of end sequences needed can be determined by the Lander and Waterman application of Poisson statistics. This number will depend on the goals of the sequencing project and particularly the degree of tolerance for a small number of gaps in the sequence results (i.e. if the tolerance for gaps is low, the number of end sequences must be high). The sequences are then computationally ''reassembled'' to provide the complete genome. For higher organisms, which contain one genome from the mother and one genome from the father, this approach yields an important dividend: Points of common DNA variation such as single nucleotide polymorphisms (to which we return at several points in this chapter) become evident.

This overall strategy, coupled with the advent of completely automated DNA sequencing machines such as the new ABI Prism 3700 DNA Analyzer (manufactured by PE Biosystems, PE Corp.), will make it possible for a single center to undertake the determination of the reference sequence of the human genome. Indeed, it gives one pause to consider that every organism of pharmacologic or toxicologic interest is now plausibly a candidate for whole genome sequencing.

TABLE 1 Published free-living organism genomes

Genome	Strain	Domain	Size (MB)	Institution	Ref.
Haemophilus influenzae Rd	KW20	B	1.83	Inst. Genomic Res.	15
Mycoplasma genitalium	G-37	B	0.58	Inst. Genomic Res.	16
Methanococcus jannaschii	DSM 2661	A	1.66	Inst. Genomic Res.	17
Synechocystis sp.	PCC 6803	B	3.57	Kazusa DNA Res. Inst.	18
Mycoplasma pneumoniae	M129	B	0.81	Univ. Heidelberg	19
Saccharomyces cerevisiae	S288C	E	13	Int. Consort.	20
Helicobacter pylori	26695	B	1.66	Inst. Genomic Res.	21
Escherichia coli	K-12	B	4.60	Univ. Wisc.	22
Methanobacterium thermoautotrophicum	Delta H	A	1.75	Genome Ther. & Ohio State Univ.	23
Bacillus subtilis	168	B	4.20	Int. Consort	24
Archaeoglobus fulgidus	DSM4304	A	2.18	Inst. Genomic Res.	25
Borrelia burgdorferi	B31	B	1.44	Inst. Genomic Res.	26
Aquifex aeolicus	VF5	B	1.50	Diversa	27
Pyrcoccus horikoshii	OT3	A	1.80	Natl. Inst. of Tech. and Eval. (Japan)	28
Mycobacterium tuberculosis	H37Rv (lab strain)	B	4.40	Sanger Cent.	29
Treponema pallidum	Nichols	B	1.14	Inst. Genomic Res./Univ. Texas	30
Chlamydia trachomatis	Serovar D (D/UW-3/ Cx)	B	1.05	Univ. Calif. Berkeley/Stanford	31
Rickettsia prowazekii	Madrid E	B	1.10	Univ. Uppsala	32
Caenorhabditis elegans			100	Wash. Univ./Sanger Cent.	33
Helicobacter pylori	J99	B	1.64	Astra Res. Cent. Boston/Genome Ther.	34
Chlamydia pneumoniae	CWL029	B	1.23	Univ. Calif. Berkeley/Stanford	35
Thermotoga maritima	MSB8	B	1.80	Inst. Genomic Res.	36
Aeropyrum pernix	K1	A	1.67	Natl. Inst. of Tech. and Eval. (Japan)	37

This is not a vision for the future or a promissory note as to what science may someday bring. The science and supporting technology are here now.

THE CHALLENGE OF MICROBIAL PATHOGENS

Microbial pathogens (e.g. tuberculosis, cholera, and malaria) are a source of great suffering and death in many developing countries. Moreover, even in nations with advanced health-care technologies and mature research-based pharmaceutical industries, the emergence of multidrug-resistant pathogens is a serious problem. Indeed, some have argued that we are not too far removed from a return to the prepenicillin era in the fight against infectious disease. The problem for research pharmacologists is formidable and getting worse. The evolving field of genomic sequencing allows unique opportunities for understanding microbial pathogenesis and control by having simultaneously at hand the genomes of the pathogen and host.

Knowing the complete genome of microbial pathogens will open up exciting opportunities to develop novel pharmaceuticals and biologics. We will know how many genes are contained in each pathogen, where they are located within that pathogen's genome, and when two or more similar genes (paralogs) exist in a single microbial genome, thereby creating the potential to confound the research pharmacologist's search for an Achilles heel. By simultaneously analyzing the microbial genome and the host genome, it will become possible to define which genes are critical for microbial survival and which are optional; why a given pathogen is virulent in the context of a specific host; how and when toxic cytokines are activated within a host; whether a given pathogen has evolved proteins with molecular mimicry capable of frustrating host immunity or inducing autoimmunity; and how a pathogen (e.g. the tubercle bacillus) is able to survive in a state of latency or dormancy, impervious to the host's immune system. The sum total of this information will make it possible to define vaccines that induce specific and effective immunity against the pathogen while minimizing untoward or toxic side effects in the host.

Malaria provides a model for the challenges and research opportunities, and Wahlgren (38) has provided an elegant perspective on this topic. Falciparum malaria is a leading killer in the African countryside, and multidrug-resistant forms of the disease have emerged on a wide scale. The disease is caused by a protozoan (*Plasmodium falciparum*), whose life cycle involves infecting human erythrocytes, which in turn causes a fulminant hemolytic anemia. One consequence is cerebral malaria, a process facilitated by the patient's own cytokines, which promote adhesion of malarial organisms or their detritus to the inside walls of blood vessels. The resulting cerebral ischemia is often lethal, and perhaps one million people, many of them children, die from this infectious kind of stroke every year.

The best known genes for malarial protection are those for which resistance alleles (nucleotide substitutions) in people in endemic regions exist in a state of balanced polymorphism. Thus, people with one copy of the S allele of the beta subunit of hemoglobin have a selective advantage living in malaria-endemic zones compared with individuals with two wild-type alleles, or with two copies of the mutant allele (in which case they would exhibit sickle cell anemia). The α- and β-thalassemias and glucose-6-phosphate dehydrogenase deficiencies are other examples of genotype polymorphisms whose evolution was driven by the selective advantage conferred on account of resistance to malaria.

More recently, certain cytokines, most notably tumor necrosis factor (TNF), have been implicated in the pathogenesis of malaria. TNF also probably plays a significant role in leishmaniasis, listeriosis, and other infectious diseases of the developing world. High circulating levels of TNF have been found in patients with cerebral malaria, particularly those whose disease runs a lethal course. TNF, among its many activities, up-regulates endothelial adhesion molecules and thereby increases the tendency of infected red cells to stick to the walls of blood vessels and interrupt blood flow. Knight et al (39) recently found that three polymorphisms in the promoter region of the *TNF* gene contribute to malarial pathogenesis and poor outcome through a complex dynamic of increased risk and counterbalancing protective effects. One such polymorphism, TNF_{-376A}, increases the secretion of TNF most likely by causing the helix-turn-helix transcription factor OCT-1 to bind to a novel region of complex protein-DNA interactions and alter gene expression in human monocytes. These workers found that a single nucleotide $G \rightarrow A$ polymorphism at position 376 upstream of the *TNF* transcriptional initiation site (the OCT-1 binding phenotype) is linked to unfavorable outcome from malarial infection in two ethnically distinct groups of people, one from the Gambia (western African) and the other from Kenya (eastern African). The results underscore the potential to link molecular events to clinical outcomes using the emerging knowledge of whole organism genomics. They are likely to influence future research into the prevention and treatment of malaria and many other diseases.

The *P. falciparum* genome is roughly 30 Mb in size and contains 14 chromosomes. Recently, after overcoming a number of theoretical and practical obstacles owing to the adenine plus thymine richness of its genome, it became possible to obtain the entire genomic DNA sequence of chromosome 2 in *P. falciparum* (40). Thus, one can expect that in the foreseeable future, the technologies for whole genome sequencing of microbial agents will make it possible to sequence the remaining 13 chromosomes and add this pathogen to the list in Table 1. The coupling of whole genome information from *P. falciparum* with whole genome information from its human host is likely to initiate an entire range of new approaches for dealing with the problem of malaria. (It is also possible that DNA sequencing of the intermediary host mosquito or a related insect genome may add still another set of strategic targets for attacking malaria.) The development of

knowledge useful in turning the tide against multidrug-resistant malaria is a crucial goal for the science of whole genome sequencing.

DNA SEQUENCE VARIATION

We all evolved in an African savanna, and we all share >99.9% of the nucleotide sequence in our genome, so it is remarkable that the extraordinary diversity of human beings is encoded by only 0.1% variation in our DNA. We are predisposed to different diseases, we respond to the environment in variable ways, we metabolize pharmaceutical agents differently, we may show differences in dose-response relationships for common drugs, and we have a range of susceptibilities to adverse side effects from therapeutic agents (even when there is no discernible difference in individual pharmacokinetics or biochemical pharmacology). Despite the overwhelming similarity in sequence, there are millions of points of DNA variation between any two randomly selected individuals.

The most common form of DNA variation is single nucleotide polymorphism (SNP) (1–14, 41), and in our discussion of malaria above we encountered the enormous implications of SNPs for pharmacology and toxicology. Put simply, a SNP is the substitution of one purine or pyrimidine nucleotide at a given location in a strand of DNA for another purine or pyrimidine nucleotide. Such substitutions can affect gene function, or they can be neutral. Neutrality is generally inferred if a SNP does not alter protein coding. In practice, this inference can be wrong. The most common substitution is a transition in which a pyrimidine is substituted for another pyrimidine, and likewise a purine is substituted for a purine. However, transversions (replacement of a purine for a pyrimidine or vice versa) can occur.

The nomenclature defining a mutation (a disease-causing change) versus a SNP is arbitrary and relative. By convention, when a substitution is present in more that 1% of a target population, it is called a variant or polymorphism. When a substitution is present in less than 1%, and especially when one can assign a clear phenotype (i.e. a disease or clinical condition), it is called a mutation. SNPs may occur across widely separated populations, or they may be relatively specific for a given population. They are virtually always biallelic. Within populations, in both human and model organisms such as *Drosophila*, essentially every SNP is in Hardy-Weinberg equilibrium proportions (A Clark, personal communication).

SNPs may occur inside or outside of a gene. If they occur within a gene, they may reside in an exon (coding region) or intron (noncoding) region. SNPs in a coding region (sometimes called cSNPs) can be either synonymous (no amino acid–altering effect) or nonsynonymous (amino acid altering). There is some level of natural selection against amino acid–altering changes (12, 13). The average person would be expected to be heterozygous for roughly 40,000 nonsynonymous (amino acid-altering) alleles (12).

As we have seen in the case of the TNF_{-376} allele, SNPs can profoundly affect gene function even if they are at a significant distance upstream of the initiation

site for gene transcription. One needs to keep in mind that enhancers (i.e. the tissue-specific control sequences on which certain regulatory substances act) may operate over at least 3 kb in either orientation ($5' \rightarrow 3'$ or $3' \rightarrow 5'$) from the start point of transcription.

It has been proposed that those SNPs that are of keenest interest in the pathophysiology of disease are nonsynonymous cSNPs, which alter the sequence of an encoded protein (12). Coding sequences comprise roughly 3% of the human genome, and therefore, it has been suggested that priority be given to shotgun sequencing of ccDNA libraries from many donors, possibly with an emphasis on specific candidate genes by direct amplification of target sequences. A word of caution is in order. It may not be appropriate to discount synonymous SNPs in the pathogenesis of important illnesses. Correlations between SNPs and clinical phenotypes are in their infancy, and it is critical that we do not assume that we know more than we do. A SNP that is ''synonymous'' for a protein coding point of view might be ''antonymous'' with respect to the folding of mRNA, the enzymatic activity on certain RNA molecules (e.g. adenosine to inosine conversion), and the function of small nucleolar RNA(snoRNA). This also applies to the set of genes involved in antisense RNA regulation, such as *Tsix,* a gene that performs antisense regulatory functions against *Xist* at the X-inactivation center (42). It is important to recall that splicing variants can also be significantly affected by DNA variation seen only at a genomic level. Moreover, the genes that may be of greatest interest in understanding the pathogenesis of common illnesses or in the development of important new drugs (particularly those that bring about true paradigmatic shifts in therapy) may be dramatically underrepresented (or undetectable) in cDNA libraries. Rarely expressed genes, or genes expressed in an extremely limited range of cells, may be poor candidates for cDNA-based detection systems. Looking for SNPs solely within or near known candidate genes may well result in missed SNPs that would be of great interest to pharmacologists and toxicologists, and scientists generally, in the long run.

Any two sets of chromosomes taken at random will differ roughly at one site per 500–1000 bp. There can be up to one potential variable site per 100 bp in an average segment of contiguous DNA if one examines DNA sequence variation among various populations (1, 5). This is one reason why the whole genome shotgun sequencing strategy discussed earlier is so powerful. Any normal individual will have two sets of chromosomes, all of which will be sequenced in their entirety. Thus, providing the full genome sequence from small numbers of normal donors, indeed from even one individual, will yield millions of SNPs for further study. However, correlations between genotypes and phenotypes (clinical conditions) will require association studies with appropriate levels of statistical power (see below).

The classic Mendelian model, in which a specific mutation in one gene produces a recognizable disease, may not apply to most common illnesses in our society. It is thought that many common illnesses have a polygenic origin, with several genes (to be more precise, gene variants) playing a comparatively small

role individually, but with a cumulative effect that leads to a detectable clinical condition or disease. There is considerable interest in using whole genome association studies, as a tool for identifying genes involved in these common disorders, to detect differences in the frequency of DNA sequence variations between unrelated affected individuals and a control group (43).

Genetic association strategies may be direct or indirect (8, 43). The direct-association strategy focuses on common variants in coding or regulatory regions. (The identification of the latter is still problematic and will be greatly facilitated by comparative genomics, which is discussed further below.) Frequencies of these variants in patients and control groups are analyzed statistically, with the goal of identifying alleles that serve as morbidity or mortality risk indicators. The indirect strategy relies on a statistical association between an illness or clinical condition. This occurs when an ''innocent'' variant is linked to the actual risk-producing variant on a stretch of DNA. The probability of a recombinatorial event occurring between any two points of DNA in a chromosome is related to the distance between the two points. If the distance is comparatively small, the two points (variants) are said to be in linkage disequilibrium. During meiosis, they are likely to share a finite journey through time, and in an individual, the presence of one marker would then predict the other. A ''neutral'' variant can then be used to detect or uncover a disease-causing variant because they are linked together in a shared haplotype (genomic segment), which is longer for new variants in a population and shorter for ancient variants. The length of the shared haplotype determines whether there is a strong or weak level of linkage disequilibrium. Thus, the indirect strategy employs a dense map of anonymous or random polymorphic markers to search the genome for statistical associations with disease. This can provide a unique tool for new gene discovery, and it does not require the investigator to intuit candidate genes or regulatory elements in advance. In some cases, one can identify a marker with a sufficiently strong predictive power for development as a diagnostic test in its own right. This approach can identify the location of genes even when they would not be considered good candidates for a disease association by first principles. Indeed, we expect the availability of the complete sequence for the human genome to yield any number of surprises in terms of assigning function and disease association to gene loci.

Although both such approaches have been used on a comparatively small scale to study genes in diseases, we will soon have unprecedented opportunities to apply such studies to the entire genome. It is widely viewed that SNPs provide the key to such approaches because of their high frequency, biallelic nature, low mutation rate, and suitability for industrial-scale automation. The precise number of SNPs necessary is still a matter for debate. Kruglyak (43) has recently proposed that roughly 500,000 evenly spaced SNPs will be required for optimal indirect whole genome association studies. This estimate is at the upper limit of various suggested numbers, and yet it is still easily within the range of SNPs that will emerge as part of the whole human genome shotgun sequencing strategy discussed earlier.

A SURVEY OF SNPS

Single base changes may directly cause gross alterations in gene function, but they may also be responsible for subtle correlations with disease. Such SNPs have been linked to a variety of cardiovascular, respiratory, allergic, neurologic, psychiatric, metabolic, bacterial, and neoplastic diseases. Some brief examples of DNA variations that pertain to cancer risk (44–53), infectious disease (38, 39, 54–62), asthma (63–71), and neuropsychiatric diseases (72–81) are summarized in Tables 2–5. In some of the tables areas of controversy are included, as a reminder of the need for more clinical research. A summary of certain alleles associated with drug metabolism is shown in Table 6 (82–100). In some cases, there is an overlap of definitions. Thus, an allele can affect cancer risk (Table 2) but it could also fit in drug metabolism (Table 6).

Sometimes, detection of a functional polymorphism leads to unexpected and dramatic biological insights. Thus, apolipoprotein E has been known for many years to play an important role in lipoprotein metabolism. The *APOE* gene provides an interesting example of genetic polymorphisms, with substantial variation in different groups. This gene affects cholesterol levels. It is surprising that one allelic variant of this gene (*APOE4*) is a significant risk factor in Alzheimer's disease (101, 102). *APOE4* appears to be directly involved in some way because the protein is seen on immunohistochemical staining in the amyloid plaques, amyloid deposits, and neurofibrillary tangles that characterize the brain lesions of Alzheimer's disease.

In this sense, SNPs offer new opportunities for prevention, diagnosis, and treatment strategies. In the not-too-distant future, physicians may be able to use advanced, miniaturized technologies in their clinics or offices to define a patient's SNP profile in order to customize a diagnosis and therapy to the specific patient's needs. We return to a discussion of pharmacogenetics further below, but several recent practical examples of these ideas are worth noting here. The thiopurine methyltransferase gene, which regulates the metabolic inactivation of azathioprine, is a good model for discussion. Substantial hematologic toxicity may accompany the use of this drug in the therapy of rheumatic diseases or transplantation rejection. Patients with a variant allele for thiopurine methyltransferase (TPMT*3A) are at substantially higher risk of hematologic toxicity compared with those with the wild-type allele (100). In another example, dealing with schizophrenia therapy, the best response to the important but potentially toxic drug clozapine was found in patients with two defined genetic polymorphisms in 5-HT_{2A} receptors (80). In another example, polymorphisms of a genetic factor expressed on platelets (*PlA2*) might determine whether use of aspirin can prevent myocardial infarctions in patients (103). Perhaps even the DNA from a buccal swab would be informative in selecting the right pharmaceutical for the right patient.

In functional terms, an SNP can be either good, bad, or neutral, depending on the selective pressures or circumstances faced by the individuals in a population.

TABLE 2 Sequence polymorphisms and cancer risk

Gene	Polymorphism(s)	Allele frequency	Clinical correlation	Reference
BRCA1	185delAG 5382insC	0.5% in Ashkenazi Jews 0.1% in Ashkenazi Jews Rare in other populations	Associated with increased risk of breast and ovarian cancer. Risk of developing cancer by age 70 is 56% for breast cancer and 16% for ovarian cancer. Male carriers have increased risk of developing prostate cancer, with 16% of carriers developing prostatic cancer by age 70.	44
BRCA2	6174delT	0.5% in Ashkenazi Jews; rare in other populations	Associated with increased risk of breast and ovarian cancer. Risk of developing cancer by age 70 is 56% for breast cancer and 16% for ovarian cancer. Male breast cancer appears to be more common in BRCA2 mutation carriers than in BRCA1 carriers.	44
APC	I1307K; Ile to Lys change at residue 1307.	3.5% in Ashkenazi Jews; rare in other populations	Carriage of the I1307K allele is associated with somatic instability in genomic APC DNA and can result in loss of APC expression and the development of cancer. Risk of colorectal and breast cancer are approximately 1.5-fold higher in I1307K carriers than in noncarriers.	45
NAT2 (N-acetyl-transferase 2)	Multiple sequence variant combinations in NAT2 form 11 haplotypes. Subjects carrying the NAT2*4 haplotype (in either the homozygous or heterozygous states) are rapid acetylators; subjects with other haplotypes are slow acetylators.	In a set of 556 German volunteers, 23.4% of alleles were of the NAT2*4 haplotype.	Individuals who were homozygous for the rapid acetylation types NAT2*4/NAT2*4 were overrepresented in lung cancer patients compared with control patients; odds ratio, 2.36 (95% confidence interval 1.05–5.32). See also discussion of NAT1 and NAT2 in Table 6.	46

MTHFR (methylene-tetraydro-folate reductase	677C→T, results in a change from alanine to valine	677C = 0.58, 677T = 0.42 in Caucasian populations	The 677T allele of MTHFR is heat labile and results in reduced enzyme activity leading to lower levels of 5-methyltetrahydrofolate. In men with adequate folate levels, homozygosity for the 677T allele reduced the risk of colorectal cancer, but in men with folate deficiency or those drinking more than 10 g of alcohol/day, 677T did not reduce cancer risk.	47, 48
AR (androgen receptor)	Highly polymorphic CAG repeat present in the first exon of the X-linked androreceptor.	Continuous distribution of CAG repeat length ranging from 12–32 repeats. Modal length of 21 CAG repeats present in 30–40% of women.	Length of the AR CAG repeat is inversely correlated with the transcriptional activity of the AR. AR CAG re-repeat length was found to modify breast cancer risk in women with BRCA1 mutations. Women carrying AR-CAG alleles of >28, >29, or >30 repeats were diagnosed with breast cancer 0.8, 1.8, or 6.3 years earlier than women with BRCA1 mutations with shorter AR-CAG alleles.	49
p53	R72P; Arg to Pro change at residue 72	Allele frequencies have been reported to vary in different study populations. In the largest single normal group studied, 626 control patients from Sweden: R72 = 0.69, P72 = 0.31	The form of p53 carrying an arginine residue at codon 72 is significantly more susceptible to degradation by the HPV E6 protein than is the variant with proline at this position. In one study (50), 72R homozygotes were seven times more likely to develop HPV-associated cancers than were 72R/72P heterozygotes (numbers of 72P homozygotes being too small to analyze separately). This result proved to be controversial, however, and could not be replicated by three subsequent studies examining the effect of the p53 R72P polymorphism on HPV-associated cervical cancer risk.	50–53

TABLE 3 Sequence polymorphisms and infectious disease

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
Chemokine receptor 2 (CCR2)	Valine at position 64 is replaced by isoleucine (V641).	V64, 90; I64, 10%; $n = 248$	In a study of 3003 HIV-positive patients, individuals carrying the I64 allele progressed to AIDS 2 to 4 years later.	54
Chemokine receptor 5 (CCR5)	32-bp deletion (Δ32)	+ 32 bp, 90.8%, Δ32, 9.2%; $n = 1408$ Caucasians The Δ32 allele is absent in Western and Central African and Japanese populations.	Carriers of the 32-bp deletion are protected from transmission of HIV.	55, 56
β-Globin	Glycine at position 6 is replaced by valine (G6V).	G6, 92%; V6, 8%; American Blacks G6, 60–80%; V6, 20–40%; African Blacks	The prevalence of malaria infection (42%) was significantly lower in individuals with the sickle-cell trait compared with their normal-hemoglobin counterparts (68%).	57, 58
Human leukocyte antigen (HLA)	HLA-B*35 haplotype HLA-C*w04 haplotype		The HLA-B*35 haplotype is associated with rapid development of AIDS in HIV-infected Caucasians. The HLS-C*w04 haplotype is associated with rapid development of AIDS in HIV-infected Caucasians.	59
Intercellular adhesion molecule 1 (ICAM-1)	Lysine at position 29 is replaced by methionine (K29M).	L29, 67%; M29, 33%; $n = 287$ Kenyans (controls) L29, 56%; M29, 44%; $n = 157$ Kenyans (cerebral malaria) L29, 76%; M29, 24%; $n = 422$ Gambians (controls)	In a case-control study of 547 subjects in Kenya, a single ICAM-1 mutation was present at high frequency. Genotypes at this locus from samples in this case-control study indicated an association of the polymorphism with the severity of clinical malaria such that individuals homozygous for the mutation (M29/M29) have increased susceptibility to cerebral malaria with a relative risk of two.	60, 61

		L29, 76%; M29, 24%; $n = 367$ Gambians (cerebral malaria)	Over 1200 children in The Gambia were typed for polymorphisms of the ICAM-1, gene. None of the polymorphisms typed was significantly associated with severe disease. These data differed significantly from the results of a previous study (Chi 2 = 8.81; $P = 0.003$) in which the ICAM-1 gene polymorphism was shown to be significantly associated with cerebral malaria. This suggests that there may be heterogeneity in genetic susceptibility to this condition between these two African populations.	
Natural resistance-associated macrophage protein 1 (NRAMP1)	There is a $(CA)_n$ repeat variation in the 5′ untranslated region.	CA repeat (201 bp), 84%	Four NRAMP1 polymorphisms were each significantly associated with tuberculosis. Subjects who were heterozygous for two NRAMP1 polymorphisms in intron 4 and the 3′ untranslated region of the gene were particularly overrepresented among those with tuberculosis, as compared with those with the most common NRAMP1 genotype (odds ratio, 4.07; 95% confidence interval, 1.86–9.12; chi-square = 14.58; $P = 0.001$).	62
	There is a G to C transversion in intron 4 (469 + 14 G/C).	469 + 14G, 92%; 469 + 14C, 8%		
	Aspartic acid at position 543 is replaced by asparagine (D543N).	D543, 93% N543, 7%;		
	There is a 4-bp deletion in the 3′ untranslated region (1749 + 55 ΔTGTG).	1749 + 55 (TGTG), 81%; 1749 + 55 (ΔTGTG), 19%; $n = 827$ Gambians		
	The D543 allele is always associated with ΔTGTG in the 3′UTR.			
Tumor necrosis factor (TNF)	There are 3 polymorphisms located in the promoter region of this gene: −238 G/A, −308 G/A, −376 G/A.	−376 G, 98.5%; −376 A, 1.5%; $n = 371$ Gambians (controls)	The -376 polymorphism is located in the OCT-1 binding site of the TNF promoter. Only the −376 A allele binds OCT-1 and it is associated with a fourfold increased susceptibility to cerebral malaria.	39
	The −376 A allele always occurs along with the −238 A allele.	−376 G, 96.9%; −376 A, 3.1; $n = 384$ Gambians (cerebral malaria)		

TABLE 4 Sequence polymorphisms and asthma

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
$Beta_2$-adrenoceptor (B2AR)	Glycine at position 16 is replaced by arginine (G16R). Glutamine at position 27 is replaced by glutamic acid (Q27E). Arginine at position 16 was associated with glutamine at position 27 in 97.8% of the haplotypes determined.	G16, 62%; R16, 38%; $n = 269$ Americans Q27, 64%; E27, 36%; $n = 269$ Americans	To assess if different genotypes of these two polymorphisms would show differential responses to inhaled B2AR agonists, 269 children who were participants in a longitudinal study of asthma were genotyped. When compared with homozygotes for G16, homozygotes for R16 were 5.3 times (95% confidence interval, 1.6–17.7) and heterozygotes (G16/R16) were 2.3 times (1.3–4.2) more likely to respond to albuterol, respectively. No association was found between the Q27E polymorphism and response to albuterol.	63
Cystic fibrosis transmembrane regulator (CFTR)	Phenylalanine at position 508 is deleted (ΔF508).	F508, 97.3%; ΔF508, 2.7%; $n = 9141$ Danes	The CFTR gene encodes a chloride channel found in epithelial cells. Individuals with mutations in both copies of this gene are affected with the chronic lung disease cystic fibrosis. Of the 250 ΔF408 mutation carriers identified in this study, 9% reported having asthma compared with 6% of the noncarriers. The odds ratio for asthma for carriers of the ΔF508 mutation was 2.0 (1.2–3.5, $P = 0.02$).	64

Fc(epsilon)RI, beta subunit (high-affinity receptor for immunoglobulin E)	Isoleucine at position 181 is replaced by leucine (I181L). Valine at position 183 is replaced by leucine (V183L). Glutamic acid at position 237 is replaced by glycine (E237G).	I181, 97%; L181, 3% (Caucasians) I181, 84%; L181, 16% (African Blacks) I181, 38%; L181, 72%; $n = 221$ Kuwaitis V183, 38%; L183, 72%; $n = 221$ Kuwaitis E237, 80%; G237, 20% (African Blacks) E237, 94.7%; G237, 5.3%; $n = 1004$ Australians	A sample of black and white asthmatic and control subjects in South Africa was studied to determine whether these variants contribute to the enhanced immunoglobulin E responses in these groups. There was a significant difference in the frequency of L181 between white asthmatics (28%) and white control subjects (3%) ($P = 0.00001$), but no difference in the frequency of I181L was observed between black asthmatics (22%) and black control subjects (16%). I181L might predispose to atopy in the white population but not in the black population. The Fc(epsilon)RIbeta polymorphism (181/183) was investigated in Kuwaiti asthmatic patients and controls. The variant sequence (L181/L183) was detected in 72% (320/442) of chromosomes analyzed. Homozygous LL genotype was detected in 48% (46/96) asthmatic subjects compared with 31% (39/125) in nonasthmatics. G237 positive subjects had a significantly elevated skin test response to grass ($P = 0.0004$) and house dust mites ($P = 0.04$), RAST to grass ($P = 0.002$), and bronchial reactivity to methacholine ($P = 0.0009$). The relative risk of individuals with G237 having asthma compared with subjects without the variant was 2.3 (95% confidence interval, 1.26–4.19; $P = 0.005$).	65–67

(*continued*)

TABLE 4 *Continued*

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
Human leukocyte antigen (HLA)	HLA-B8; HLAS-A10 haplotype; HLA-DQ2 haplotype		In a study of 76 Greek asthmatic patients (35 children/41 adults), increased frequency of HLA-B8 was found in the adults and an increased frequency of HLA-A10 was found in the children. The HLA-DQ2 allele is found more frequently in asthmatic children than in control subjects (60% vs 34%, $P = 0.013$) with a relative risk of 2.8.	68, 69
5-Lipoxygenase (ALOX5)	$(GGGCGG)_{3\text{-}6}$ repeat variation in the promoter region	3 repeats, 3.8%; 4 repeats, 172.%; 5 repeats, 77.2%; 6 repeats, 1.8%; $n = 221$ Americans	Individuals with asthma carrying the 5-repeat allele respond to treatment targeted at this enzyme (18–23% improvement in FEV). Individuals without the 5-repeat allele showed no response.	70
Platelet-activating factor; acetylhydrolase (PAFA)	Valine at position 279 is replaced by phenylalanine (V279F).	V279, 82%; F279, 18%; $n =$ 263 Japanese (controls); V279, 77%; F279, 23%; $n =$ 266 Japanese (asthmatics)	A missense mutation (V279F) in the PAF acetylhydrolase gene results in the complete loss of activity. The prevalence of PAF acetylhydrolase deficiency is higher in Japanese asthmatics than in healthy subjects and the severity of this syndrome is highest in homozygous-deficient subjects. PAF acetylhydrolase gene is a modulating locus for the severity of asthma.	71

TABLE 5 Sequence polymorphisms and drug response—central nervous system

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
D_2 dopamine receptor	−241 A/G in 3′ UTR		The polymorphism in the promoter region (−141 C) does not affect clozapine response. The serine 310 variant has been reported to be associated with the adverse drug reaction of neuroleptic malignant syndrome. In an in vitro study, the A96 variant showed a reduction of ~50% in the binding of dopamine, chlorpromazine, and clozapine. The binding of other neuroleptic drugs (haloperidol, thioridazine, thiothixene, and risperidone) was not affected. Sequence variation at positions 310 and 311 did not affect drug binding.	72–74
	−141 C ins/del in the promoter region			
	Valine at position 96 is placed by alanine (V96A)			
	Proline at position 310 is replaced by serine (P310S)	0.004 Caucasians		
	Serine at position 311 is replaced by cysteine (S311C)	0.16 Southwestern American Indians; 0.03 Caucasian; 0.23 Japanese		
D_4 dopamine receptor	A null mutation occurs in ~2% of the general population		Binding site of atypical neuroleptics clozapine and olanzapine. The 7-repeat allele of the exon 3 VNTR is associated with dependence on opiate drugs. The effect of other polymorphisms on drug efficacy is not yet known.	75–77
	Exon 1 contains a 12-bp duplication and a 13-bp deletion			
	Exon 3 contains a 48-bp repeat (2–10 copies) and a G194V polymorphism	0.64 4-repeat, 0.20 7-repeat, 0.08 2-repeat alleles		

(*continued*)

TABLE 5 *Continued*

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
D_5 dopamine receptor	Leucine at position 88 is replaced by phenylalanine (L88F).		In an in vitro study, the N351D polymorphism resulted in a ten-fold decrease in dopamine binding affinity, and a three-fold decrease in R(=)-SKF-38393 binding. The L88F polymorphism showed a slight decrease in the binding of SCH-23390 and risperidone, and a small increase in dopamine binding.	78
	Alanine at position 269 is replaced by valine (A269V).	0.008 Caucasians		
	Proline at position 330 is replaced by glutamine (P330Q).	0.1 Asians		
	Asparagine at position 351 is replaced by aspartic acid (N351D).	0.008 Caucasians		
	Serine at position 453 is replaced by cysteine (S453C).	0.008 Caucasians		
Dopamine transporter (DAT1)	VNTR in 3′ UTR		480-bp allele has been associated with attention-deficit hyperactivity disorder.	79
5-HT_{2A} serotonin receptor	Thymine at position 102 is replaced by cytosine (102 T/C).	C allele 0.5 in Caucasians	Receptor target of antipsychotics risperidone, ketanserin, clozapine, and olanzapine. Retrospective analysis of studies showed an association between the 102 T/C and H452Y polymorphisms and clozapine response.	80
	Threonine 25 replaced by asparagine (Thr452Asn)	0.02 Caucasians		
	Alanine 447 replaced by valine (Ala447Val)	0.01 Caucasians		
	Histidine at position 452 is replaced by tyrosine (H452Y).	0.09 Caucasians		
Serotonin transporter (5-HTT)	44-bp insertion/deletion in promoter region	S allele 0.43 in Caucasians	s allele: reduced transcriptional efficiency, associated with anxiety-related traits.	81

TABLE 6 Sequence polymorphisms and drug metabolism

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
Cytochrome P-450 2D6 (CYP2D6)	There are 48 positions in the CYP2D6 gene where sequence variations have been reported, resulting in 53 different haplotypes. The allele frequencies for the most common polymorphisms are listed in the next column.	V11, 93.3%; M11, 6.7% P34, 81.6%; S34, 18.4% L91, 84.2%; M91, 15.8% H94, 84.2%; R94, 15.8% 1085 C 84.1%; 1085 G 15.9% 1127 C 97.5%; 1127 T, 2.5% 1749 G 47.4%; 1749 C 52.6% 1935 G 82.9%; 1935 A 17.1% 2637 A 98.3%; 2637 ΔA 1.7% K281, 97.3%; ΔK281, 2.7% R296, 65.7%; C296, 34.3% 3916 G 98.8%; 3916 A, 1.2% S486, 47.1%; T486, 52.9% Null allele 6.9%; $n = 1344$ Europeans	More than 25 commonly prescribed drugs are metabolized by CYP2D6, including codeine, debrisoquine, indoramin, phenformin, and a number of antiarrythmics, antidepressants, beta-blockers, and neuroleptics. Sequence changes in CYP2D6 can either increase or decrease the rate at which the enzyme functions. Extensive metabolizers need lower doses to achieve therapeutic response, and may suffer adverse effects if a normal dose is given. Poor metabolizers require greatly elevated doses to achieve response and may not respond to some drugs at all.	82–84
Cytochrome P-450 2C9 (CYP2C9)	The wild-type sequence is designated CYP2C9*1. Arginine at position 144 is replaced by cysteine (R144C). This variant allele is designated CYP2C9*2. Isoleucine at position 359 is replaced by leucine (I359L). This variant allele is designated CYP2C9*3.	CYP2C9*1, 79%; CYP2C9*2, 12.5% CYP2C9*3, 8.5%; $n = 100$ Caucasians	Cytochrome P-450 CYP2C9 is responsible for the metabolism of S-warfarin. The CYP2C9*2 and CYP2C9*3 allelic variants are associated with impaired hydroxylation of S-warfarin. These patients have difficulty at induction of warfarin therapy and are potentially at a higher risk of bleeding complications. The V_{max} values for phenytoin in patients who are heterozygous for the I359/L359 allele are 40% lower than those in patients with the wild-type CYP2C9 allele (I359/I359 homozygotes).	85–87

(*continued*)

TABLE 6 *Continued*

Gene	Polymorphism(s)	Allele frequencies	Clinical correlation	Reference
Cytochrome P-450 2C19 (CYP2C19)	M1 allele; 1-bp mutation in exon 5 splice site M2 allele: G636A transition in exon 4	M1 allele: 30.7%, $n = 39$ Japanese; 16.7%, $n = 45$ Caucasians M2 allele: 8.9%, $n = 39$ Japanese; 0%, $n = 45$ Caucasians	CYP2C19 catalyzes the 4-hydroxylation of S-phenytoin, hexobarbitone, diazepam, omeprazole, proguanil, and R-warfarin. The M1 1-bp mutation lacks the heme binding domain and is catalytically inactive. It is associated with the poor metabolizer phenotype. The A636 transition in exon 4 creates a stop codon and is also associated with the poor metabolizer phenotype.	86, 88, 89
Glutathine S-transferase Mu (GSTMI)	The presence of a null allele results in decreased total GST activity.	Normal, 46.5%; null, 53.5%; $n = 213$ whites normal, 72.4%; null, 27.6%; $n = 203$ blacks	GSTM1 is a detoxification enzyme with high specificity for epoxides. Individuals who are homozygous for the null allele demonstrated a slightly increased risk for prostate cancer, and smokers who carry the null allele are at increased risk for bladder cancer.	90, 91
N-acetyl transferase 1 (NAT1)	There are 29 positions in the NAT1 gene where sequence variations have been reported, resulting in 24 different haplotypes.	R64, 95%; W64, 5%; $n = 85$ Caucasians R187, 96%; Q187, 4%; $n = 85$ Caucasians	NAT1 can further metabolize hydroxylamine metabolites to N-acetoxy derivatives. Normal NAT-1/fast NAT-2 genotype may be protective for susceptibility to smoking-induced bladder cancer.	92–94
N-acetyl transferase 2 (NAT2)	There are 11 position in the NAT2 gene where sequence variations have been reported, resulting in 26 different haplotypes.	I114T: 55:45, $n = 968$ Caucasians; 76:33, $N = 61$ Indians; 70:30, $N = 214$ African Americans; 72:28, $n = 148$ Hispanics; 99:1, a $n = 224$ Japanese; 95:5, $n = 254$ Chinese R197Q: 55:45, $n = 968$ Caucasians; 67:33, $n = 61$ Indians; 70:30, $n = 214$ African Americans; 72:28, $n = 148$ Hispanics; 99:1, $n = 224$ Japanese; 95:5, $n = 254$ Chinese	NAT2 polymorphisms are associated with higher incidence or severity of adverse drug reactions to isoniazid, hydralazine, procainamide, and sulfamethazole. Slow acetylators {NAT2*5B (I114T), and NAT2*6A [C282T(Y94Y), R197Q]} who smoke may have an eightfold higher risk of bladder and lung cancer due to their inability to detoxify aromatic amines in tobacco smoke. Rapid acetylators (NAT2*4) are at increased risk of colon cancer from acetylation of heterocyclic amines found in cooked meats.	93–95

Paraoxonase (PON1)	Methionine at position 55 is replaced by leucine (M55L). Glutamine at position 192 is replaced by arginine (Q192R).	Q192, 84.5%; R192, 15.5%; n = 166 Indians Q192, 63.2%; R192, 36.8%; n = 105 Turks Q192, 73%; R192, 27%; n = 248 Saudis	High serum paraoxonase levels protect against the neurotoxic effects of organophosphate insecticides as well as the nerve agents soman and sarin. The R192 allele specifies high enzymatic activity, whereas the Q192 variant specifies low activity. The M55L polymorphism affects the level of mRNA, with M55L heterozygotes showing an excess of the L55 allele.	96, 97
Thiopurine S-methyl transferase (TPMT)	TMPT has several mutant alleles that exhibit low enzyme activity compared with the wild-type enzyme (TPMT*1). Alanine at position 80 is replaced by proline (A80P). This variant allele is designated TPMT*2. Alanine at position 154 is replaced by threonine (A154T). This variant allele is designated TPMT*3B. Tyrosine at position 240 is replaced by cysteine (Y240C). This variant allele is designated TPMT*3C. Variant TPMT*3A contains both the T154 and C240 amino acid changes.	Haplotype frequencies: TPMT*1, 94.5%; TPMT*2, 0.18%; TPMT*3A, 3%; TPMT*3B, 0.35%; TPMT*3C, 0.71%; TPMT*3D, 0.18%; TPMT*4, 0%; TPMT*5, 0.18%; all other, 0.88%; n = 283 (mixed ethnic groups)	Prevention of transplant rejection and therapy of rheumatoid arthritis by the immunosuppressive drug azathioprine is limited by hemotologic toxicity (leucopenia or agranulosis). This toxicity is a particular problem in patients with low TPMTase activity (homozygotes for variant alleles; ~1% of the population). Other drugs affected by variation in TPMT include mercaptopurine and thioguanine.	98–100

Evolution affects the phenotypes encoded by genotypes, but it cannot operate through the genotype directly (11). Phenotypes that confer a selective advantage at one point in a population's history may at a different point do the opposite. Thus, there is an increased risk for deep venous thrombosis in carriers of the prothrombin $G \rightarrow A^{20210}$ gene variant (polymorphism). The *A* allele is deleterious because it is associated with increased prothrombin levels and an approximately threefold increased risk for deep venous thrombosis compared with individuals homozygous for the *G* allele (104). The median age at the time of the first thrombotic episode is 38 years. Deep venous thrombosis is a serious and potentially life-threatening condition for human beings in western society today. But there was likely a time in human history in which the life expectancy of an individual was well under 38 years. The capacity to undergo blood coagulation quickly and decisively might well have been a selective advantage if wounds from predatory animals or human combatants were a common event in everyday life. Indeed, even as the twentieth century draws to a close, current events do not necessarily support the lack of a selective advantage for the *A* allele.

PHARMACOGENETICS

Research pharmacologists and toxicologists have begun applying the lessons of modern genetics and molecular biology at perhaps a much faster pace than many other medical sciences. This is clear in the rapid practical applications of pharmacogenetics and supporting technology (83, 84, 105). Yet, as with any genetics testing, there are a number of regulatory and policy issues that require further dialogue (106, 107). Pharmacogenetics, a term originally coined in the 1950s, may now be viewed as the study of correlations between an individual's genotype and that same individual's ability to metabolize an administered drug or compound. Genotypic variation, often in the form of SNPs, exists for many of the enzymes that metabolize important drugs. Extensive metabolism of a drug is a general characteristic of the normal population. Poor metabolism, which typically is associated with excess accumulation of specific drugs or active metabolic products, is generally an autosomal recessive trait requiring a functional change, such as a frameshift or splicing defect in both copies of the relevant gene. Ultraextensive metabolism, which may have the effect of diminishing a drug's apparent efficacy in a given individual, is generally an autosomal dominant state derived from a gene duplication or amplification. Some representative examples of genetic polymorphisms that affect drug metabolism are shown in Table 6. For some fields, such as cancer chemotherapy, several common drugs show wide polymorphism-related metabolic varations, with 30-fold or greater interindividual variability reported (108).

Drug metabolism is often divided into two components: phase I (oxidative) and phase II (conjugative). The cardinal phase I enzymes belong to the cytochrome P-450 (CYP) superfamily of inducible mixed-function monooxygenases

located within hepatic endoplasmic reticulum. Thirty or more forms of P-450 enzymes have been identified, each with distinct enzymatic activity. It is difficult to overstate the importance of these enzymes in clinical practice. More than 25 prescription drugs are metabolized by one member of the family, the CYP2D6, alone. It has been estimated that genetic polymorphisms of the *CYP2D6* gene alter clinical care and outcome of nearly one fifth of patients in some ethnic groups. In the case of certain drugs or carcinogens, the P-450 enzymes play a role in generating active moieties from otherwise inactive or poorly active starting compounds (pro-drugs or pro-carcinogens), which may add a layer of complexity in any global overview of these enzymes. Phase II enzymes include the glutathione transferases, N-acetyltransferases, UDP-glucuronyl-transferases, and the sulfotransferases. Genetic polymorphisms in these enzymes can also have important pharmacologic and clinical sequelae. They may also participate in relative cancer risk and causation, depending on their effectiveness in eliminating certain carcinogens (109) (see also Tables 2 and 6). However, with our current state of knowledge, there may be complex or counter-intuitive relationships between cancer and certain polymorphisms, such as those seen in the N-acetyltransferases.

In a sense, every clinical pharmacologist already relies on the tools of medical genetics and genotyping. The common technologies for determining alleles of pharmacogenetic interest are readily adaptable to most reference labs. There are two commonly used techniques. The first is amplification of a specific genetic region by polymerase chain reaction, followed by an analysis of restriction fragment length polymorphism (which reflects alterations in nucleotide sequence). The second involves allele-specific polymerase chain reaction, which depends on oligonucleotides capable of hybridization with common or variant alleles. Only a successful (i.e. precise) hybridization to the known target sequence yields an amplification product, which can in turn be detected on agarose gels. However, it is expected that improvements in microfluidic and chip-based technologies will make it even easier and more convenient, and thus more important, to assess genotypes of important metabolic enzymes. These technologies may also eventually make it possible to address the metabolic pathways for biologic response modifiers, for which very little is known.

The availability of the complete sequence for the human genome, including new knowledge regarding DNA variation, will alter the scope and definition of pharmacogenetics. New genomic knowledge will make it possible to examine families of drug efflux pump genes (e.g. P-glycoprotein) in the gut epithelium and central nervous system for their effects on oral bioavailability and central nervous protective effects, particularly with respect to xenobiotic agents (110). This knowledge may make it possible to clearly predict important pharmacokinetic parameters of certain drugs well in advance of empirical testing. In addition to the possibility of identifying new genes (and new alleles) from known metabolic gene families, pharmacologists will gain unprecedented ways of analyzing polymorphisms in the target receptors of important drugs, as well as in genes involved in the metabolic pathways affected by new agents. As an example, about

6% of patients with asthma do not carry a wild-type allele at the 5-lipoxygenase (*ALOX5*) core promoter locus (see Table 4). In such patients, the *ALOX5* pathway agents (i.e. leukotrienes) do not make a major contribution to their small airway disease, and these asthmatics do not improve with a drug whose mechanism is to inhibit this pathway (70). Also, a common polymorphism of the leukotriene C_4 synthase promoter appears to be a risk factor for aspirin-induced asthma (111). This clearly illustrates the principle that variants in the promoter region of a therapeutic target gene can predict clinical response. This concept is a significant expansion of pharmacogenetics beyond polymorphisms in drug-metabolizing enzymes.

Thus, on several fronts, pharmacologists and toxicologists will be able to address issues of efficacy and safety not reflected in the traditional pharmacokinetic or pharmacodynamic profiles of drugs. Even when patients have identical pharmacokinetic profiles for a given drug, we currently have no way of knowing that there will be comparable interpatient efficacy or safety because the science of pharmacogenetics can now only look at one part of the picture, and even that picture generally excludes important classes of therapeutics, such as biologic-response modifiers and monoclonal antibodies. Some examples of polymorphisms in a target receptor, with important pharmacologic consequences, are shown in Tables 4 and 5. In the future, the growing knowledge, based on the foundation of the complete human genome, will make it possible for pharmaceutical developers to select therapeutic agents according to the individual allele profiles of the intended patient. Said another way, patients will someday get only the drugs they need, and no other. But what is certain now is that pharmacology and toxicology will become dependent on the emerging bioinformatics and computational sciences linked to genomics. Data acquisition and management across multiple disciplines will require new tools and skill sets, necessitating changes in the curricula of pharmacology teaching centers toward a much more computational orientation. It will also be necessary for new cross-relational databases to emerge to serve the needs of pharmacologists and toxicologists in the new genomics era.

COMPARATIVE GENOMICS

The availability of whole genome sequence information in the human is important in its own right, but the full power of this knowledge requires the additional availability of whole genome sequence information from model organisms, especially *Drosophila melanogaster* (common fruit flies) and mice. *Drosophila* has been at the forefront of genetics research for nearly 80 years. This species is an important model for combining genetics, electrophysiology, and molecular biology. A brief listing of some medically important fruit fly genes is shown in Table 7. We expect that our group, in collaboration with the Berkeley Drosophila Genome Project, will publish the complete sequence of *Drosophila* in the near future. Many gene families, ranging literally from A to Z (aldolases to zinc finger

TABLE 7 Model organisms, gene names, and human disease

Fruit flies[a]	Mice	Humans
Congenital heart defects Tinman	Csx/NKx2.5	Csx/NKx2.5
Huntington's polyglutamine repeats cause neural degeneration	Polyglutamine repeats	Polyglutamine repeats
Alzheimer's APPL	APP	APP
Amyotrophic lateral sclerosis SOD	SOD1	SOD1
Cancer Cdx(2) caudal	cdx2	cdx2
Diabetes insulin receptor	Insulin receptor (knockout mice)	Insulin receptor
Retinitis pigmentosa rhodopsin	Rhodopsin	Rhodopsin
NF MERLIN	NF2	NF2
Ataxia telangiectasia mei-41	ATM	ATM

[a]Human disease. APP, amyloid precursor protein; SOD, superoxide dismutase, NF, neurofibromatosis.

transcription factors), suggest that a single invertebrate gene corresponds to a handful of equally related vertebrate genes on different chromosomes. Thus, gene duplication is the engine that drove vertebrate evolution (112).

In the case of vertebrate organisms, not only does homology to human genes exist, there is also something more meaningful called synteny. This means there are related genes arrayed on chromosomes with an evolutionary history common to human counterparts, in a comparable order in terms of exons and gene regulatory elements. Thus, a mouse chromosomal region with such a common evolutionary history and genetic arrangement is said to be syntenic, and the relevant mouse and human genes are said to be orthologs. Synteny and the capacity to overlay complete human and mammalian genome sequences (especially mouse) will affect gene discovery and our understanding of gene structure and function in ways without precedent. We expect the completion of other genomes with importance to pharmacology and toxicology, e.g. rat and canine genomes, not too far behind those of humans and mice. These advances will allow an integration of information from transgenic animals and gene knockout models in far-reaching ways and will stimulate novel strategies for currently intractable therapeutic problems.

The implications of these advances can perhaps be illustrated with a few examples. The homeobox genes are an interesting case. An excellent website by Gaunt can be consulted (http://www.bi.bbsrc.ac.uk/world/sci4Alll/gaunt/dud/gaunt.html). We can examine homeobox (*Hox*) genes (113), a family of genes that are conserved in evolution in detail. These genes direct the development of the body plan and body parts of many morphologically distinct species. The homeobox is a highly conserved 180-bp nucleotide sequence shared by this family of transcription factors. The homeobox in turn encodes the DNA binding region, called the homeodomain, a DNA binding motif.

Homeobox genes are found in clusters, as shown at the top of Figure 1 (see color insert). The 5′ → 3′ order is reflected in the spatial correspondence of the genes along the posterior → anterior axis in *D. melanogaster,* and this is likewise conserved in other species. There is a rigid correspondence between the order of *Hox* genes within their clusters and that of their expression domains along the body of the embryo. Homeobox genes have a high degree of sequence similarity (denoted by color coding) in Figure 1 and can be found thoughout the animal kingdom at the same positions in the cluster. Shown below *Drosophila* in Figure 1 are the corresponding clusters of orthologs in mice, demonstrating a general preservation of these genes, along with the expected pattern of gene duplication in vertebrates in different chromosomes. Verterbrates have four Hox clusters, called *Hoxa*, *Hoxb*, *Hoxc*, and *Hoxd.* A more complete understanding of these genes will offer the pharmacologist many new strategies for preventing and treating birth defects and possibly organ or even limb repair.

Cross-species homologues can show similar function, even in organisms separated by hundreds of millions of years of evolution. Thus, dysfunction of the homeotic gene *Pax-6* can cause an eyeless phenotype in *Drosophila* and small eye syndrome in mice, whereas in children, mutations in *Pax-6* results in a complete loss of the iris and also in disturbances of the lens, cornea, and retina, which can contribute to blindness (114, 115). One must use caution in extrapolating these observations too broadly, but they serve as a reminder of the remarkable power of comparative genomics from fruit fly to human.

Knowledge from comparative genomics may make it possible to rethink therapeutic strategies for currently untreatable disorders, including those that arise from cytogenetic abnormalities. For example, is it possible that if an appropriate intervention were made early enough, Down's syndrome (trisomy, critical region chromosome 21q22.2–q22.3) might not inevitably lead to mental retardation? Asked another way, do the models of phenylketonuria and congenital hypothyroidism apply? Is there a way of testing whether an extra dose of a gene causes mental retardation, and could a pharmaceutical agent somehow neutralize the effects of such an extra dose if given early enough? One could note, with some justification, the futility of such hypotheticals in the absence of knowledge of the genes responsible for the disorder, and this underscores the importance of completing the human genome sequence effort quickly.

By utilizing available knowledge of synteny and comparative genomics, a novel family of protein kinases, called Dyrk, was found to represent interesting candidates for a role in this syndrome (116). Human chromosome 21 and the syntenic region of mouse chromosome 16 are shown in Figure 2 (see color insert) (117, 118). The gene *Dyrk1A* is located in the Down's syndrome critical region of chromosome 21. *Dyrk1A* reveals homologies with *minibrain,* a gene in *Drosophila* whose mutations yield reduced neuronal number and defective learning behavior. Mouse models of Down's syndrome, which involve a partial trisomy 16, have been created (119, 120). And perhaps most interesting, mice transgenic for a 180-kb DNA segment derived from the human Down's syndrome critical

region had defects in learning and memory (121, 122). Thus, it may soon be possible to identify both the gene(s) responsible for the most significant feature of the syndrome and in vivo systems for designing and testing interventions. These types of research opportunities will be multiplied thousands of times with the completion of the reference human genome program.

THE CHIMPANZEE GENOME

Approximately 5 million years ago, humans and chimpanzees shared a common ancestor. At the nucleotide sequence level, humans and chimpanzees differ by approximately 1.5% (50 Mb), on average. It has been argued that portions of human and chimpanzee genes are sometimes so similar that differences may fall within the range of normal DNA sequence variation (123). The overlapping synteny between humans and chimpanzees means that the completion of a reference human genome will greatly simplify the tasks of sequencing and assembling a reference genome for chimps.

Chimp DNA variation casts light on human DNA variation. Of nearly all human SNPs studied, one or another of the nucleotide allele set is shared with chimpanzees, thus making it possible to differentiate between ancestral and more recent alleles. Hacia et al (124) found that at the vast majority of SNP sites scored, all three nonhuman primates (chimps, pygmy chimps, and gorillas) exhibited one or the other of the human nucleotides, and roughly 75% of the time, the more common allele in humans was the ancestral form.

On the basis of considerable DNA sequence information (especially from the noncoding sequences of the genomic region called the beta-globin gene cluster), Goodman (125) has proposed creating a new phylogenetic classification of primates. In this proposed classification, humans would share their genus with chimpanzees and bonobos (pygmy chimpanzees). Thus, the subtribe *Homina* would contain *Homo* (*Homo*), humans, and *Homo* (*Pan*), chimpanzees and bonobos.

The science of comparative primate genomics can yield a number of interesting insights into the molecular events that contributed to human evolution (125–129). For example, mutations in *cis*-regulatory control elements changed γ-globin gene expression from strictly embryonic to fetal in profile, and there is evidence that natural selection favored such a distinct fetal hemoglobin during the evolution of the anthropoid primates. There were also amino acid–changing substitutions, including those in codons specifying 2,3-diphosphoglycerate–binding sites. A noteworthy consequence at the protein level was a loss of 2,3-diphosphoglycerate–binding capacity, yielding a fetal hemoglobin molecule that binds oxygen with increased affinity. Such a change facilitates the transfer of oxygen from mother to fetus. These changes, taken together, made a prolonged gestation and extensive prenatal brain development possible in anthropoid primates.

The changes in both regulatory elements and coding exons that permitted a prolonged fetal gestation provide but a glimpse of what can be learned when

complete genomic information becomes available. These changes are particularly instructive because they are not in any immediate or obvious way associated with brain structure or related anatomical development, and yet these differences in hemoglobin structure and kinetics of expression enabled the evolution of intelligence in primates.

Two questions emerge at this point. What are the genes that define human beings as a unique primate species and how can we apply this knowledge to medical science? McConkey & Goodman (130) have suggested that the first priority would be to identify genes that are situated in close proximity to the archaic chromosomal rearrangements occurring in human evolution after the divergence of humans and chimps from their last common ancestor. The fusion of two ape chromosomes to form human chromosome 2 and a small pericentric inversion in human chromosomes 1 and 18 might be excellent clues about where to search for candidate genes and regulatory elements that uniquely facilitated human evolution (131, 132). In any event, there will be exceptional opportunities to study the genes and pathways involved in affect and emotion; cognition and memory; language, speech, and gesticulation; and attention span, sexuality, craniofacial and neurodevelopment, gait, aging, and resistance to infectious agents (notably retroviruses). The implications for pharmaceutical sciences and such disciplines as rehabilitative medicine are profound.

SOCIETAL RESPONSIBILITY

It is worth discussing the limits of genomics. Understanding the human genome will change science and medicine in profound ways. It will, however, not solve or explain every important problem of the public health. Other components of society will need to express determination and commit resources for solutions. One such problem is the need to respect patient privacy. Another is racism in whatever guise, which is invariably an enemy of science. In our efforts to provide a reference sequence for the human genome, we believe there is neither ''good'' nor ''bad'' DNA, only human DNA. Every individual will have a finite number of genetic flaws in his or her genome. Our task is to use the modern tools of pharmacology and genomics to ameliorate these flaws or undo their consequences. There is very good reason for optimism on this front, as we have tried to convey in this chapter. Yet many societies, including our own, have at various times embraced eugenics and other irrational genetic theories of race or ethnicity as the justification for neglect, oppression, or worse. However, the embrace of these negative philosophies is not inevitable. Opening incomparable opportunities for preventing and curing illnesses through genomic science is one way of refuting these pseudoscientific viewpoints. We believe that various governmental and private agencies, especially the National Institutes of Health, need to redouble their efforts to provide resources in the arena of ethics, education, and genomic research. Furthermore, some illnesses have their roots in certain external envi-

ronmental factors for which genes may not affect outcomes in practical terms. Poverty may be viewed as one such factor. It is, therefore, important for society at large to recognize that advances in science alone, including the reference human genome project, will need to be coupled with programs to address these larger societal issues.

ACKNOWLEDGMENTS

We wish to thank Mary Whiteley, Trevor Woodage, Anibal Cravchik, and Emily Winn-Deen for their helpful scientific input, and Beth Hoyle for her excellent editorial assistance in preparing this manuscript.

Visit the Annual Reviews home page at www.AnnualReviews.org.

LITERATURE CITED

1. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, et al. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* 63(2):595–612
2. Kittles RA, Long JC, Bergen AW, Eggert M, Virkkunen M, et al. 1999. Cladistic association analysis of Y chromosome effects on alcohol dependence and related personality traits. *Proc. Natl. Acad. Sci. USA* 96(7):4204–9
3. Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, et al. 1998. Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* 62(5):1171–79
4. Harris EE, Hey J. 1999. X chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. USA* 96:3320–24
5. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, et al. 1998. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.* 19(3):233–40
6. Hästabacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat. Genet.* 2:204–11
7. Kruglyak L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nat. Genet.* 17:21–24
8. Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273(5281):1516–17
9. Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A. 1998. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res.* 8:111–23
10. McKeigue PM. 1998. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning or parental admixture. *Am. J. Hum. Genet.* 63:241–51
11. Weiss KM. 1998. In search of human variation. *Genome Res.* 8(7):691–97
12. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22:231–38
13. Halushka MK, Fan J-B, Bentley K, Hsie L, Shen N, et al. 1999. Patterns of single-nucleotide polymorphisms in candidate

genes for blood-pressure homeostasis. *Nat. Genet.* 22:239–47
14. Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. 1998. Shotgun sequencing of the human genome. *Science* 280:1540–42
15. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512
16. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. 1995. The minimal gene complement of *Mycoplasma genitalium. Science* 270:397–403
17. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* 273:1058–73
18. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, et al. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3:109–36
19. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae. Nucleic Acids Res.* 24:4420–49
20. The yeast genome directory. 1997. *Nature* 387(Suppl.):5
21. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* 388:539–47
22. Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–74
23. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, et al. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* DH: functional analysis and comparative genomics. *J. Bacteriol.* 179: 7135–55
24. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis. Nature* 390:249–56
25. Klenk H-P, Clayton RA, Tomb J-F, White O, Nelson KE, et al. 1997. The complete genome sequence of the hyperthermophilic, sulfate-reducing archeon *Archaeoglobus fulgidus. Nature* 390: 364–70
26. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi. Nature* 390:580–86
27. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, et al. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus. Nature* 392:353–58
28. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, et al. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* 5:147–55
29. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–44
30. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, et al. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281:375–88
31. Stephens RS, Kalman S, Lammel CJ, Fan J, Marathe R, et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis. Science* 282:754–59
32. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UCM, et al. 1998. The genome sequence of *Rickettsia prowazekii* and

the origin of mitochondria. *Nature* 396:133–40
33. The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans:* a platform for investigating biology. *Science* 282:2012–18
34. Alm RA, Ling L-SL, Moir DT, King BL, Brown ED, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori. Nature* 397:176–80
35. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, et al. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis. Nat. Genet.* 21:385–89
36. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima. Nature* 399:323–29
37. Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, et al. 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* 6:83–101
38. Wahlgren M. 1999. Creating deaths from malaria. *Nat. Genet.* 22(2):120–21
39. Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, et al. 1999. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat. Genet.* 22(2):145–50
40. Gardner MJ, Tettelin H, Carucci DJ, Cummings LM, Aravind L, et al. 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum. Science* 282:1126–32
41. Landegren U, Nilsson M, Kwok PY. 1998. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.* 8(8):769–76
42. Lee JT, Davidow LS, Warshawsky D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* 21(4):400–4
43. Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22(2):139–44
44. Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, et al. 1997. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N. Engl. J. Med.* 336:1401–8
45. Woodage T, King SM, Wacholder S, Hartge P, Struewing JP, et al. 1998. The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nat. Genet.* 20:62–65
46. Cascorbi I, Brockmoller J, Mrozikiewicz PM, Bauer S, Loddenkemper R, Roots I. 1996. Homozygous rapid arylamine N-acetyltransferase (NAT2) genotype as a susceptibility factor for lung cancer. *Cancer Res.* 56:3961–66
47. Chen J, Giovannucci E, Kelsey K, Rimm EB, Stampfer MJ, et al. 1996. A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res.* 56:4862–64
48. Ma J, Stampfer MJ, Giovannucci E, Artigas C, Hunter DJ, et al. 1997. Methylenetetrahydrofolate reductase polymorphism, dietary interactions and risk of colorectal cancer. *Cancer Res.* 57: 1098–102
49. Rebbeck TR, Kantoff PW, Krithivas K, Neuhausen S, Blackwood MA, et al. 1999. Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. *Am. J. Hum. Genet.* 64:1371–77
50. Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, et al. 1998. Role of a p53 polymorphism in the development of human papilloma-virus-associated cancer. *Nature* 393:229–34
51. Helland A, Langerod A, Johnsen H, Olsen AO, Skovlund E, Borresen-Dale

AL. 1998. p53 polymorphism and risk of cervical cancer. *Nature* 396:530–32
52. Josefsson AM, Magnusson PKE, Ylitalo N, Quarforth-Tubbin P, Ponten J, et al. 1998. p53 polymorphism and risk of cervical cancer. *Nature* 396:530–32
53. Hildesheim A, Schiffman M, Brinton LA, Fraumeni JF Jr, Herrero R, et al. 1998. p53 polymorphism and risk of cervical cancer. *Nature* 396:530–32
54. Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, et al. 1997. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science* 277: 959–65
55. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, et al. 1996. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382:722–25
56. Wilkinson DA, Operskalski EA, Busch MP, Mosley JW, Koup RA. 1998. A 32-bp deletion within the CCR5 locus protects against transmission of parenterally acquired human immunodeficiency virus but does not affect progression to AIDS-defining illness. *J. Infect. Dis.* 178:1163–66
57. Ntoumi F, Mercereau-Puijalon O, Ossari S, Luty A, Reltien J, et al. 1997. *Plasmodium falciparum:* sickle-cell trait is associated with higher prevalence of multiple infections in Gabonese children with asymptomatic infections. *Exp. Parasitol.* 87:39–46
58. Lawrenz DR. 1999. Sickle cell disease: a review and update of current therapy. *J. Oral. Maxillofac. Surg.* 57:171–78
59. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, et al. 1999. HLA and HIV-1: heterozygote advantage and B35-Cw04 disadvantage. *Science* 283: 1748–52
60. Fernandez-Reyes D, Craig AG, Kyes SA, Peshu N, Snow RW, et al. 1997. A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum. Mol. Genet.* 6:1357–60
61. Bellamy R, Kwiatkowski D, Hill AV. 1998. Absence of an association between intercellular adhesion molecule 1, complement receptor 1 and interleukin 1 receptor antagonist gene polymorphisms and severe malaria in a West African population. *Trans. R. Soc. Trop. Med. Hyg.* 92:312–16
62. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. 1998. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N. Engl. J. Med.* 338:640–44
63. Martinez FD, Graves PE, Baldini M, Solomon S, Erickson R. 1997. Association between genetic polymorphisms of the beta2-adrenoceptor and response to albuterol in children with and without a history of wheezing. *J. Clin. Invest.* 100:3184–88
64. Dahl M, Tybjaerg-Hansen A, Lange P, Nordestgaard BG. 1998. F508 heterozygosity in cystic fibrosis and susceptibility to asthma. *Lancet* 351:1911–13
65. Green SL, Gaillard MC, Song E, Dewar JB, Halkas A. 1998. Polymorphisms of the beta chain of the high-affinity immunoglobulin E receptor (Fc epsilon RI-beta) in South African black and white asthmatic and nonasthmatic individuals. *Am. J. Respir. Crit. Care Med.* 158:1487–92
66. Hijazi Z, Haider MZ, Khan MR, Al-Dowaisan AA. 1998. High frequency of IgE receptor Fc epsilon RI-beta variant (Leu181/Leu183) in Kuwaiti Arabs and its association with asthma. *Clin. Genet.* 53:149–52
67. Hill MR, Cookson WOCM. 1996. A new variant of the beta subunit of the high-

affinity receptor for immunoglobulin E (Fc epsilon RI-beta E237G): associations with measures of atopy and bronchial hyperresponsiveness. *Hum. Mol. Genet.* 5:959–62

68. Apostolakis J, Toumbis M, Konstantopoulos K, Kamaroulias D, Anagnostakis V, et al. 1996. HLA antigens and asthma in Greeks. *Respir. Med.* 90:201–4
69. Gerbase-DeLima M, Gallo CA, Daher S, Sole D, Naspitz CK. 1997. HLA antigens in asthmatic children. *Pediatr. Allergy Immunol.* 8:150–52
70. Drazen JM, Yandava CN, Dube L, Szczerback N, Hippensteel R, et al. 1999. Pharmacogenetic association between *ALOX5* promoter genotype and the response to anti-asthma treatment. *Nat. Genet.* 22:168–70
71. Stafforini DM, Numao T, Tsodikov A, Vaitkus D, Fukuda T, et al. 1999. Deficiency of platelet-activating factor acetylhydrolase is a severity factor for asthma. *J. Clin. Invest.* 103:989–97
72. Arranz MJ, Li T, Munro J, Liu X, Murray R, et al. 1998. Lack of association between a polymorphism in the promoter region of the dopamine-2 receptor gene and clozapine response. *Pharmacogenetics* 8:481–84
73. Ram A, Cao Q, Keck PE Jr, Pope HG Jr, Otani K, et al. 1995. Structural change in dopamine D2 receptor gene in a patient with neuroleptic malignant syndrome. *Am. J. Med. Genet.* 60:228–30
74. Cravchik A, Sibley DR, Gejman PV. 1999. Analysis of neuroleptic binding affinities and potencies for the different human D2 dopamine receptor missense variants. *Pharmacogenetics* 9:17–23
75. Nothen MM, Chicon S, Hemmer S, Hebebrand J, Remschmidt H, et al. 1994. Human dopamine D4 receptor gene: frequent occurrence of a null allele and observation of homozygosity. *Hum. Mol. Genet.* 3:2207–12
76. Gelernter J, Kranzler H, Coccaro E, Siever L, New A, Mulgrew CL. 1997. D4 dopamine-receptor (DRD4) alleles and novelty seeking in substance-dependent, personality-disorder, and control subjects. *Am. J. Hum. Genet.* 61:1144–52
77. Kotler M, Cohen H, Segman R, Gritsenko I, Nermanov L, et al. 1997. Excess dopamine D4 receptor (D4DR) exon III seven repeat allele in opioid-dependent subjects. *Mol. Psychiatry* 2:251–54
78. Cravchik A, Gejman PV. 1999. Functional analysis of the human D5 dopamine receptor missense and nonsense variants: differences in dopamine binding affinities. *Pharmacogenetics* 9:199–206
79. Cook EH Jr, Stein MA, Krasowski MD, Cox NJ, Olkon DM, et al. 1995. Association of attention-deficit disorder and the dopamine transporter gene. *Am. J. Hum. Genet.* 56:993–98
80. Arranz MJ, Munro J, Sham P, Kirov G, Murray RM, et al. 1998. Meta-analysis of studies on genetic variation in 5-HT2A receptors and clozapine response. *Schizophr. Res.* 32:93–99
81. Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, et al. 1996. Association of anxiety-related traits with a polymorphism in the seratonin transporter gene regulatory region. *Science* 274:1527–31
82. Marez D, Legrand M, Sabbagh N, Guidice JM, Spire C, et al. 1997. Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles. *Pharmacogenetics* 7:193–202
83. May DG. 1994. Genetic differences in drug disposition. *J. Clin. Pharmacol.* 34:881–97
84. Linder MW, Prough RA, Valdes R Jr. 1997. Pharmacogenetics: a laboratory tool for optimizing therapeutic efficiency. *Clin. Chem.* 43:254–66
85. Aithal GP, Day CP, Kesteven PJ, Daly AK. 1999. Association of polymorphisms in the cytochrome P450 CYP2C9 gene with warfarin dose requirement and

risk of bleeding complications. *Lancet* 353:717–19
86. Hashimoto Y, Otsuki Y, Odani A, Takano M, Hattori H, et al. 1996. Effect of CYP2C polymorphisms on the pharmacokinetics of phenytoin in Japanese patients with epilepsy. *Biol. Pharm. Bull.* 19:1103–5
87. Stubbins MJ, Harries LW, Smith G, Tarbit MH, Wolf CR. 1996. Genetic analysis of the cytochrome P450 PYP2C9 locus. *Pharmacogenetics* 6:429–39
88. Karam WG, Goldstein JA, Lasker JM, Ghanayem BI. 1996. Human CYP2C19 is a major omeprazole 5-hydroxylase, as demonstrated with recombinant cytochrome P450 enzymes. *Drug Metab. Dispos.* 24:1081–87
89. Inoue K, Yamazaki H, Shimada T. 1998. Linkage between the distribution of mutations in the CYP2C18 and CYP2C19 genes in the Japanese and Caucasian. *Xenobiotica* 28:403–11
90. Murata M, Shiraishi T, Fukutome K, Watanabe M, Nagao M, et al. 1998. Cytochrome P4501A1 and glutathione S-transferase M1 genotypes as risk factors for prostate cancer in Japan. *Jpn. J. Clin. Oncol.* 28:657–60
91. Bell DA, Taylor JA, Paulson DF, Robertson CN, Mohler JL, Lucier GW. 1993. Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen metabolism gene glutathione S-transferase M1 (GST M1) that increases susceptibility to bladder cancer. *J. Natl. Cancer Inst.* 85:1159–63
92. Okkels H, Sigsgaard T, Wolf H, Autrup H. 1997. Arylamine N-acetyltransferase 1 (NAT1) and 2 (NAT2) polymorphisms in susceptibility to bladder cancer: the influence of smoking. *Cancer Epidemiol. Biomarkers Prev.* 6:225–31
93. Spielberg SP. 1996. N-acetyltransferases: pharmacogenetics and clinical consequences of polymorphic drug metabolism. *J. Pharmacokinet. Biopharm.* 24:509–19
94. Butcher NJ, Ilett KF, Mincnin RF. 1998. Functional polymorphism of the human arylamine N-acetyltransferase type 1 gene caused by C190T and G560A mutations. *Pharmacogenetics* 8:67–72
95. Grant DM, Hughes NC, Janezic SA, Goodfellow GH, Chen HJ, et al. 1997. Human acyltransferase polymorphisms. *Mutat. Res.* 376:61–70
96. Davies HG, Richter RJ, Keifer M, Broomfield CA, Sowalla J, Furlong CE. 1996. The effect of the human serum paraoxonase polymorphism is reversed with diazoxon, soman and sarin. *Nat. Genet.* 14:334–36
97. Furlong CE, Li WF, Costa LG, Richter RJ, Shih DM, et al. 1998. Genetically determined susceptibility to organophosphorus insecticides and nerve agents: developing a mouse model for the human PON1 polymorphism. *Neurotoxicology* 19:645–50
98. Otterness D, Szumlanski C, Lennard L, Klemetsdal B, Aarbakke J, et al. 1997. Human thiopurine methyltransferase pharmacogenetics: gene sequence polymorphisms. *Clin. Pharmacol. Ther.* 62: 60–73
99. Escousse A, Guedon F, Mounie J, Rifle G, Mousson C, D'Athis P. 1998. 6-Mercaptopurine pharmacokinetics after use of azathioprine in renal transplant recipients with intermediate or high thiopurine methyl transferase activity phenotype. *J. Pharm. Pharmacol.* 50:1261–66
100. Black AJ, McLeod HL, Capell HA, Powrie RH, Matowe LK, et al. 1998. Thiopurine methyltransferase genotype predicts therapy-limiting severe toxicity from azathioprine. *Ann. Intern. Med.* 129(9):716–18
101. Kamboh MI. 1995. Apolipoprotein E polymorphism and susceptibility to Alzheimer's disease. *Hum. Biol.* 67(2):195–215
102. Roses AD. 1996. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* 47:387–400

103. Cooke GE, Bray PF, Hamlington JD, Pham DM, Goldschmidt-Clermont PS. 1998. PlA2 polymorphism and efficacy of aspirin. *Lancet* 351(9111):1253
104. Margaglione M, Brancaccio V, Giuliani N, D'Andrea G, Cappucci G, et al. 1998. Increased risk for venous thrombosis in carriers of the prothrombin G → A^{20210} gene variant. *Ann. Intern. Med.* 129(2):89–93
105. Daly AK. 1995. Molecular basis of polymorphic drug metabolism. *J. Mol. Med.* 73:539–53
106. Gutman S. 1999. The role of Food and Drug Administration regulation of in vitro diagnostic devices—applications to genetics testing. *Clin. Chem.* 45(5):746–49
107. Holtzman NA. 1999. Promoting safe and effective genetic tests in the United States: work of the task force on genetic testing. *Clin. Chem.* 45(5):732–38
108. Krynetski EY, Evans WE. 1998. Pharmacogenetics of cancer therapy: getting personal. *Am. J. Hum. Genet.* 63:11–16
109. Nebert DW, McKinnon RA, Puga A. 1996. Human drug metabolizing enzyme polymorphisms: effects on risk of toxicity and cancer. *DNA Cell Biol.* 15:273–80
110. Sparreboom A, van Asperen J, Mayer U, Schinkel AH, Smit JW, et al. 1997. Limited oral bioavailability and active epithelial excretion of paclitaxel (Taxol) caused by P-glycoprotein in the intestine. *Proc. Natl. Acad. Sci. USA* 94(5):2031–35
111. Sanak M, Simon HU, Szczeklik A. 1997. Leukotriene C4 synthase promoter polymorphism and risk of aspirin-induced asthma. *Lancet* 350:1599–600
112. Spring J. 1997. Vertebrate evolution by interspecific hybridisation—are we polyploid? *FEBS Lett.* 400(1):2–8
113. Burglin TR. 1996. Homeodomain in proteins. In *Encyclopedia of Molecular Biology and Molecular Medicine*, ed. RA Meyers, 3:55–76. Weinheim, Ger.: VCH Verlagsgesellschaft
114. Glaser T, Walton DS, Maas RL. 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nat. Genet.* 2(3):232–39
115. Quiring R, Walldorf U, Kloter U, Gehring WJ. 1994. Homology of the *eyeless* gene of Drosophila to the *Small eye* gene in mice and Aniridia in humans. *Science* 265(5173):785–89
116. Becker W, Joost HG. 1999. Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. *Prog. Nucleic Acid Res. Mol. Biol.* 62:1–17
117. Ohira M, Seki N, Nagase T, Suzuki E, Nomura N, et al. 1997. Gene identification in 1.6-Mb region of the Down syndrome region on chromosome 21. *Genome Res.* 7(1):47–58
118. Guimera J, Casas C, Pucharcos C, Solans A, Domenech A, et al. 1996. A human homologue of Drosophila minibrain (MNB) is expressed in the neuronal regions affected in Down syndrome and maps to the critical region. *Hum. Mol. Genet.* 5(9):1305–10
119. Reeves RH, Gearhart JD, Littlefield JW. 1986. Genetic basis for a mouse model of Down syndrome. *Brain Res. Bull.* 16(6):803–14
120. Haydar TF, Blue ME, Molliver ME, Krueger BK, Yarowsky PJ. 1996. Consequences of trisomy 16 for mouse brain development: corticogenesis in a model of Down syndrome. *J. Neurosci.* 16(19):6175–82
121. Smith DJ, Rubin EM. 1997. Functional screening and complex traits: human 21q22.2 sequences affecting learning in mice. *Hum. Mol. Genet.* 6(10):1729–33
122. Smith DJ, Stevens ME, Sudanagunta SP, Bronson RT, Makhinson M, et al. 1997. Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects

associated with Down syndrome. *Nat. Genet.* 16(1):28–36
123. Clark AG. 1999. Chips for chimps. *Nat. Genet.* 22(2):119–20
124. Hacia JC, Fan J-B, Ryder O, Jin L, Edgemon K, et al. 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide assays. *Nat. Genet.* 22:164–67
125. Goodman M. 1999. Molecular evolution '99: the genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* 64:31–39
126. Czelusniak J, Goodman M, Hewett-Emmett D, Weiss ML, Venta PJ, Tashian RE. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* 298(5871):297–300
127. Chiu C-H, Schneider H, Slightom JL, Gumucio DL, Goodman M. 1997. Dynamics of regulatory evolution in primate β-globin gene clusters: cis-mediated acquisition of simian γ fetal expression patterns. *Gene* 205:47–57
128. Chiu C-H, Gregoire L, Gumucio DL, Muniz JAPC, Lancaster WD, Goodman M. 1999. Model for the fetal recruitment of simian γ-globin genes based on findings from two new world monkeys *Cebus apella* and *Callithrix jacchus* (Platyrrhini, primates). *J. Exp. Zool.* 285:27–40
129. Hayasaka K, Fitch DHA, Slightom JL, Goodman M. 1992. Fetal recruitment of anthropoid γ-globin genes. *J. Mol. Biol.* 224:875–81
130. McConkey EH, Goodman M. 1997. A human genome evolution project is needed. *Trends Genet.* 13:350–51
131. Dutrillaux B. 1979. Chromosomal evolution in primates: tentative phylogeny from *Microcebus murinus* (Prosimian) to man. *Hum. Genet.* 48:251–314
132. Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* 215:1525–30

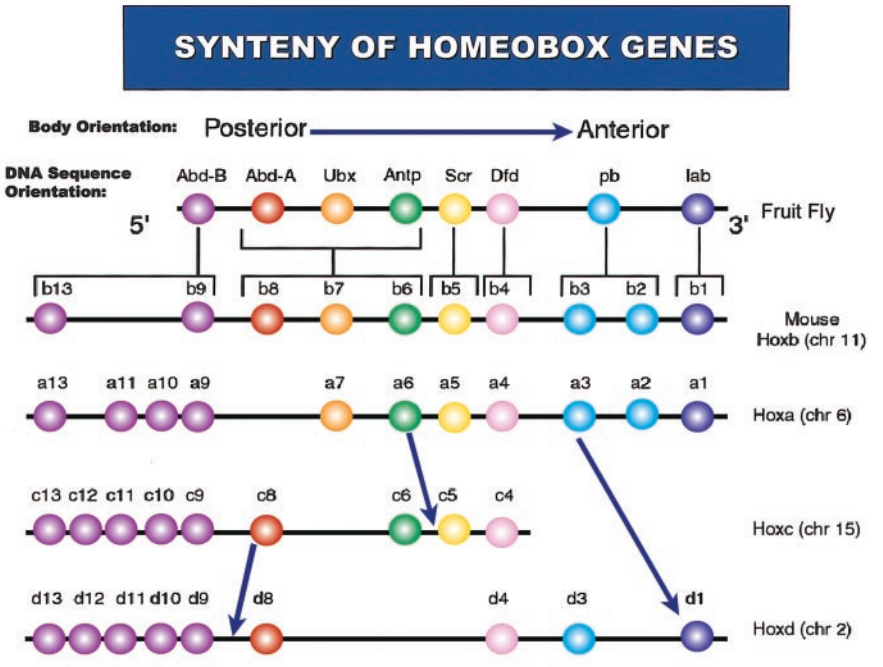

Figure 1 Synteny of homeobox (*Hox*) genes. (*Color coding*) Shows evolutionary relationships of orthologs. *Abd-B*, Abdominal B; *Abd-A*, abdominal A; *Ubx*, ultrabiothorax; *Antp*, antennapedia; *Scr*, sex combs reduced; *Dfd*, deformed; *Pb*, proboscipedia; *Lab*, labial.

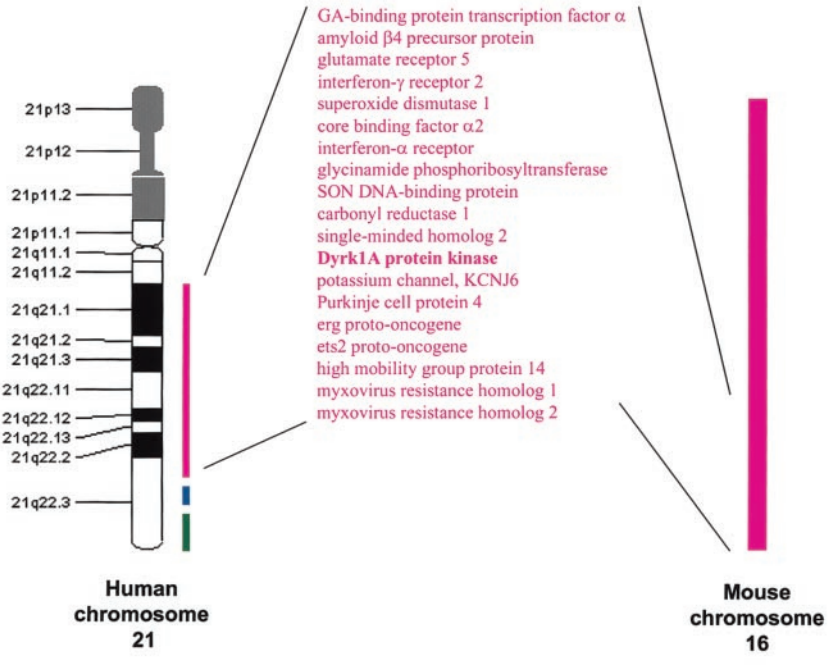

Figure 2 Synteny (*red*) between human chromosome 21 and mouse chromosome 16. *Dyrk1A* protein kinase is in a Down's syndrome critical region and is a candidate gene for causation. (See also http://www.ncbi.nlm.nih.gov/Homology/.)