Progress in Drug Research, Vol. 58 (E. Jucker, Ed.)

Drugs, the human genome, and individual-based medicine

By Jay A. Glasel

Global Scientific Consulting LLC
15 Colton St.
Farmington, CT 06032, USA
<jaglasel@consult-globalsci.com>

Jay A. Glasel

studied Chemistry and Physics at the California Institute of Technology and received his Ph.D. in Chemical Physics at the University of Chicago. He is managing member of Global Scientific Consulting LLC in Farmington, Connecticut and Professor Emeritus of Biochemistry at the University of Connecticut Medical-Dental School. He is senior author of an extensive list of research articles in scientific publications.

Summary

The so-called "Genomic Revolution" has made possible the high-resolution sequencing of the DNA making up the human genome. One of the main conclusions of the currently available sequencing data is that individuals differ genetically from one another *via* sequence variations in their genomes. When affected genes are transcribed and translated, some of these sequence variations result in protein products that may affect the functioning of the proteins. This has led to widespread optimism that information on an individual's pattern of sequence variations will lead to drugs that target that individual's variant proteins and make "individual-based medicine" possible. In this chapter some of the assumptions underlying the proposed production of individual drug treatments are examined. The assumptions are viewed in the light of very recent experimental evidence about the sequence patterns found in humans. Also discussed are ancillary ethical problems in cataloging and using databases containing individuals' sequence data, what human genomic sequences are revealing about the use of animal models in developing drugs, and how evidence is mounting that the human genome is only one element serving to maintain an organism's interaction with its environment.

Contents

Key words

Pharmacogenetics, pharmacogenomics, epigenetics, human genome, single nucleotide polymorphisms, molecular genetics, individual medicine, drug development.

Glossary of abbreviations

Anchored re-sequencing strategy: An expensive approach to obtaining a large number of SNPs using short sequences of DNA from individuals of different ethnic groups that are derived from a plasmid library. These sequences are then compared to the genome draft sequence to obtain a large number of SNPs in a very short amount of time.

Annotation (annotated sequence): A term used with very little consistency among researchers working on genome-related projects. As commonly used, annotation entails assembling information of several distinctive types, starting with refined DNA sequence data but extending beyond that level to varying degrees of complexity. For example, completed DNA sequence information may be segmented into distinct intervals that may be demarcated in terms of their encoding specific types of product, such as proteins. Next, a particular gene may be annotated in such terms as its protein coding region, its transcript, promoter region, and so forth. At a higher level of annotation, a protein that is encoded by a particular gene may be annotated in terms of its physical attributes, such as molecular weight, membrane spanning regions, structural domains, or three-dimensional structure. Annotation at the level of comparative biology may include information linking a particular protein from a specific organism to similar proteins from other organisms or to members of similar protein families. Genes may also be annotated at a functional level, in terms of their respective roles in cellular metabolism, a particular systematic enzyme number (EC) designation, protein-protein interactions, and expression profiles.

Contigs (contig map): Overlapping (contiguous) DNA segments assembled from sequenced fragments of a chromosome. A contig map is the sequence of DNA in a chromosome constructed by isolating contigs.

Epigenetics: The study of heritable changes in gene expression that occur without a change in DNA sequence.

EST: Expressed sequence tag, several hundred base pairs of known cDNA sequence flanked by PCR primers. Because they are derived from cDNA libraries (i.e., from reverse transcription of cellular mRNAs), ESTs represent portions of expressed genes. ESTs are useful for identifying full-length genes and a landmark for mapping.

Finished sequence: Complete sequence of a clone or of a genome with an accuracy of at least 99.99% and no gaps.

Functional genomics: The study of genomes to determine the biological function of all the genes and their products.

Genetic map: The ordering of genes on chromosomes according to their recombination frequency during meiosis. The unit of distance is centimorgans (cM) which corresponds to a 1%

chance of recombination. Genetic mapping preceded physical mapping and when a genome is completely sequenced, the genetic map becomes redundant.

Genetic linkage map: A map of the relative positions of genes and other regions on a chromosome, determined by how often loci are inherited together.

Genome draft sequence: The sequence produced by combining information from individual overlapping sequenced pieces of DNA with linkage information and positioning the sequence along the physical map of the chromosomes.

Haplotype: Long stretches of DNA (lengths as long as 100,000 bases have been found) at a given location on a chromosome. Also, a specific pattern of alleles or sequence variations that are closely linked (i.e., likely to be inherited together) on the same chromosome. Recently, it has been found that many such blocks come in a few different versions.

Linkage equilibrium, disequilibrium: When there is no preferential association of alleles at linked loci, they are said to be in equilibrium; when there is a nonrandom association of alleles at linked loci in a population, the population is said to be in linkage disequilibrium. As sometimes stated oppositely: at 0% disequilibrium, there is no association between alleles, at 100% disequilibrium, alleles are always found in association.

Microsatellite: A type of satellite DNA that consists of small repeat units (usually two, three, four, or five base pairs) that occur in tandem.

Minisatellite: A type of satellite DNA that consists of tandem repeat units that are each about 20 to 70 base pairs in length.

Missense (mutation): A DNA sequence change that results in a single amino acid change in the translated gene product.

Neutral hypothesis and neutral model: The hypothesis is that the large majority of polymorphisms within species and fixed substitutions between species are the result of the random drift of neutral mutations and not natural selection. Deleterious mutations are assumed to occur, but are quickly eliminated. The model assumes a random-mating population of constant size where new mutations occur at sites not previously mutated (a good approximation when mutation rates at all sites are relatively low).

Nonsense (mutation): A DNA sequence in which an mRNA stop codon is produced or removed, thus resulting in a premature termination of translation or an elongation of the protein product.

π (Pi): A measure of nucleotide diversity (per-site heterozygosity) in the sequences of single strands of DNA. It is the probability that a pair of chromosomes drawn from a population will differ at a nucleotide site.

Pseudogene: A nonfunctional sequence of DNA similar to a functional one. Pseudogenes are probably remnants of once-functional genes that accumulated mutations.

Radical (change in an amino acid residue): A commonly used, somewhat arbitrarily defined description of an amino acid substitution in a protein that is predicted to, or does, change the protein's observed properties. As discussed in this chapter, the concept has been given a more quantitative form.

Recombination frequency: The proportion of meioses in which recombinants between two loci are seen. The proportion rises the farther apart the recombining loci are. The recombination frequency is used to estimate genetic distances between loci.

RFLP: Restriction fragment length polymorphism; genetic variation in the length of DNA fragments produced by restriction enzymes. RFLPs are useful as markers on maps.
Satellite DNA: A portion of the DNA that differs enough in base composition that it is physically separable from other DNAs by its different density; usually contains highly repetitive DNA sequences.
Silent (mutation): A DNA sequence change that does not change the amino acid sequence in the translated protein product due to the degeneracy of the genetic code.
SNP: Single-nucleotide polymorphism; common DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered. Other types of polymorphisms include VNTRs (variable number of tandem repeats; a type of polymorphism created by variations in the number of minisatellite repeats in a defined region of the sequence) and MRPs (microsatellite repeat polymorphism; a type of genetic variation in populations consisting of differing numbers of microsatellite repeat units at a locus).
SNP map: A collection of SNPs that can be superimposed over the existing genome map, creating greater detail, and facilitating further genetic studies.
STS: Sequence tagged site, a unique stretch of DNA whose location is known. STSs serve as landmarks for mapping and assembly.
Trans-splicing: A process whereby a transcript, translatable into a protein, is produced from both the sense and antisense strands of a DNA double helix. Previously observed in viruses, trans-splicing has now been found in a eukaryote (*Drosophila*) and may take place in humans.

1 Introduction

This chapter presents a critical look at the practical impact that increasing knowledge of human molecular genetics will have on drug development in the next 5–10 years. Biotechnological advances are taking place with such rapidity that longer-term projections are worthless guesswork.

Many of the terms used to describe human molecular genetics have not yet entered into common use in pharmacological/pharmaceutical literature. Therefore, a glossary has been included at the start of the chapter. It is hoped that the glossary will be useful in guiding some readers through this chapter, the references quoted therein, and the developing literature on drugs and genetics.

1.1 Drugs, the genome, proposals for "individual medicine"

Within the last five years, the pharmacological literature has been flooded with publications hailing what is usually taken for granted: determining sin-

gle nucleotide polymorphisms (SNPs) in an individual's genes will soon become a reality, and will inevitably result in what some have called "individual medicine." That is, drug and other clinical treatments will begin to be targeted to individuals on the basis of these individuals' own specific genetic variations. An examination of this widespread belief forms the basis for this chapter.

1.1.1 The genetic basis for the concept of individual medicine

The fundamental information upon which any practical application of individual medicine is based has been assumed to be the identification of genetic loci and functional significance of SNPs in an individual's genome.

The human genome is diploid, but the SNP information that is sought is haplotypic. In the most familiar conventional analyses of genomic DNA variation, both strands on each chromosome are analyzed together so that at every position in the DNA sequence where a variation occurs, the DNA sequence is heterozygous (two bases are identified instead of one). If multiple variant loci are present, there will be multiple heterozygous positions and the information of which variant is present on which DNA molecule is lost in these analyses. For example, if one allele is dominant (the most frequent case) it has been assumed that it will be the drugs directed against it, or its product, that are important to develop.

In conventional human genetic studies, the pattern of inheritance provides key information that allows most or all of the haplotypes to be reconstructed because, except for very rare recombination events, DNA molecules are inherited intact from parents. This allows the determination of which particular variant alleles have been coinherited from one parent and thus must lie on the same DNA molecule. The kinds of multi-generation linkage studies traditionally used in human genetics provide sufficient haplotype information that disease genes can be located based on the pattern of inheritance of nearby markers. It should be noted that such experiments are not the typical design used in clinical studies. Instead, case-control data are the norm, and this dramatically increases the difficulty of reconstructing haplotype information, because without this information, the statistical power of association of a particular DNA marker with a nearby phenotype becomes much weaker.

However, over several decades, studies of both types have resulted in the discovery of many variant genes whose products, or lack of products, are believed to cause disease in humans and animal models. These are the so-called "monogenic" diseases.

The technical ability to obtain gene sequence information from large numbers of chromosomes from large numbers of individuals' DNA has been recognized as the stumbling block to obtaining enough SNP information to be used for developing individual medicine. The background to this problem has been reviewed [1, 2] as have been technological advances to solve it [3].

At present, the practical situation is that technologies exist to obtain detailed SNP information for a limited number of genes from a relatively large population of individuals, or limited information from a large number of genes from a relatively small population of individuals. The information now available on both these types of studies will be discussed later in this chapter. The greatest effort now being made is an attempt to define all the SNPs in the human genome as obtained from DNA obtained from a few individuals.

1.1.2 Present progress in obtaining genomic SNP information

To the end of finding all SNPs in the human genome, the expensive anchored re-sequencing approach is the only strategy known that is rigorous enough to find the number of SNPs necessary to completely specify individual genetic variations. This strategy has been adapted cooperatively by a number of pharmaceutical firms along with public research facilities in a venture called the "SNP Consortium." The Consortium began pilot studies in January of 1999 and hopes to locate and map more than 150,000 SNPS by the end of 2001. All information collected by the groups involved is being published on The Consortium web site (http://snp.cshl.org).

1.1.3 Will genomics have an impact on clinical and pharmaceutical medicine?

At present, published opinions concerning the success of applying genomic sequencing data to produce new drugs – indeed how important the data will

be to clinical medicine in general – range from deep scepticism [4] to a call to "wake up and get ready" [1], with a heavy emphasis on the latter position.

1.2 The need for new ideas in drug development

Both ends of this scientific opinion spectrum agree on one basic fact: there is a need for new ideas in pharmaceutical development. Despite the increasing intensity of drug development efforts throughout the world, some challenges must be faced concerning the efficacy of established drugs and the development of efficacious new ones. The facts are these:

- On average, for each specific drug treatment, less than 50% of patients experience a benefit (for example, it has been stated that 35% of people do not respond to β-blockers and as many as 50% do not respond to tricyclic antidepressants).
- Although accurate quantitative figures are controversial [5–9] there is general agreement that a large number of patients who take prescribed drugs have severe adverse drug reactions (ADRs) [10, 11].
- According to one estimate [5], drug-related side-effects were between the fourth and sixth leading cause of death in the United States in 1997–about 106,000 deaths that year–and nothing has reduced these figures since.
- It has been reported that 30–50% of pharmacies dispense drugs with interactions that could be life-threatening or life-altering without warning to patient or prescriber.
- ADRs are a major cause for removing prescribed drugs from the market [12]. In developing new drugs, costs per compound typically run at $7 million in phase I, but jump to $43 million in phase III; abandoning an unpromising or dangerous compound at the earlier phase of development could save a pharmaceutical company $36 million per drug.

Not all directions are negative in bringing efficacious drugs to market and having them stay there:

- Many technical developments (e.g., combinatorial libraries, computer docking models, high-throughput screening, etc.) have led to an increase in the number of compounds under development for drug use. For exam-

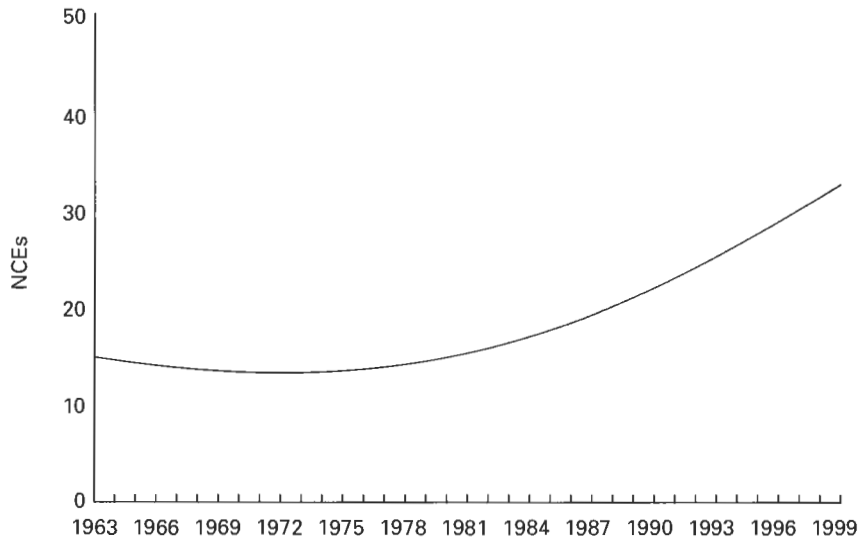


Fig. 1
New chemical entity (NCE) approvals in the United States from 1963 to 1999 as a quadratic fit to raw annual approval data. Adapted from [14].

ple, between 1995 and 1999, the number of compounds increased by 35% to 7434 [13].

- In the United States, several years following the 1962 Amendments to the Federal Food, Drug and Cosmetic Act of 1938, the number of NCEs (New Chemical Entities) approvals began to rise (Fig. 1). This upturn is likely to be due to a combination of the technical developments, such as those just cited, with governmental and economic changes [14].
- Following the Prescription Drug User Fee Act (PUDFA) of 1992 in the United States, drug approval times have fallen (Fig. 2) [15].
- Viewed by therapeutic classes, there is a variation in clinical phase times between the first half of the 1990s to the second half [15] (Fig. 3) with significant drops in the times for AIDS antiviral, antiinfective, anticancer, and respiratory classes, respectively, along with modest increases for CNS, endocrine, anesthetic/analgesic and cardiovascular classes, respectively.

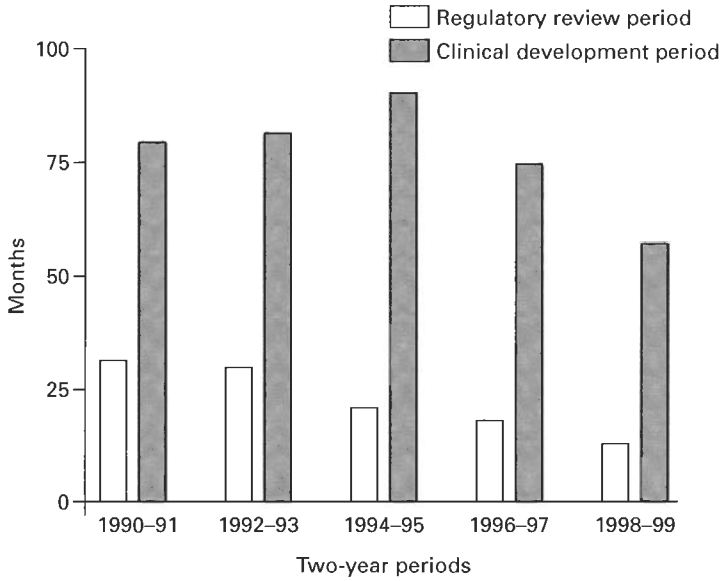


Fig. 2
Mean clinical development period [time from the date of investigational new drug application (IND) file to the date of new drug application (NDA)] and the mean regulatory review period [time from NDA submission to approval] after the PUDFA in five two-year periods. Adapted from [15].

On balance, despite these positive developments, the rate at which approved new drugs are entering clinical use is falling, and has been for some time: during the period 1995–99, the number of new product launches declined 22% to 56, and at an estimated cost of $500 million Research and Development cost per drug [13]. Between 1961 and 1995, 131 approved drugs were withdrawn from the market in western European countries and the United States because of severe ADRs [16]. This is in the face of the fact that, compared to 20 years ago, a growing proportion of drugs is being prescribed and used long-term for treating chronic conditions, thus requiring the drugs to be free of side-effects [13].

To accomplish the aims of this chapter, recent results bearing on a few important medical areas – asthma and its treatment, the responses of humans to glucocorticoids, and the roles of drug-clearing enzymes – will be used as illustrations of some of the opportunities and some of the problems associ-

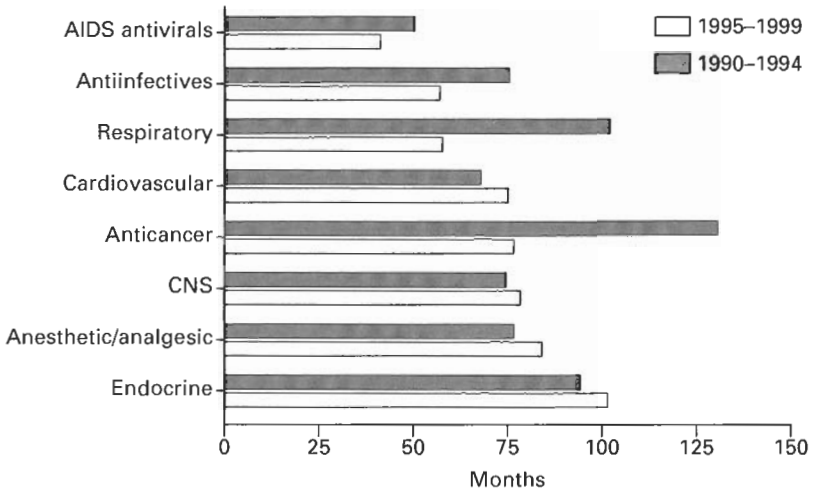


Fig. 3
Mean clinical phase for NCEs approved in 1990–94 and 1995–99, grouped by therapeutic class. Adapted from [15].

ated with producing individual medicine. However, it will be useful to begin by briefly reviewing how the relation between genes and health or disease are currently regarded in biomedical science.

2 Viewpoints on health and disease

2.1 Theoretical basis for "individual medicine"

The underlying reason why the belief in the future of individual medicine has taken such a hold on projections about the future of drug research is what may be termed a "genetocentric" viewpoint of human biology – one that permeates contemporary biological and medical thinking.

This genetocentric viewpoint is bolstered by a belief that the increasingly powerful technological tools becoming available to molecular geneticists will solve all technical problems. In this chapter, the further development of these tools is taken for granted. However, the genetocentric viewpoint is not taken for granted and its background, basis, and implications for development of individual medicine is examined using some specific recent examples.

2.1.1 The genetocentric viewpoint of health and disease

The genetocentric viewpoint is based on the reductionist concept of an organism as an automatic control entity. This idea, that organisms are controlled by instructions encoded into their genomes, has been a powerful influence in the explosive development of molecular biology and, up to now, on hopes for decisive clinical applications of genetic discoveries. According to a strict interpretation of this viewpoint, organisms are no different from subtle machines: the whole is the sum of its parts (in physical terms, organisms are "linear systems") which are arranged in such a way that an internal energy source can move them in accordance with a built-in program of purposeful action – a program encoded in the genome of each individual organism.

From the genetocentric viewpoint, prediction of health, diagnosis of disease, and targeting of drugs should largely be attainable from genomic information.

In an influential article about the Human Genome Project, the Head of the National Institutes of Health's effort, Francis Collins, stated the genetocentric case succinctly, "Largely, but not entirely, at the behest of our genes, we fare better or worse." [17].

Yet, it is a universally acknowledged fact that life expectancy in wealthy countries has increased greatly in the past century, and clearly this increase could not be due to genetic factors. This fact has led many to speculate that environmental factors are of equal or greater importance to human health than is genetic makeup. Despite some dissenting arguments of this sort, it is widely accepted in both the scientific and lay communities that advances in molecular biology corroborate Collins's claim: the idea that, in principle, organisms are no more than complicated physical systems and therefore biological principles can be reduced to physical and chemical laws.

Probably many (if not most) scientists, physical and biological alike, regard philosophical concepts like reductionism as woolly-minded and having little application to their work or to the practical implications of molecular genetics, especially in medicine. The immense outpouring of literature within the last two years on the implications of the results of current genetic research – particularly those stemming from the human genome projects – for the future of drug development and treatment largely assumes that it will be based on the genetic diversity of individuals as revealed by their individual genomic sequences.

But there is an alternate view of the nature of organisms that, while accepting the importance of the information contained in their genomes, looks at health and disease differently.

2.1.2 The epigenetic viewpoint of health and disease

This is an anti-reductionist viewpoint and contends that molecular genetics cannot explain all, or even the most important, aspects of living forms. Leaving aside the faith-based aspects of this view, there is increasing evidence and awareness that living processes all take place in a cellular environment and that this fact may largely influence the applications of ideas based strictly on genetic results.

Recently, this viewpoint – that living systems are non-linear ones – has come to be expressed with increasing frequency [18, 19] as new evidence mounts indicating that reductionism is not adequate to describe their health or diseases. Pertinent to the present topic, this viewpoint questions whether analysis of disease on the basis of single SNP, or even single gene, causality can be expected to help in the development of new drugs [19].

As discussed in the next section, evidence of this nature was first noticed by cell biologists but has recently been advanced in another context.

2.2 Cytoplasmic inheritance and epigenetics

As a background to understanding the consequences of the assumption that individualized drug treatments will be a necessary outcome of accurate sequencing of the human genome, this section includes a brief discussion of some often overlooked experiments in cell biology.

2.2.1 Cytoplasmic inheritance

In an elegant series of experiments performed over 30 years ago on ciliated paramecia, Sonneborn and collaborators first showed that the form and arrangement of preexisting parental cilia determine the form and arrangement of the new cilia on their progeny [20]. The cilia are small "hairs" that

are the locomotive organelles for the cells (mostly, they use them to swim toward food). On a molecular level, cilia are structurally complex although constructed mainly from proteins. The distribution of cilia on a cell varies among different taxonomic groups.

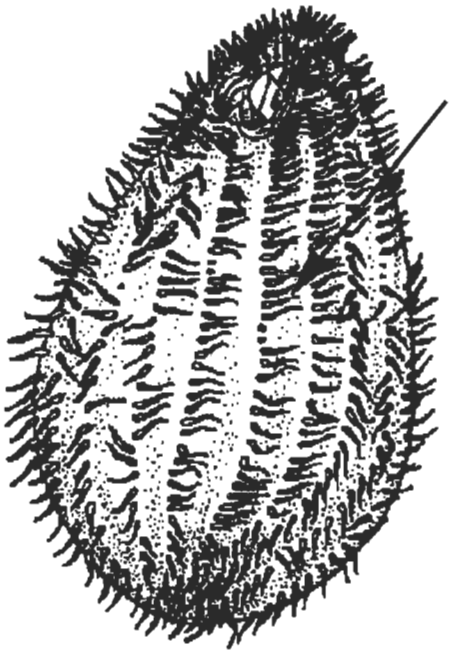
The architectural complexity of ciliates sets them apart from mammalian cells because besides being cells, they are complete, independent organisms. Adaptations to particular habitats over evolutionary time have resulted in both intracellular and extracellular structures seldom, if ever, found in higher eukaryotic cells. On the other hand, the metabolic pathways known for these animals are essentially no different from those found in other eukaryotic cells and they have been studied as good models for higher eukaryotic cells.

Asexual reproduction is the norm for ciliates: it involves nuclear division followed by binary fission of the cell that produces two identical daughter cells that each share some of the parental cytoplasm.

Work by Frankel on the ciliate *Tetrahymena pyriformis* [21] has given the most easily illustrated experimental results on the nongenetic inheritance of structural forms. *Tetrahymena* has its cilia arranged in rows that have a definite orientation along the longitudinal axis of the cell (Fig. 4). The motion of these cilia allow the cells to swim in more or less straight lines. Frankel used micromanipulation techniques on single cells to invert the orientation of one or more rows of their cilia by 180° with respect to the rest of the rows. After the operation, their swimming patterns were changed but the altered *Tetrahymena* could survive perfectly well and were allowed to divide normally.

In typical experiments, the swimming and ciliary patterns of the progeny *Tetrahymena* were observed over hundreds of cell divisions. These experiments showed that the inverted rows were perpetuated over those generations and passed on to all progeny, directly demonstrating the stability and determinism of a structural intracellular rearrangement in the absence of differences in genes or gene action. Occasionally, the "mutant" cells' cilial patterns reverted to normal after some generations, but the fastest reversions took 30–40 cell generations, during which they transmitted the variant organization of their cilia to a billion to a trillion progeny cells.

These cell biological experiments do not mean that genes play an unimportant role in cell organization: genes are, of course, needed to make the macromolecules that make the altered structures that are perpetuated. But the



Normally, the rows of cilia work together to rhythmically propel the cell in one direction. When one or more rows are reoriented 180 degrees from normal, the cell's swimming pattern is altered. This allows easy visual detection of cells with reoriented rows.

Fig. 4
Sketch of cilia in *Tetrahymena pyriformis*. Adapted from [21].

experiments raise some important points about cell replication in higher organisms – points that are often forgotten in genetocentric thinking.

The main point is that subcellular structures and organelles are not assembled in a vacuum, but in a structured cellular environment, the molecules of which are an essential part of a cell's means for locating, orienting and patterning new molecular structures or assemblies of molecules. The experiments just described show that physical interactions between existing structures and newly synthesized ones may result in hereditary extragenetic variations.

Thus, not only protein structures may be passed along to progeny extragenetically (as is the case for prions). The "blueprints" (i.e., information) for the cellular construction of complex structures made from multiple proteins can be passed to progeny as well.

The concept of non-genetically determined interacting signaling pathways has also been applied to human disease.

2.2.2 Epigenetics

The existence of non-genomic interactive networks in cells such as those described in the previous section is at the core of the epigenetic viewpoint. This viewpoint has recently been reviewed [19, 22]. Epigenetic ideas about regulation in cells and organisms emphasize the dependency of organisms on environments and the adaptive variations possible when are respond to environmental changes. Recently, an entire issue of the journal *Science* was devoted to a review of epigenetics. The electronic edition of the journal (http://www.sciencemag.org/feature/plus/sfg/resources/res_epigenetics.shtml) contains many connecting links to sites offering publications and information on epigenetics.

In epigenetic regulation, large networks of intracellular genes, gene products and groups of cells such as tissues and organs all absorb environmental signals and are active in producing adaptive behavior. These networks may seek out new pathways of response to varying inputs. The process is analogous to what is found in cell populations, in neural network activity, and in physiological organ system feedback interactions of diverse descriptions. It is a concept of organism based on past, conserved genetic adaptation coupled with current epigenetic regulation. Most importantly, it is a concept that attempts to integrate biology with molecular biology.

In epigenetics, genetic and higher-order networks operating within environmental limits defined during evolution determine healthy behavioral outcome, not single genes. On the other hand, when the world presents epigenetic information for which the genome and its interactive epigenetic network have no adequate response – for example, an overly oxidative environment – the result is maladaptation, regressive-state change in cell behavior and, finally, end-state disease. In this viewpoint, disease is a result of the organism's frustrated attempts to adapt phenotypically to a hostile environment or set of elements for which there is no adequate response basis.

In epigenetics, responses to environmental change may be seen as analogous to the functioning of the world-wide web (the Internet). The original intention in the development of the Internet was to allow military information transfer in the United States to take place even when some information relay points had been destroyed by enemy attack. The Internet was designed to allow packets, containing parts of the total message, to be transferred over varying routes and to be re-directed when some "nodes" (computer relay

points) are overloaded or non-functional. The packets are reassembled into the whole message at their final destination. Only in the cases of massive overload or destruction of the web does the system irreversibly break down.

Further underlying the epigenetic viewpoint is a model that predates molecular biology. It holds that evolution selects not single genes but genetic interaction [18, 23]. In this model, it is the physiological product of many genes interacting with one another, with many gene products and with environmental and developmental signals that provides advantage in, and is selected during, evolution.

In the epigenetic viewpoint, it would be very difficult, in cases where phenotype is clearly polygenic, to identify a particular single gene or even a set of genes that would always produce the same outcome, even under identical environmental circumstances. If one or a few genes, each with small effect, were deleted in an interactive network, one would expect multiple rearrangements of the entire network. In slightly differing circumstances, therefore, these rearrangements have some ability to produce adaptive phenotypes.

If there is even partial validity to the epigenetic viewpoint, there are broad implications concerning drug development. The main implication is that even in the case of monogenic diseases, targeting even the most efficacious drug to one gene or one gene product, may not suffice to either alleviate the disease or prevent ADRs. So the question immediately arises: Is there evidence for the validity of the epigenetic viewpoint of the organism in human disease?

2.2.3 Single- and multi-genic disease

Contemporary genetic research has been extremely successful in identifying the mutated genes for monogenic diseases and familial hypercholesteremia was one of the first to be identified. It furnishes a good example of the possible role of epigenetic concepts in how disease should be regarded.

Familial hypercholesterolemia arises from mutations in the low density lipoprotein receptor gene, carries an increased risk of coronary heart disease, and is one of the commonest autosomal dominant disorders with over 600 known genetic variants. But even for this monogenic disease, the relation between genotype and phenotype appears to be modified both by other genes and environmental effects. In a recent study [24] carried out in the

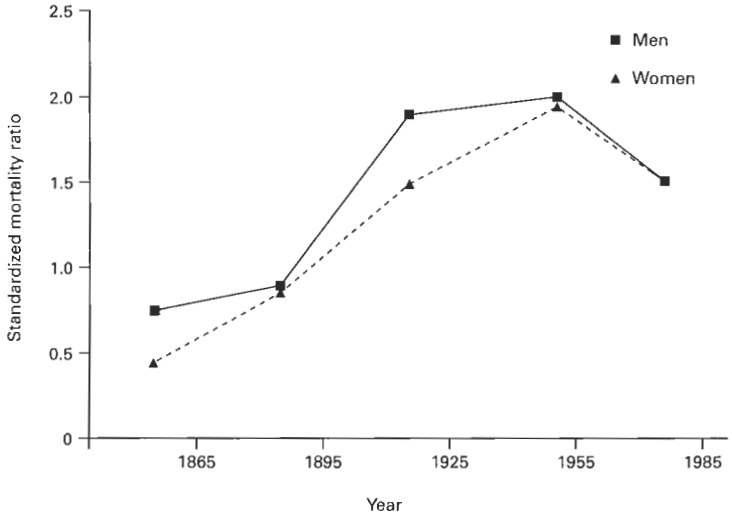


Fig. 5
Mortality from familial hypercholesterolemia according to sex and time. Adapted from [24].

Netherlands using historical records and designed to estimate all-cause mortality from untreated familial hypercholesterolemia free from selection for coronary artery disease, it was found that risk of death varies significantly among patients with the disease. The study concluded that the large variability in risk of mortality over time and between branches of the pedigree points to a strong interaction with environmental factors (Fig. 5). In fact, the mortality in the pedigree did not differ significantly from national rates until early in the 20th century. The importance of environmental factors in familial hypercholesterolemia has been found in other studies from the United States [25] and Finland [26].

The overall conclusion from these studies is that even in monogenic disease, other genes and the environment can be important and must be taken into account in risk analysis and clinical treatment of patients.

Furthermore, while new monogenic diseases are being discovered with increasing frequency and reported with much publicity, they affect only small segments of the population – altogether they have a frequency of about 2% of live births [27]. Also, there is no evidence that major diseases like cancers, diabetes and cardiovascular disease result from either single-gene muta-

tions or any other unitary cause. On the contrary, it is believed that the major human diseases are polygenic, have environmental determinants, and that the mechanisms of these diseases cannot be analyzed by a strictly genetic approach [18, 28].

Reductionism as currently applied to medicine asserts that, while they may have an environmental component, monogenic diseases such as hypercholesterolemia and more complex ones all have separate genetic components that can be discovered and used to pursue treatment strategies. This linear view of gene-disease causality thus finds itself in serious debate with a significant amount of evidence supporting the epigenetic viewpoint which sees complex traits, including disease, as highly interactive and impossible to separate into genetic components.

To understand how interactive networks may affect drug developments and treatments, we need to review what is currently known about drugs and the genome.

3 Drugs and human genomics

Under most current proposals to implement individual medicine, drug developers would apply SNP data for the human genome to identify affected genes and their variant proteins as drug targets. Thus, global attention has been forcefully directed toward building SNP databases and developing cost-effective haplotyping strategies for humans. The complexities associated with SNP-based drug development are just beginning to be realized.

3.1 An abbreviated history of pharmacogenomics

In the published literature, two terms describing individual responses to drugs – "pharmacogenetics" and "pharmacogenomics" – have often been used interchangeably or in confusing ways. Traditionally, pharmacogenetics has been defined as the study of the linkage between an individual's genotype and his or her ability to metabolize a foreign compound, while pharmacogenomics is a broader study of the identity of the genes or loci responsible for different individuals' different responses to a drug. Since the two

studies are growing closer together and their molecular basis – genetic variation – is the same, it seems redundant to use them both. In this chapter we will use pharmacogenomics as the umbrella term covering both genetic variations that affect drug metabolism and drug targeting.

3.1.1 Drugs and ethnicity

At the start of the discussion a social-political problem needs to be addressed.

Some opinions concerning the applications of human genetics to individual medicine have attempted to draw a clear distinction between genetic studies that identify ethnic groups who are at risk for specific diseases and genetic studies that identify individuals who may respond differently to the same drug used to treat their disease [29]. As will be described below, this ignores the fact that individual responses to drug treatments were first observed and reported on the basis of ethnicity [30]. The quality of recent quantitative studies [31] of human genetic variation indicates that a good description of the structure of human variation across populations and genomic regions could be available in the short term. However, the same studies emphasize that ethnic variations in genetic diversity are large and cannot be accounted for by the neutral model [32, 33]. The question is whether or not worldwide public opinion will allow these differences to be uncovered on a large scale and used as a basis for clinical treatments.

Attempts to dissociate ethnicity from drug development and treatment are understandable but unrealistic. Genetic testing has become highly controversial because of its social, economic and political implications. Altruistically, one would hope that this would not complicate the use of human genetics to develop healing treatments. Unfortunately, the political history of the last part of the twentieth century and continuing to the present tells us that all ethnically linked medicine becomes controversial and pharmacogenomics will be no exception.

3.1.2 Ethical considerations

Nor should the ethical problems associated with gathering data for pharmacogenomic purposes be ignored [34]. For example, the controversy resulting

from the proposed formation and uses of the Icelandic Health Sector Database has been reviewed [35–37] and similar databases have been proposed for the United Kingdom [38, 39]. Development and use of databases consisting of individuals' genetic data may seem logical to drug researchers and pharmaceutical firms [40], but may not seem so even to individuals who might ultimately benefit from the use of the data.

3.2 Metabolic pharmacogenomics

In humans, variations in clinical responses to drugs began to be observed in the 1950s. These included:

- inherited ethnic differences in hemolysis after average primaquine doses used in antimalarial therapy [30] where the differences are associated with differences in erythrocyte glucose-6-phosphate dehydrogenase activities [41];
- inherited differences in rates of acetylation of the drug isoniazid [42];
- the association of prolonged apnea that sometimes occurred after administration of the muscle relaxant, suxamethonium chloride (succinyl chloride), with levels of plasma pseudocholinesterase [43].

In retrospect, these results are not terribly surprising in light of data from the late 1930s showing that sometimes different strains of animals of the same species differ enormously in their responses to drugs [44].

At about the same time as human variations in drug responses were first observed – that is, at the dawning of the molecular genetic age – drug-metabolizing enzymes, present in all eukaryotic cells (but first identified in hepatocytes) began to be identified. Genetically derived variations in these enzymes' activities were soon suggested to be the cause of some ADRs [45].

Cytochromes P450 (members of the CYP superfamily) came to be identified as "detoxification" proteins and the study of P450 enzymes and their genetic variations became central to what was called "pharmacogenetics" [46–48].

From results of rapid advances in the technology of gene sequencing, it is now clear that SNPs exist in most, if not all, CYP isozymes, but that the majority of the polymorphisms exist in four CYPs and that these four are respon-

sible for about 40% of all drug metabolism mediated by CYP isozymes. We will return to this subject later.

3.3 What is the extent of human genomic heterozygosity?

As more genes have been sequenced, it has become apparent that genomic DNA polymorphisms in individuals are far more common than was believed even 10 years ago. In effect, this has blurred the distinction between mutations and polymorphisms. Because of the redundancy of the genetic code, many of these SNPs result in no primary sequence alterations in the proteins the genes encode. However, some SNPs (missense variants) do result in changes in protein primary sequences that may increase, decrease, abolish, or have no effect upon the protein's function. The structures of these proteins would have to be determined or predicted accurately for them to furnish potential drug targets. Even more recently, it has become clear that many SNPs occur in promoter regions where they affect expression levels of gene products.

We now summarize and compare the available data on heterozygosity estimates for the human genome.

3.3.1 Results from the human genome projects

Before actual sequencing results were available, widely quoted estimates of human genome-wide nucleotide diversity were about one SNP for every 1000 base-pairs [31, 49]. A standard estimate for the average size of a human protein is 300–400 amino acids. This corresponds to a total transcript length of approximately 1.2–1.5 kb and leads to a prediction of 1–2 SNPs per total exon content of an average human gene.

Using statistical methods, an estimation of nucleotide diversity from ascertained SNPs has been made from the private draft sequence of the human genome [50]. Several features were reported:

- Agreement with the previous estimates of nucleotide diversity of about 1 SNP per 1000 base-pairs.

- 75% of the SNPs found are intergenic.
- Considerable unexplainable variability of SNP density across the human genome exists and G + C content accounts for only a small part of the variation.
- Based on sequences of 10,239 known genes, missense SNPs are only about 0.12% of the total SNP count and of these, only 47% are radical protein changes (the meaning of radical changes is discussed below in Section 3.2.3).
- The results from the National Institutes of Health draft sequence [51] are in qualitative agreement with those from the private study, at least for estimates of nucleotide diversity. However, the authors don't make estimates of the frequency of potential protein structural changes.
- A total of about 35,000 protein-encoding genes in humans was predicted.

Some of these observations would be likely to have a significant impact on the use of drugs to treat disease.

- The variable SNP density makes it important to identify "hot-spots" in DNA encoding proteins involved in drug-responsiveness because the particular proteins will be the most likely to vary in a population.
- The low frequency of differences in protein activities due to SNPs seemingly presents the possibilities of favorable practical outcomes from pharmacogenomics developments as much more promising because of the decreased likelihood of multiple protein activity changes in drug response and metabolic pathways.
- Optimism about the promise of individual medicine is also based on reasoning that the smaller number of protein differences defining individual responses to drugs may also mean that targeting drugs to non-variable proteins could result in more uniform responses in populations.

In summary, drawing high-resolution conclusions on the basis of the draft sequences is currently risky [52]. Even the number of genes encoded by the human genome – one of the most widely publicized original conclusions – is in dispute [53] at the time of writing.

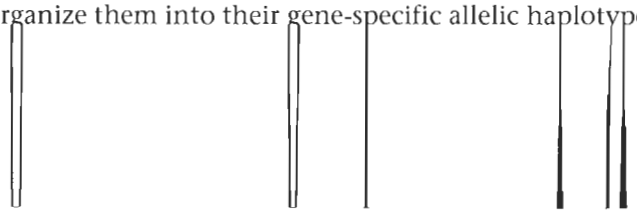
The primary reason for not weighting the results of the draft sequences of the whole genome too heavily when discussing SNPs is one that has not been widely commented on: both draft sequences are based on DNA obtained from

only a few anonymous subjects (< 10 in each case). In the private project, the selected donors were two males and three females, including one African-American, one Asian (Chinese), one "Hispanic-Mexican," and two Caucasians. For the NIH Project, the eight libraries used all derived from male DNA, ethnic mix not stated. Indeed, it is widely believed that, because of political considerations, the NIH is very reluctant to deal with the ethnic component of genomic diversity.

These facts are important because conclusions drawn concerning human genetic diversity on the basis of such small samples are statistically unreliable. Also, as discussed in the next section, the results from the Genome Projects do not agree well in some important details with the results of a recent but smaller-scale sequencing project that was specifically designed to investigate the impact of ethnicity on genomic diversity [54].

3.3.2 SNP evidence from direct sequencing

The result of a project of significance to future efforts in the pharmacogenomics area of drug development has been recently reported. This study [54] undertook a systematic discovery of gene-based sequence variation in 82 unrelated individuals, whose ancestors were from various geographical origins. The sample size and composition were sufficient to detect, with high certainty, globally distributed variants present at a frequency of at least 2% and population-specific variants present at a frequency of at least 5%. The population sample, using the definitions of the U.S. Census Bureau, included an approximately equal number of self-described Caucasians, African-Americans, Asians, and Hispanic-Latinos. The project's goal was to identify SNPs and to organize them into their gene-specific allelic haplotypes.



The exons (coding regions, 5'UTR and 3'UTR), up to 100 bases into the introns from the exon-intron boundaries (including the splice junctions), and the 5' upstream genomic regions were sequenced for 313 genes. These genes were chosen from ones for which complete genomic organization is publicly available. To assist in assessing the quality of the sequence information and to validate the construction of haplotypes, a three-generation Caucasian family and a two-generation African-American family were included. For evolutionary comparisons, the corresponding genomic regions from a chimpanzee were sequenced but will not be discussed here.

Table 1.
Nucleotide diversity estimates from the Human Genome Project draft sequences and a previous estimate.

Source	$\pi_{av} \times 10^4$	SNP/kb
[50]	8.98	1113
[51]	7.51	1331
[54]	5.8	1724
prev. estimate [31]	8	1250

Nucleotide sequence variation is conventionally estimated using a normalized measure of heterozygosity (π), representing the likelihood that a nucleotide position will be heterozygous when compared across two chromosomes selected randomly from a population. Table 1 gives the results from the three sequencing efforts discussed here [50, 51, 54] and a previous estimate [31].

For π, the results from the three recent sequencing studies agree reasonably well with previous estimates of about 1 SNP for every 1000 kb surveyed in two human chromosomes.

However, more detailed analysis of the results available from sequencing in the study of 313 human genes [54] reveals another scenario. This may be taken to be predictive of the results for similar analyses of the whole genomes of different individuals. Figures 6 A, B summarize some of these detailed results.

Figure 6 (A) shows that about 33% of the SNPs found were found in all four populations and that the number of population-specific SNPs for the African-American population was almost three times higher than the next highest figure (for Asians). Figure 6 (B) shows that the population distribution of haplotypes was similar to the population distribution of SNPs.

Other highly significant genetic conclusions from this study were:

- An average of about 12.7 SNPs and 14 different haplotypes per gene were observed.
- A significant fraction (48%) of population-specific haplotypes was seen only in the African-American sample.
- Many genes did not have one predominant haplotype and for 35% of the genes, no single haplotype had a frequency ≥ 50%. This indicates that the idea that there is one predominant "wild-type" form of a gene and various

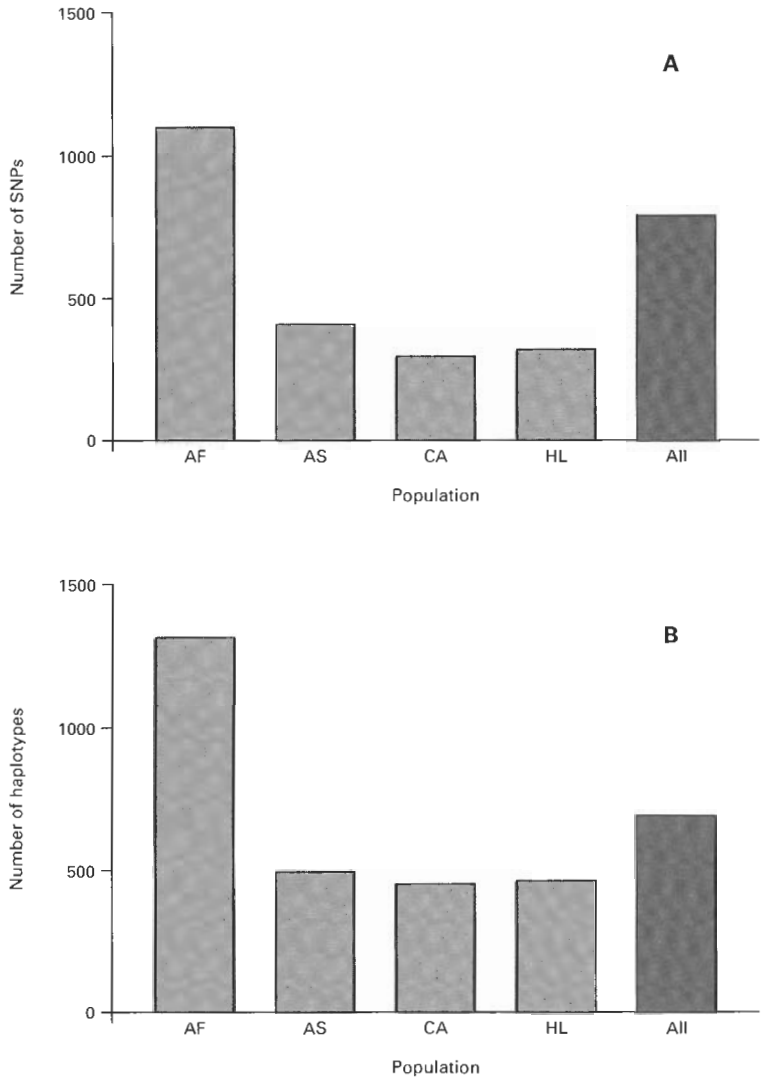


Fig. 6
(A) The distribution of total SNPs in 313 genes among population samples. The gray bars show the number of SNPs that are variable in each population, the black bar shows the number of SNPs that are variable in all four populations(B) The distribution of total haplotypes in 313 genes among population samples. The gray bars show the number of haplotypes that are variable in each population, the black bar shows the number of haplotypes variable in all four populations. Population codes are AF, African-American; AS, Asian; CA, Caucasian; HL, Hispanic-Latino; All, all four populations. Adapted from [54].

other rare or "mutant" forms is misleading. Instead, there are multiple haplotypes, each of which are observed in multiple populations and which account for a large fraction of human variability.
- Complex or unknown patterns of human migration and evolution complicate the distribution and interpretation of human genomic variation.
- The average number of polymorphic sites per kilobase of DNA found was 3.4 in the coding regions, 5.3 in the 5'UTR, 5.9 in the 5' upstream region, 6.5 in the exon-intron boundaries, and 7.0 in the 3'UTR, results that also differ from those found from the Human Genome Project data.
- The pattern of variation reflects the recent expansion of the human population.

Focusing on the impact these conclusions will have on pharmacogenomics:

- On average, there are many more SNPs and haplotypes per gene than previously predicted or estimated from the draft sequences of the human genome.
- There are large ethnic variations in genomic diversity.
- There is no "wild-type" human genome.
- The changes in population and distribution of individuals are changing the patterns of genetic diversity.
- Haplotypes generally supply more information on heterozygosity than individual SNPs.
- Of the polymorphic sites within the coding regions (about 30% of the total sites found), about 50% coded for an amino acid change of which 42% were essentially conservative, and 12% essentially radical.

3.3.3 Conservative and radical changes in polypeptides

Predictions of functional changes in proteins due to amino acid substitutions are currently more an art than a science. This situation stems from a lack of experimental structural data on substituted proteins to back up the predictions. In an attempt to determine which substitutions could be important, an integration of several chemical properties of residues – their chemical properties, polarities, and molecular volumes – has been made [55]. This integration comprises one of the few tools currently available to predict structural variations in residue-substituted proteins. Table 2 summarizes this effort: the

Table 2.
Chemical "distances" between amino acid pairs used in predicting conservative to radical changes in protein structures due to amino acid substitutions of one member of a pair by another. Adapted from [55].

Arg	Leu	Pro	Thr	Ala	Val	Gly	Ile	Phe	Tyr	Cys	His	Gln	Asn	Lys	Asp	Glu	Met	Trp	
110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177	Ser
	102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101	Arg
		98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61	Leu
			38	27	68	42	95	114	110	169	77	76	91	103	108	93	87	147	Pro
				58	69	59	89	103	92	149	47	42	65	78	85	65	81	128	Thr
					64	60	94	113	112	195	86	91	111	106	126	107	84	148	Ala
						109	29	50	55	192	84	96	133	97	152	121	21	88	Val
							135	153	147	159	98	87	80	127	94	98	127	184	Gly
								21	33	198	94	109	149	102	168	134	10	61	Ile
									22	205	100	116	158	102	177	140	28	40	Phe
										194	83	99	143	85	160	122	36	37	Tyr
											174	154	139	202	154	170	196	215	Cys
												24	68	32	81	40	87	115	His
													46	53	61	29	101	130	Gln
														94	23	42	142	174	Asn
															101	56	95	110	Lys
																45	160	181	Asp
																	126	152	Glu
																		67	Met

Table 3.
Categories of amino acid substitutions based on chemical distances, D, between pairs of residues. Adapted from [56].

Descriptor of structural effect	D value (Tab. 2)
Conservative	$0 < D \leq 50$
Moderately conservative	$51 < D \leq 100$
Moderately radical	$101 < D \leq 150$
Radical	$D > 151$

Table presents values of a parameter (D) for amino acid pairs and D is interpreted as a "chemical distance" between pairs. The D values have been further broken down somewhat arbitrarily into categories that express in words the dissimilarities between the chemical characteristics at each substitution site [56]. Table 3 gives these categories.

This categorization is, at best, only moderately convincing. For example, the SNP which converts hemoglobin A (HbA) to hemoglobin S (HbS, Gluβ6 → Val), resulting in a medically significant change in the properties of the blood containing large amounts of HbS (sickle cell hemoglobin) instead of HbA, would be classified as "moderately radical" according to Table 3. On the other hand, many HbA variants resulting from single amino acid substitutions are functionally unchanged, even though the substitution is in the same moderately radical category. The conclusion is that it is currently very difficult to predict functional changes in proteins from missense codon substitutions.

The result of applying these amino acid classifications to the proteins encoded by the 313 genes is shown in Table 4.

Clearly, there is a greater frequency of SNP-encoded amino acid substitutions that might lead to variant protein functions than were found from preliminary analyses of the draft sequences of the genome. This means that more quantitative structural and biochemical functional analyses than previously estimated would have to be performed before variant proteins could be regarded as drug targets.

3.3.4 Summary of preliminary SNP work

Problems associated with pharmacogenomics based purely on SNP data are just beginning to be realized. Much effort is being expended on building SNP

Table 4.
Coding region SNP consequences on amino acid substitutions.

Type of amino acid change	(%)
None	44.4
Conservative/moderately conservative	42.8
Moderately radical/radical	11.9
To stop codon	0.9

None means a silent nucleotide substitution. Change types categorized according to Grantham [55] values as: conservative/moderately conservative < 100, moderately radical/radical > 101. Adapted from [54].

databases and developing cost-effective genotyping strategies, while little regard is given to the matters of complexity, functionality, utility, and ethnic diversity of genetic associations revealed in patient populations.

By focusing on diversity in one gene – the one encoding the β_2-adrenergic receptor (β_2-AR) – instead of 313 different genes, a recent study [57] illustrates what efforts may be needed to uncover associations between genetic variants and drug response. Preceded by background on the clinical problem, we turn to a discussion of that study in the next section.

4 The molecular genetics of asthma

4.1 Asthma, the disease

In light of the discussion just presented, asthma serves as a good exemplar of the problems facing the application of pharmacogenomics to clinical medicine:

- Asthma affects 5 to 7 percent of the population of North America and may affect 150 million or more individuals worldwide [58, 59]. It is one of the most common and economically burdensome chronic diseases. However, a clear definition of the disease is lacking and no single symptom, physical finding or laboratory test serves to diagnose asthma.
- Only a few new therapeutic agents based on novel mechanisms of action have been developed over the last two decades. In addition, there is no evidence that morbidity and mortality are decreasing.

- Since asthma is often triggered by an allergic response, the environmental milieu of patient populations can play a significant role in expression of the disease.
- The varied clinical presentations of asthma have led to the concept that there may be multiple asthmatic phenotypes, which adds to the complexity of genetic studies.
- The varied clinical presentations of asthma have led to the belief that there may well be a multigenic component to the disease.
- Even though asthma has long been recognized as having a significant heritable component, genetic studies have been difficult to perform and the results have been difficult to interpret, although an increasingly large amount of conventional genetic data is now being accumulated.

4.2 Drug treatment for asthma

Ephedrine, obtainable from plants of the genus *Ephedra*, particularly the Chinese species *E. sinica*, has been used in China for more than 5,000 years to treat asthma and hay fever. This drug initiates its actions on cells by binding to the β_2-AR.

About 60 to 70 percent of patients diagnosed with asthma have mild disease for which a variety of synthetic agonists to the β_2-AR, introduced over 35 years ago and which revolutionized the treatment of asthma, suffice today. However, that leaves the treatment of millions of patients with more severe asthma subject to ongoing controversy about treatment with agonists which sometimes have very adverse effects [60].

The β_2-AR is a member of the G protein-coupled receptor superfamily that mediate the actions of catecholamines in many tissues. The β_2-AR expressed on bronchial smooth muscle acts to relax contracted muscle, resulting in bronchodilation, and beta agonists are the most effective acute treatment for reversal of bronchospasm in asthma. The bronchodilating response to beta agonists is known to exhibit significant inter-individual variation [58] and reports have been recently been published reporting on, and speculating on, the role of SNPs in responses of the β_2-AR to agonists used in asthma treatments [61, 62].

For future reference in section 6 of this chapter: much of the research on asthma has involved traditional animal models of asthma, inbred animal

models of hyperreactivity, and genetically altered mice. Most researchers in this area regard transgenic and gene-knockout mice as powerful tools for probing the mechanisms of disease. The strategy has been to use mouse models, when candidate genes for asthma are identified, in assessing the role of a given pathway in the pathogenesis of asthma.

As a major clinical problem and the object of intense past drug development efforts, asthma affords a good example of how molecular genetics might affect future efforts. A recent landmark paper [57] describes the haplotypes in the gene encoding the β_2-AR.

4.3 The genome and asthma

4.3.1 SNPs in the β_2-AR gene

The sequence variations in the gene were studied in a multiethnic population consisting of Caucasian, African-American, Asian, and Hispanic-Latino groups, and associations of the haplotypes found with bronchodilating responses to the agonist, albuterol, in asthmatics were studied [57].

The control sequencing data were gathered from an index repository of apparently normal individuals. This repository consisted of 23 Caucasians, 19 African-Americans, 20 Asians, and 15 Hispanic/Latinos. This number of individuals in each ethnic group provided for a >90% probability of detecting a SNP with an allele frequency between 0.05 (Caucasian group) and 0.08 (Hispanic-Latino group). To determine whether haplotypes of the β_2-AR gene were associated with the bronchodilatory response to the agonist albuterol, 121 Caucasian patients with asthma, enrolled from an outpatient facility, were used.

From sequencing, 13 SNPs organized into 12 haplotypes within a span of 1.6 kb were identified in the β_2-AR gene. Taken at face value, this is a promising result with respect to drug targeting: if SNPs were distributed randomly in individuals at a frequency of one SNP per kb of DNA, the number of different SNP combinations over a haplotype could be large – $2^{\#SNPs}$ – which in this case corresponds to $2^{13} = 8{,}192$ possible combinations. The actual number of haplotypes comprise only a fraction of this – a result that seems to be the general pattern for genes that have been investigated so far. This has been used as evidence to predict that haplotype blocks may be at least 10 times longer than

Table 5.
Localization of SNPs and identification of haplotypes of the β_2-AR. Adapted from [57].

SNP Position	–1023	–709	–654	–468	–406	–367	–47	–20	46	79	252	491	523
Alleles:	G/A	C/A	G/A	C/G	C/T	T/C	T/C	T/C	G/A	C/G	G/A	C/T	C/A
Haplotype													
1	A	C	G	C	C	T	T	T	A	C	G	C	C
2	A	C	G	G	C	C	C	C	G	G	G	C	C
3	G	A	A	C	C	T	T	T	A	C	G	C	C
4	G	C	A	C	C	T	T	T	A	C	G	C	C
5	G	C	A	C	C	T	T	T	G	C	G	C	C
6	G	C	G	C	C	T	T	T	G	C	A	C	A
7	G	C	G	C	C	T	T	T	G	C	A	T	A
8	G	C	A	C	C	T	T	T	A	C	A	C	C
9	A	C	G	C	T	T	T	T	A	C	G	C	C
10	G	C	G	C	C	T	T	T	G	C	A	C	C
11	G	C	G	C	C	T	T	T	G	C	G	C	C
12	A	C	G	G	C	T	T	T	A	C	G	C	C
Location:	5′	5′	5′	5′	5′	5′	*	5′	**	***	silent	****	silent

Nucleotide numbers are relative to the start codon at position + 1. Allele refers to the two nucleotide possibilities at each SNP position; 5′ means 5′ upstream of the β_2-AR open reading frame; *Cys/Arg substitution; **Gly/Arg substitution; ***Gln/Glu substitution; ****Thr/Ile substitution.

had been previously predicted, so that sequences as long as 50,000 bases may contain small numbers of haplotype blocks that would account for variations in these genes in 80–90% of the population [63]. The results of the study under discussion here are contrary to this expectation. For the β_2-AR gene, and by extension to other genes, the actual case is much more complicated.

Tables 5 and 6 show the haplotype blocks found and their frequencies. Examination of Table 5 shows that no individual SNP serves to adequately predict a haplotype. Thus, it would be hazardous to select an individual SNP as a marker for a haplotype in this case. That conclusion may hold true in general. This would enormously complicate identification of haplotypes.

4.3.2 SNPs in the β_2-AR gene; clinical correlates

Using a cohort of 121 Caucasians diagnosed with asthma, the work under discussion was extended to an examination of the predictive value of the SNPs

Table 6.
Five most common β_2-AR haplotype pairs found in an asthmatic cohort of 121 Caucasians. All other less common pairs total 12.2%. Adapted from [57].

Haplotype pair	%
2/4	30.6
2/2	20.7
2/6	18.2
4/4	11.6
4/6	6.6

and haplotypes in response to albuterol. No differences were found in the frequencies of the haplotypes between index and this asthmatic population. The five most common haplotype pairs in the asthmatic cohort and their frequencies are shown in Table 6.

From their sequencing work on the β_2-AR gene in the control group, the authors conclude:

- They identified 13 SNPs in a contiguous region of the 5′ upstream and coding sequence regions of the gene.
- 12 distinct haplotypes are represented in their population consisting of four major United States ethnic groups.
- A striking divergence in ethnic distribution exists for several haplotypes.

Their results are therefore in excellent agreement with the number of SNPs and haplotypes per gene observed in the study of 313 genes [54].

The results from albuterol treatment were:

- Five haplotype pairs were common in Caucasian asthmatics.
- Specific pairs of haplotypes were required to show maximum predictive value in treatments with albuterol (Fig. 7) and there are clear differences in the *in vivo* response to the drug based on these pairs.
- Individual haplotypes showed intermediate predictive value.
- None of the individual SNPs had any predictive value for response to the drug.

Examination of Figure 7 with reference to Table 5 indicates that differences in response to albuterol for the three most common haplotype pairs (present

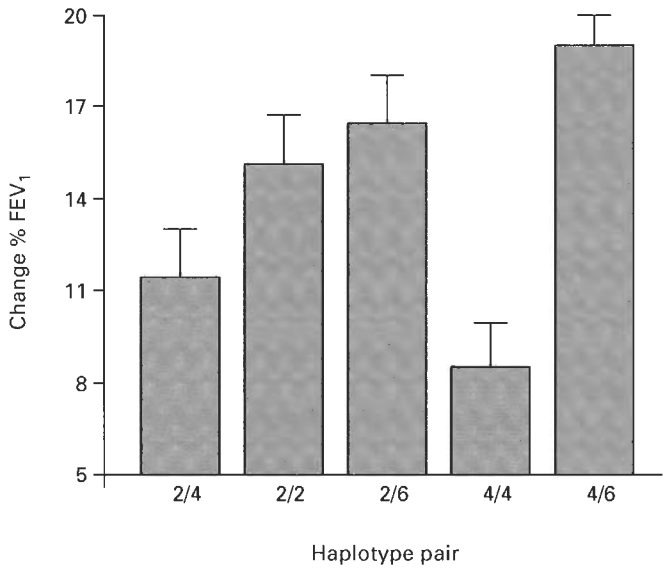


Fig. 7
In vivo responses to albuterol distinguished by haplotype pairs. Data from 121 Caucasian subjects diagnosed with asthma. FEV_1 = forced expiratory volume in 1 second. Adapted from [57].

in 69.5% of the patient population) varied minimally. The maximum differences in response for the most common haplotype pairs was a 50% better response for patients with the 2/4 pair than with the 4/4 pair and about a 120% better response for patients with the 4/6 pair than with the 4/4 pair. Using a cultured human embryonic kidney cell line that expresses β_2-AR, it was then shown that haplotype 2 significantly increased expression of β_2-AR protein and β_2-AR mRNA compared to haplotype 4, results consistent with the *in vivo* results.

4.3.3 Conclusions from this study

Although it was not discussed in these terms, the authors present an overall summary that is in agreement with the epigenetic viewpoint: their results indicate that the unique interactions of multiple SNPs within a haplotype ultimately affect biological and therapeutic phenotypes and that

individual SNPs may have poor predictive value for patient responsiveness to drugs.

Their results also highlight some other clinical problems in applying SNP data to drug development and treatment. With respect to treatments:

- The difference in efficacy between the most and least responsive pairs of haplotypes to treatment with albuterol was found to be significant but comprises only about 20% of the asthmatic Caucasian population used in this study. This leaves 80% of patients whose responses range between poor and good; what are to be recommended treatments for these patients?
- The very large differences in haplotype frequencies (>20-fold) between the important United States ethnic groups described here have been found in all similar investigations so far. This study did not look at whether or not identical haplotypes from different ethnic groups exhibit the same responses to albuterol *in vivo* or whether or not the same correlations with *in vitro* studies exist. If they don't, it is possible that other factors that interact with the β_2-AR haplotypes may exist.
- The results send no clear signals as to how individual drug development might proceed. In fact, the data show that although its impact may be limited, haplotyping may be useful mainly for predicting individual responses to currently used drugs in individuals.
- Clearly, the results from β_2-AR haplotype analysis would have to be used in conjunction with haplotype analysis of drug-metabolizing enzymes before it could apply to a clinical scenario. As has been remarked, "The mantra of pharmacogenomics has been 'one drug may not fit all'. It might be the case that one pharmacogenomic test might not fit all" [64].

Obviously, major conclusions cannot be made with certainty on the basis of one study, landmark that it may be. However, because of its basic agreement with the later study on 313 genes [54], it seems likely that outcomes from studies of SNP organization in many other genes will be similar to the β_2-AR gene results. Recent studies on sequence-based haplotype analysis of the interleukin-11 (IL-11), tissue-type plasminogen activator (t-PA), and angiotensin genes [65–67] certainly show that individual SNPs have little predictive value in clinical correlations.

We now turn to a case of a gene of major importance in health and disease – the gene encoding the glucocorticoid receptor – where haplotype maps

have not yet been reported in the open literature, but where consideration of non-genomic pathways must be a major consideration when such maps become available.

5 Genomic and non-genomic effects mediated by the glucocorticoid receptor

5.1 Genomic effects

Glucocorticoids exert a strong influence on human metabolism and especially on mesenchymally derived tissues such as the hematopoietic system, the lymphatic system, and the thymus. In lymphoblastic leukemia cells, glucocorticoid treatment leads to cell cycle arrest and subsequent apoptosis. The molecular sequence of events leading to these phenotypic changes is unknown, but they depend on a functional glucocorticoid receptor (GCR).

Glucocorticoids are also involved in a variety of human functions including behavior, cardiovascular status, regulation of blood pressure, inflammation and metabolism. A common medical dilemma is the identification of those individuals with a strong family history of high blood pressure and associated diseases, such as diabetes, heart disease and stroke, that require very close monitoring and early drug treatment or lifestyle changes.

Specific for glucocorticoid hormones, the GCR was the first steroid hormone receptor to be highly purified and shown to be a DNA-binding protein.

Gene transcription is affected by these hormones in several steps as follows: (1) glucocorticoids bind to GCR complexes located in the cytoplasm; (2) the GCR complexes release associated (mainly heat-shock) proteins upon binding the ligand; (3) the GCR-ligand complex translocates to the nucleus and binds to specific recognition sites; (4) gene transcription is enhanced or inhibited depending on the type of the GCR and the specific genomic site.

The characteristic nucleotide sequences of the DNA sites, called response elements, that bind several major hormone receptors have been determined. The sequences of the consensus response elements for the glucocorticoid and estrogen receptors are 6-bp inverted repeats separated by any 3 base-pairs.

Essential hypertension is a complex multifactorial disorder in which genetic predisposition interacts with environment to yield the ultimate phenotype. In turn, hypertension acts as a risk factor for the development of ischaemic heart disease and stroke. It is now accepted that several loci contribute to the genetic component of hypertension. Although the field is complicated by numerous unconfirmed or contradictory findings, the evidence that genes are involved in renal sodium handling or aspects of corticosteroid production and action is remarkably consistent.

As an example of the commercial interest in studying the GCR gene for commercial development, Gemini Genomics has been issued a patent (U.S. Pat. No. 6,156,510) for its method of diagnosing a predisposition to hypertension or sustained high blood pressure. The hope is that diagnostic tests based on this patent will enable the early detection of those at risk, allowing physicians to recommend treatments that will minimize the consequences of high blood pressure disease, such as stroke, kidney disease, and cardiovascular disease. The patent covers proprietary methods of detecting mutations in a GCR gene.

Gemini's patent is based on their study which revealed an apparent association between certain SNPs in the glucocorticoid receptor gene and hypertension. The study compared the glucocorticoid receptor genes of individuals who had high blood pressure, who were taking anti-hypertensive medication, and who had two parents with hypertension, with those of individuals whose blood pressure was normal, who were not taking anti-hypertensive medication, and who had two parents with normal blood pressures.

While the diagnostic uses of SNPs in GCR-encoding genes may be important, the existence of important non-genomic effects of glucocorticoids will impede pharmacogenomic studies of these hormones.

5.2 Non-genomic effects

The presence of non-genomic effects of glucocorticoids can be experimentally tested:

- The genomic actions take place on a long time-scale compared to non-genomic effects. The fastest genomic effects take place on the order of min-

utes, but many actions take place on a time-scale of days. Rapid non-genomic effects can be determined by washout experiments.
- The genomic actions are mediated by the GCR. Non-genomic effects may be determined by blockading the receptor with antagonists.
- The genomic actions result in protein synthesis. Non-genomic effects may be determined by inhibitors of protein synthesis.

The slowness of genomic effects from these hormones may be partly characteristic of glucocorticoids because they require translocation of the ligand-bound receptors to the nucleus (requiring 10–30 min). However, in general all genomic effects require time for transcription and protein synthesis to take place. Drug effects that have very rapid actions should always be suspected of having non-genomic interactions.

Although these transcriptional effects of glucocorticoids are well known, there is growing evidence suggesting that hormonal steroids use multiple transduction pathways, either genomic or non-genomic or both. In particular, glucocorticoids affect neural function also *via* mechanisms that do not involve a genomic component. Non-genomic mechanisms appear to be activated by a series of other steroids as well (progesterone, estrogens, testosterone, aldosterone, vitamin D3, and neurosteroids) [68].

Evidence has accumulated suggesting that non-genomic mechanisms activated by glucocorticoids involve various brain areas, neurotransmitter, and second messenger systems as well as behaviors. The non-genomic effects of glucocorticoids may be one of the major keys to understanding the role of these hormones in controlling brain function.

Nevertheless, non-genomic effects of glucocorticoids are still poorly understood. Moreover, discriminating between genomic and non-genomic mechanisms is difficult in some cases.

From animal model and human studies, glucocorticoids have been implicated in non-genomic early stress responses. It appears that specific changes induced by glucocorticoids in early phases of the stress reaction are especially important. The meaning and consequences of stress may be different in early and late phases. While the early response often needs an active response, chronic stressors lead to behavioral depression. Non-genomic mechanisms appear to be crucial for the active phase of the stress response.

The targets of glucocorticoid hormones and pharmacological agents present in the central nervous system overlap. The GABAergic, cholinergic, sero-

tonergic and other systems are important targets of drug action, and all these systems appear to be influenced by glucocorticoids *via* both genomic and non-genomic mechanisms. Disparate data suggest that stress in general and glucocorticoids in particular affect both the desired and adverse effects of pharmacological agents. For example:

- The anxiolytic efficacy of benzodiazepines is affected by experiential background .
- The effects of serotonergic drugs were affected by the intrinsic HPA-axis-stimulating effects of some drugs, housing and testing conditions, handling history, and stress.
- Amphetamine increased locomotor activity of subordinate squirrel monkeys, but caused hypoactivity in dominant group members. Desipramine has antidepressant-like effects in the Porsolt test in rats previously exposed to forced swimming only. Uncontrollable, but not controllable, stress was shown to potentiate morphine's rewarding properties. The behavioral effects of kainate and bicuculline applied to the hypothalamic attack area (subfornical intermediate hypothalamus) depend on the previous aggressive experience of subjects. Some human observations indicate that the interaction between stress and the efficacy of drugs is clinically meaningful.

All the above manipulations affect acute stress responsiveness. The effect of isolation and handling history on pharmacological efficacy might be especially relevant, because they do not change background levels of glucocorticoids but change the reactivity of the HPA-axis. Therefore, it has been hypothesized [68] that non-genomic glucocorticoid effects may have contributed to the interaction between stress and the efficacy of drugs. This assumption is supported by the fact that some effects of glucocorticoids on drug efficacy were expressed rather quickly. The anxiolytic effect of buspirone in the social interaction test was abolished by corticosterone applied 10 min before testing. In addition, acute glucocorticoid treatments decrease, while a chronic elevation in plasma glucocorticoids increases anxiety, i.e., acute and long-term effects of glucocorticoids appear to be opposite.

The non-genomic mechanisms of glucocorticoids interact with genomic mechanisms. They (1) prepare the field for genomic mechanisms, (2) contribute to the early occurrence of effects supported later by genomic mecha-

nisms, and (3) induce early adaptive changes that are opposite to the delayed genomic effects. Such opposing ways of interaction can be reconciled only by assuming that the nature of the interaction depends on the nature of the challenge and/or on the brain sites involved. It is noteworthy that opposing ways of interaction are not always contradictory. One can hypothesize that preparing the field for delayed responses may be useful when the early response proved inefficient (i.e., there is a need for activating additional mechanisms).

Clearly, the development of specific agonists and antagonists for non-genomic glucocorticoid actions – drugs unavailable now – would be of major clinical importance. Some evidence suggests that some unknown molecular properties may make steroids effective or non-effective in triggering non-genomic effects: for example, triggering could be initiated by glucocorticoids specifically binding to proteins other than the GCR [68]. However, whatever the mechanisms are for the non-genomic actions of these hormones, the actions themselves will have to be taken into account in developing drugs that alter the genomic actions.

In conclusion, the actions of glucocorticoid hormones, once believed to be a classic case of a small signalling molecule acting directly on the genome, are now known to have a non-genomic component. These multiple modes of interaction are in accord with the epigenetic viewpoint. Since some genomic effects of the hormones counter the non-genomic effects [68], especially in the central nervous system, taking both into account in treatments involving these hormones or their antagonists is needed.

6 Animal models in the genomic era of drug research

The sequencing of the human genome and growing awareness of human genomic diversity has brought another problem into the drug development process: the differences in diversity between the genomes of inbred animal strains and humans. In this chapter emphasis has been placed on the need to consider ethnic diversity and epigenetic mechanisms to understand genomic sequence data in humans. We now turn to how human genomic data may affect drug research involving animal models.

Preclinical drug development relies heavily on animal models – typically inbred strains of mice and rats – for drug-target validation. This strategy

increases data reproducibility due to the small genetic heterogeneity of the animals. However, this advantage is often offset by the failure of later clinical trials on humans, which must be partly due to the failure of the inbred animals to adequately model human genetic diversity [69]. Furthermore, as discussed earlier in this chapter, results already obtained for the human genome reveal that there is no such thing as a human wild-type genome. It would be foolish to expect that this won't be true for other animal genomes as well.

It is not known, nor is it presently feasible to determine, what fraction of human and mouse or rat genomes contain similar haplotypic patterns [69]. Thus, drug efficacy trials using a disease model in a single inbred mouse strain are analogous to a similar trial on a small inbred population of humans – such trials throw away diversity information now known to be important in connecting haplotype to phenotype. In general, the knockout and transgenic models suffer the same problem of not accurately reflecting the human genome because these strains are typically produced from an inbred background.

For example, the paternally expressed gene 3 (PEG3 gene), common to humans and mice, plays an important role in mouse maternal behavior. In the knockout model when the PEG3 gene is removed from the mouse genome, mothers ignore their offspring and let them die. This must be associated with the fact that in mice, the gene product is expressed primarily in the brain while, in humans, expression is mainly in the ovaries and placenta. The same gene is clearly acting in different ways in the two organisms.

In September 1999, the NIH in the United States announced that it was setting up a Mouse Genome Sequencing *Network* (as distinct from a well-financed Project), which aims to sequence the mouse genome in draft form by the year 2003. This network includes several research groups who are working at sequencing a variety of mouse chromosomes. In 2000, The Celera Corporation announced that they had begun a private mouse genome project of their own based on the methods they used in their human genome sequencing project [50]. Both these projects will presumably be based on DNA obtained from one, or a few, inbred strains.

An annotated mouse cDNA project [70] that is based on sequencing every transcript encoded by the C57BL/6J inbred strain of mice, is the project furthest along. However, it suffers not only from being confined to sequences from an inbred strain, but also, because of the use of cDNA, from exclusion of promoter and intron sequence data.

So far as characterizing mouse SNPs and haplotypes are concerned, nothing like the human SNP Consortium exists and the discovery and genotyping of mouse SNPs lags far behind the human project [71]. Thus, application of human SNP and haplotypic data to mouse models will remain uncertain for many years. Suggestions for interim solutions to this problem include:

- If the relevant mouse and human genes are found to exhibit comparable variation, the new drug target should be studied, preferably, in parallel in several mice strains that could partially resemble the genetic polymorphism of humans in analogy to studying drug efficacy in parallel in patients from different ethnic origins.
- Transgenic mice expressing the spectrum of the human allele repertoire could be prepared.
- Increase the use in drug trials of mouse strains showing higher levels of natural polymorphism compared with classical inbred strains, such as the PWD/Ph and PWK/Ph strains that are more closely related to wild mice.

7 Conclusions and future directions

7.1 Metabolic pharmacogenomics

It seems clear that this branch of pharmacogenomics will continue to develop at a rapid pace. Its impact is already being felt in isolated cases. For example, about 2,400 children are diagnosed with acute lymphoblastic leukemia (ALL) each year in the United States. Treatment of ALL usually involves a combination of chemotherapeutic drugs including 6-mercaptopurine. Children presenting with an inherited deficiency in thiopurine methyltransferase (TMPT, the drug that metabolizes 6-mercaptopurine and other thiopurine drugs), need to be treated with lower concentrations of the drug. About one in 300 children are homozygotes in the deficient enzyme, and one in ten are heterozygotes. These patients are at risk with standard doses of 6-mercaptopurine. A DNA-based test has been developed to identify them so that they may be treated with lower doses of the drug [72].

Technological advances in testing for polymorphisms have led to a test for all the known 18 polymorphisms in the cytochrome p450 2D6 (CYP2D6)

enzyme. Polymorphisms in CYP2D6, resulting in deficiencies in enzymatic activity, occur in about 7–10 percent of the population with wide ethnic variations. The enzyme is responsible for metabolizing about 25% of all drugs, including many of the top 100 best-selling ones, including tricyclic antidepressants and the antidepressant Prozac. Research versions of the test currently cost in the neighborhood of $100 in the United States and prices will surely drop [72]. This is certainly inexpensive for a test that predicts individual responses to such a wide variety of drugs.

Because the effects of deficiencies in metabolizing drugs are relatively easy for patients and physicians to understand, because it will have positive impact on both physicians' and patients' insurance premiums, and because data from an individual's test do not have to be entered into a large database, it seems likely that testing individuals for cytochrome p450 and some other metabolizing enzyme polymorphism patterns will shortly become standard.

7.2 Personalized drugs

As more data on the human genome and its roles in human life are analyzed, the epigenetic viewpoint of human biology and medicine appears to be growing stronger. Even in its present unfinished form, the human genome sequence, accompanied by increasing data from a variety of sources, is showing genetic interactions that were previously unknown. For example, in all of the cases below, polymorphisms in a given gene will produce protein targets potentially having multiple functional roles:

- The one-gene/one-polypeptide concept is dead. Gene interactions are now recognized as more important than gene numbers. Many discontinuous genes can encode more than one protein by splicing together alternate exon combinations [73]. On the basis of a draft sequence of the human genome, it has been estimated that at least 10,000 human genes take part in this process [51].
- Introns sometimes take part in protein production. For example, the gene for prostate-specific antigen (PSA) has five exons and four introns but encodes a second protein (called PSA-linked molecule) which is partly encoded by the fourth intron of the PSA gene [74].

- Although the frequency of occurrence is not yet known, some transcripts are translated into precursor proteins which are then differentially cleaved into proteins of different functions [75, 76].
- Trans-splicing takes place (although with as yet unknown frequency in humans) and can be vital to a eukaryotic organism. For example, a gene in *Drosophila* that controls chromatin structure in early development, transcribes RNA from four exons on one strand and two exons in the opposite orientation on the opposing strand in a process that eventually yields four proteins with different functions [75, 76].
- The frequency of overlapping genes is also not currently known in humans although, since many eukaryotes use this mechanism for producing different proteins, it is difficult to believe that this mechanism is not employed in humans.
- The concept that data currently being consolidated by the SNP consortium will inevitably lead to new drug developments has not been borne out by data reviewed in this chapter. The concept may need to be folded into one that might be called the "Haplotype consortium."

In conclusion, much more evidence needs to be gathered before it can be stated unequivocally that personalized drug treatments are even theoretically possible. What is presently known about haplotype variations lends little confidence to the idea that SNP data will clearly lead the way to new drug development.

There are no adequate structural data available on what variant proteins produced by particular haplotypes might form attractive drug targets. It is surprising that more effort is not being put into systematic, quantitative studies of structural modifications to proteins due to missense substitutions. Structural data are going to be at least as important as raw SNP data in determining if there are drug targets with any advantages.

7.3 Animal models

Pharmaceutical development now depends, and will depend as far as can be seen into the future, on use of animal models. However, firm genomic data are rapidly emerging, showing how inadequate are the inbred, knockout, and transgenic animal models for developing efficacious drugs for humans. It is

therefore somewhat surprising that a higher priority in the form of international collaborations has not been placed on completing the mouse genome project. Furthermore, as the epigenetic viewpoint of human biochemistry and physiology takes hold, questions immediately arise as to whether the same interaction networks that occur in humans are the same as those occurring in animals.

7.4 Ethical considerations

As we have seen, as more data on the human genome is brought forth, epigenetic thinking and considerations have become more important in foreseeing future drug development and treatments. We know that single genes do not determine most of the effects of medications and not all responses have inherited roots. We know that some drug responses are caused by environmental factors such as age or diet, and teasing out this nature-*versus*-nurture difference may prove difficult. We know from actual cases that amassing databases of human genetic profiles creates enormous ethical and privacy problems. Among the basic questions in dealing with these problems are: do the benefits outweigh the complications? If they do, can the public be convinced of this?

Not the least part of the problem would be the development of adequate quality assurance of individual sequence data. The general case of quality control in molecular genetic testing has been recently reviewed [77]. Some features of genetic testing are worth mentioning here: (1) unlike other forms of laboratory tests, genetic composition of an individual remains unaltered through life so the results of a single test can stay with them for life, including false positives and false negatives, (2) because of the present perception that genetic predisposition is an all-or-none proposition, genetic test results can have profound psychological implications for individuals and their supporting family members, (3) if genetic test databases are ever to enter mainstream medical practice, there must be absolute public confidence in the accuracy of the results.

The worldwide status of the art is nowhere near achieving a high enough level of confidence in genotyping, much less haplotyping. For example, Figure 8 shows the status in Europe for well-established genotyping for cystic fibrosis. The apparent asymptotic approach to 90% correctness as the number of laboratories grows larger is not impressive.

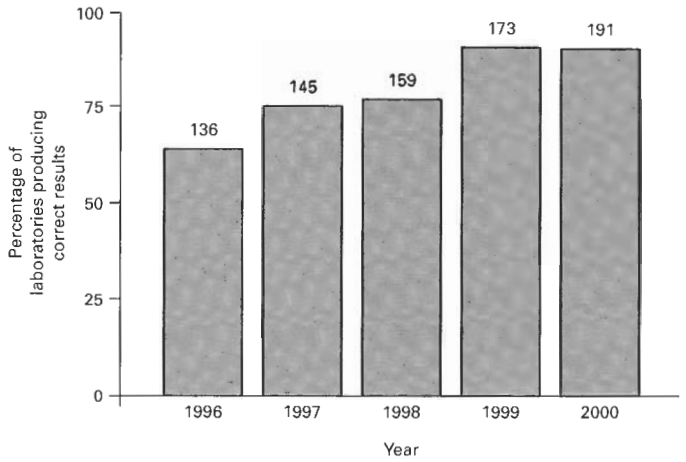


Fig. 8
Errors in European laboratories during the five years of European external quality assessment schemes for cystic fibrosis. Numbers across the tope indicate the total number of participating laboratories. Adapted from [77].

References

1 C.R. Cantor: Mol. Diagn. *4*, 287 (1999).
2 A.D. Roses: Novartis Found. Symp. *229*, 63 (2000).
3 M.M. Shi, M.R. Bleavins and F.A. de la Iglesia: Mol. Diagn. *4*, 343 (1999).
4 S. Jones: Genetics in Medicine: Real Promises, Unreal Expectations. Milbank Memorial Fund, New York 2000.
5 J. Lazarou, B.H. Pomeranz and P.N. Corey: JAMA *279*, 1200 (1998).
6 M. Kvasz, I.E. Allen, M.J. Gordon, E.Y. Ro, R. Estok, I. Olkin and S.D. Ross: Med. Gen. Med. E3 (2000).
7 T.R. Einarson: Ann. Pharmacother. *27*, 832 (1993).
8 D.W. Bates, D.J. Cullen, N. Laird, L.A. Petersen, S.D. Small, D. Servi, G. Laffel, B.J. Sweitzer, B.F. Shea, R. Hallisey et al.: JAMA *274*, 29 (1995).
9 D.W. Bates, N. Spell, D.J. Cullen, E. Burdick, N. Laird, L.A. Petersen, S.D. Small, B.J. Sweitzer and L.L. Leape: JAMA *277*, 307 (1997).
10 I.R. Edwards and J.K. Aronson: Lancet *356*, 1255 (2000).
11 I.R. Edwards: J. R. Coll. Physicians Lond. *34*, 48 (2000).
12 D.B. Jefferys, D. Leakey, J.A. Lewis, S. Payne and M.D. Rawlins: Br. J. Clin. Pharmacol. *45*, 151 (1998).
13 J.K. Borchardt: Mod. Drug Disc. *4*, 35 (2001).
14 J.A. DiMasi: Drug Inf. J. *34*, 1169 (2000).
15 K.I. Kaitin and J.A. DiMasi: Drug Inf. J. *34*, 673 (2000).

16 S. Aldridge: Gen. Eng. News *21*, 29 (2001).
17 F.S. Collins: N. Engl. J. Med. *341*, 28 (1999).
18 R.C. Strohman: Perspect. Biol. Med. *37*, 112 (1993).
19 R.C. Strohman: Toward an Epigenetic Biology & Medicine, http://www.mentalhelp.net/perspectives/articles/art05963.htm.
20 J. Beisson and T.M. Sonneborn: Proc. Natl. Acad. Sci. USA *53*, 275 (1965).
21 S.F. Ng and R.J. Williams: J. Protozool. *24*, 257 (1977).
22 A.P. Wolffe and M.A. Matzke: Science *286*, 481 (1999).
23 S. Wright: Phys. Rev. *21*, 487 (1941).
24 E.J. Sijbrands, R.G. Westendorp, J.C. Defesche, P.H. de Meier, A.H. Smelt and J.J. Kastelein: BMJ *322*, 1019 (2001).
25 R.R. Williams, S.J. Hasstedt, D.E. Wilson, K.O. Ash, F.F. Yanowitz, G.E. Reiber and H. Kuida: JAMA *255*, 219 (1986).
26 A.F. Vuorio, H. Turtola, K.M. Piilahti, P. Repo, T. Kanninen and K. Kontula: Arterioscler. Thromb. Vasc. Biol. *17*, 3127 (1997).
27 T. McKeown: The Role of Medicine: Dream, Mirage or Nemesis? Princeton University Press, Princeton 1979.
28 T. McKeown: The Origins of Human Disease. Basil Blackwell, Inc., New York 1988.
29 A.D. Roses: Lancet *355*, 1358 (2000).
30 R.S. Hockwald, J. Arnold, C.B. Clayman and A.S. Alving: JAMA *149*, 1568 (1952).
31 M. Przeworski, R.R. Hudson and A. Di Rienzo: Trends Genet. *16*, 296 (2000).
32 M. Kimura: Proc. Natl. Acad. Sci. USA *63*, 1181 (1969).
33 M. Kimura: Nature *217*, 624 (1968).
34 A.M. Issa: Trends Pharmacol. Sci. *21*, 247 (2000).
35 H. Rose: The Commodification of Bioinformation: The Icelandic Health Sector Database, Wellcome Trust, London 2000.
36 G.J. Annas: N. Engl. J. Med. *342*, 1830 (2000).
37 D. Winickoff: N. Engl. J. Med. *343*, 1734; discussion 1735 (2000).
38 R. Fears and G. Poste: Science *284*, 267 (1999).
39 R. Fears, D. Roberts and G. Poste: BMJ *320*, 933 (2000).
40 K.Y. Kreeger: The Scientist *15*, 32 (2001).
41 P.E. Carson, C.L. Flanagan, C.E. Ickes and S. Alving: Science *124*, 484 (1956).
42 H.B. Hughes, J.P. Biehl, A.P. Jones and L.H. Schmidt: Am. Rev. Tuberc. *70*, 266 (1954).
43 H. Lehmann and E. Ryan: Lancet *211*, 124 (1956).
44 F. Bernheim and M.L.C. Bernheim: J. Pharm. Exp. Ther. *64*, 209 (1938).
45 A.G. Motulsky: JAMA *165*, 835 (1957).
46 E.S. Vesell: Pharmacol. *61*, 118 (2000).
47 E.S. Vesell: J. Clin. Pharmacol. *40*, 930 (2000).
48 D.W. Nebert: Clin. Genet. *56*, 247 (1999).
49 A. Chakravarti: Nat. Genet. *19*, 216 (1998).
50 J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt et al.: Science *291*, 1304 (2001).
51 E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh et al.: Nature *409*, 860 (2001).
52 N. Katsanis, K.C. Worley and J.R. Lupski: Nature Genetics *29*, 88 (2001).
53 J.B. Hogenesch, K.A. Ching, S. Batalov, A.I. Su, J.R. Walker, Y. Zhou, S.A. Kay, P.G. Schultz and M.P. Cooke: Cell *106*, 413 (2001).

54 J.C. Stephens, J.A. Schneider, D.A. Tanguay, J. Choi, T. Acharya, S.E. Stanley, R. Jiang, C.J. Messer, A. Chew, J.-H. Han et al.: Science *293*, 489 (2001).

55 R. Grantham: Science *185*, 862 (1974).

56 W.H. Li, C.I. Wu and C.C. Luo: J. Mol. Evol. *21*, 58 (1984).

57 C.M. Drysdale, D.W. McGraw, C.B. Stack, J.C. Stephens, R.S. Judson, K. Nandabalan, K. Arnold, G. Ruano and S.B. Liggett: Proc. Natl. Acad. Sci. USA *97*, 10483 (2000).

58 J.M. Drazen, E. Israel, H.A. Boushey, V.M. Chinchilli, J.V. Fahy, J.E. Fish, S.C. Lazarus, R.F. Lemanske, R.J. Martin, S.P. Peters et al.: N. Engl. J. Med. *335*, 841 (1996).

59 S.B. Liggett and D.A. Meyers (eds): The Genetics of Asthma. Marcel Dekker, Inc., New York 1996.

60 P.M. O'Byrne and H.A. Kerstjens: N. Engl. J. Med. *335*, 886 (1996).

61 S.B. Liggett: Am. J. Respir. Crit. Care Med. *156*, S156 (1997).

62 E. Israel, J.M. Drazen, S.B. Liggett, H.A. Boushey, R.M. Cherniack, V.M. Chinchilli, D.M. Cooper, J.V. Fahy, J.E. Fish, J.G. Ford et al.: Int. Arch. Allergy Immunol. *124*, 183 (2001).

63 L. Helmuth: Science *293*, 583 (2001).

64 J. McCarthy: Trends Pharma. Sci. *21*, 461 (2000).

65 Y. Shinohara, Y. Ezura, H. Iwasaki, I. Nakazawa, R. Ishida, M. Kodaira, M. Kajita, T. Shiba and M. Emi: J. Hum. Genet. *46*, 494 (2001).

66 I. Nakazawa, T. Nakajima, T. Ishigami, S. Umemura and M. Emi: J. Hum. Genet. *46*, 367 (2001).

67 N. Sato, T. Katsuya, T. Nakagawa, K. Ishikawa, Y. Fu, T. Asai, M. Fukuda, F. Suzuki, Y. Nakamura, J. Higaki et al.: Life Sci. *68*, 259 (2000).

68 G.B. Makara and J. Haller: Prog. Neurobiol. *65*, 367 (2001).

69 D. Gurwitz and A. Weizman: Drug Discov. Today *6*, 766 (2001).

70 J. Kawai, A. Shinagawa, K. Shibata, M. Yoshino, M. Itoh, Y. Ishii, T. Arakawa, A. Hara, Y. Fukunishi, H. Konno et al.: Nature *409*, 685 (2001).

71 K. Lindblad-Toh, E. Winchester, M.J. Daly, D.G. Wang, J.N. Hirschhorn, J.P. Laviolette, K. Ardlie, D.E. Reich, E. Robinson, P. Sklar et al.: Nat. Genet. *24*, 381 (2000).

72 J. Wallace: The Scientist *15*, 10 (2001).

73 W. Gilbert: Nature *271*, 501 (1978).

74 N. Heuze, S. Olayat, N. Gutman, M.L. Zani and Y. Courty: Cancer Res. *59*, 2820 (1999).

75 X. Zhang, J. Zhao, C. Li, S. Gao, C. Qiu, P. Liu, G. Wu, B. Qiang, W.H. Lo and Y. Shen: Nat. Genet. *27*, 151 (2001).

76 S. Xiao, C. Yu, X. Chou, W. Yuan, Y. Wang, L. Bu, G. Fu, M. Qian, J. Yang, Y. Shi et al.: Nat. Genet. *27*, 201 (2001).

77 E. Dequeker, S. Ramsden, W.W. Grody, T.T. Stenzel and D.E. Barton: Nature Rev. Gen. *2*, (2001).