Gene Therapy (2002) 9, 658–661

www.nature.com/gt

SESSION I – GENETIC BASIS OF INHERITED DISEASE

Genomics and medicine: hopes and challenges

J Kaplan
Hopitâl Cochin, Laboratoire de Génétique, Paris, France

The major impact of the completion of the human genome sequence is the understanding of disease etiology with deduced therapy. The catalog of monogenic diseases should be easily completed through in silico cloning. The major challenge today is to decipher the polygenic and multifactorial etiology of common diseases, such as cancer, cardiovascular, nutritional, allergic, auto-immune, degenerative disorders. In fact every gene, when mutated, is a potential disease gene, and we end up with the new concept of 'reverse medicine', by which we will derive new morbid entities and pathogenic pathways from the knowledge of the structure and function of every gene. A new molecular and supramolecular integrated physiology will help build a molecular pathophysiology of the different syndromes, from which etiologic therapy will derive. The revolution in nosology, the problem of predictive medicine when therapeutic curative or preventive measure are lagging behind, are some of the novel challenges that molecular medicine has to face.
Gene Therapy (2002) **9,** 658–661. DOI: 10.1038/sj/gt/3301739

Keywords: *complex traits; ethics; Human Genome Project; molecular medicine; proteomics*

The project of sequencing the human genome, launched in 1990, was originally expected to require 20 years to be completed. Actually the first draft of the whole human genomic sequence has been obtained after only 10 years. The major incentive in this enterprise was medical, and now that the 30 000 genes contained in the human genome are at hand, one may wonder how medicine will benefit from the success of the Human Genome Project. A word of caution about timing is necessary. The enormous quantity of sequence data recently accumulated and expected to merge ultimately into 24 huge contigs, one per chromosome, is not the end of genomics, but its beginning. Annotating the genome, ie extracting knowledge from raw sequence data will take a long time, possibly exceeding that needed to acquire them.

Cognitive issues

First of all we hope to learn the 'syntax', ie the logic of genomes including: regulatory processes, dynamics (recombination, transposition), variations (polymorphisms, with special emphasis on SNPs or single nucleotide polymorphisms), modulations, whether epistatic (modifying genes) or epigenetic (methylation and imprinting) and interactions with environmental factors. The ultimate goal is to extract from the dull unidimensional DNA sequence the complex tridimensional structure of proteins and RNAs interacting within cell architectural and functional networks.

Thus the next challenge is to go from genomics to proteomics, and from proteome to pathology.

Correspondence: J Kaplan, Hopitâl Cochin, Laboratoire de Génétique, Bâtiment Cassini, 27 rue de Fg St Jacques, 75679 Paris Cedex 14, France

Comparative genomics

Comparative genomics is very useful since there are more similarities than differences in the structure and function of genomes along the evolutionary tree. Hence, the genome of a worm such as *Caenorhabditis elegans*, a fly such *Drosophila melanogaster*, even a yeast such as *Saccharomyces cerevisiae*, are useful models for understanding the function and dysfunction of proteins.

Physiology revisited

The deciphering of the proteic actors within and around the cells will put physiology on molecular grounds helping to integrate molecular and cell biology, hence allowing us to dissect the pathophysiology of all syndromes.

Disease genes: from 'reverse genetics' to 'reverse medicine'

The identification of causative genes of all inherited diseases is one of the most expected consequences of the deciphering of genome sequences. There are several degrees of complexity. The first level is monogenic diseases (often called 'Mendelian' diseases by physicians, which is undoubtedly a misnomer, because it implies that polygenic diseases are non-Mendelian which is *stricto sensu* untrue). About 8000 different monogenic diseases have been itemized in the McKusick catalog, among which 1000 distinct causative gene have already been characterized by positional cloning. Most are rare diseases, also coined orphan diseases, with an incidence $<1/10\,000$. The deciphering of the whole genome and the reconstitution of the whole proteome is expected to ensure the identification of the many morbid genes that are still uncovered. This will be accomplished through the candidate gene strategy, comprising regional candidates, now reachable by simple computer interrogation (*in silico* cloning), and functional candidates when, in the

absence of regional assignment, one bets on the protein function correlated to the hypothesized pathophysiology. Regardless of the strategy used, the ultimate validation relies on the finding of pathogenic mutations in the suspected gene. The validation cannot be assessed on the loss or decrease of the corresponding protein, since this phenomenon can be secondary to a mutation residing in another gene.

The anticipated inflation of new proteins, each representing a possible candidate for morbidity, is one of the major challenges of the so-called 'post-genomic' era. The strategy of gene invalidation or induced mutations in orthologous genes in animal models (mouse, Drosophila, *C. elegans*) will be of great help. Characterizing the disease(s) related to each protein, the opposite of reverse genetics, is sometimes coined as 'reverse medicine'.

Classical nosology in pieces

Once a morbid gene is characterized, besides the immediate diagnostic benefit, the new information gathered is exploited to understand the pathophysiology. This implies investigation of all possible pathogenic mutations in the newly discovered gene (single identical mutation disorders, like sickle cell anemia, are the exception rather than the rule), and to correlate them to the phenotype, ie to the clinical manifestations and severity. From the experience acquired during the past 10 years, it is clear that there is no simple 'automatic' correlation between genotype and phenotype. A given mutation may give rise to a disease of varying degrees of severity. Moreover a given morbid gene may cause, when mutated, several distinct diseases (usually but not always explained by different mutations). The aphorism one gene → one disease, or even one mutation → one disease, is no longer tenable. Figure 1 shows some examples of the phenotypic heterogeneity of some inherited muscle diseases.

One of the main benefits expected from the current progress in genomics and proteomics is to understand the molecular basis of this lack of simple Boolean, or binary, correlation. Apart from environmental influences, there are endogenous factors encrypted in the genome itself, such as modifying genes, or polymorphisms in both coding and non-coding sequences, and some so-called neutral alleles may modulate the expression of a key protein. It is now clear that monogenic disases, in which only one gene is affected by an etiological mutation, can no longer be considered as monofactorial disorders. Finding these additional modulating factors is thus another crucial challenge.

Conversely it is not possible to deduce a specific phenotype from genotypic information, because we now know of many instances in which a given disease with a stereotyped phenotype may be caused by several alternative genes. Figure 2 illustrates the phenomenon of genetic heterogeneity in limb-girdle dystrophy, which may be produced by at least nine different genes.

From an etiological standpoint, both phenotypic and genetic heterogeneity invalidate the classical nosology, based upon anatomoclinical criteria. However, it is difficult to propose a more adequate nosology, since none of the emerging possible criteria of classification, based on the primary protein defect, the cell compartment involved, or the type of dysfunction, is comprehensive enough to qualify for replacing the old nosology.

The challenge of polygenic diseases, or 'complex traits'

These diseases are caused by mutations, or sequence variations, occurring simultaneously in several genes, and contrary to monogenic diseases, the segregation of the trait does not follow a strictly Mendelian mode. This is due to the fact that each individual genetic factor involved in the pathogenic assortment segregates according to Mendel's laws, but independently. The other difference is that, in contrast to the rare prevalence of monogenic disorders, complex diseases are quite common. They represent the major cause of morbidity and lethality, such as cardio-vascular disease, cancers, nutritional disorders (diabetes, obesity), auto-immune diseases (such as multiple sclerosis), degenerative disorders (Parkinson's disease, Alzheimer's disease). Due to the frequent and variable contribution of environmental factors, they are often considered as acquired diseases. In fact there are genetic components, the weight of which may vary considerably, with a continuum from relatively simple and rare situations, where few genes are coparticipating, to the complex cases where many genes are simultaneously contributing with an equivalent weight.

The identification of these contributing genes is not a simple task. About 5 years ago, there were great expectations concerning the identification of the genetic factors involved in common diseases, because it was believed that the strategies that proved to be so successful in the positional cloning of monogenic disorders would be applicable to polygenic disorders. This original belief has

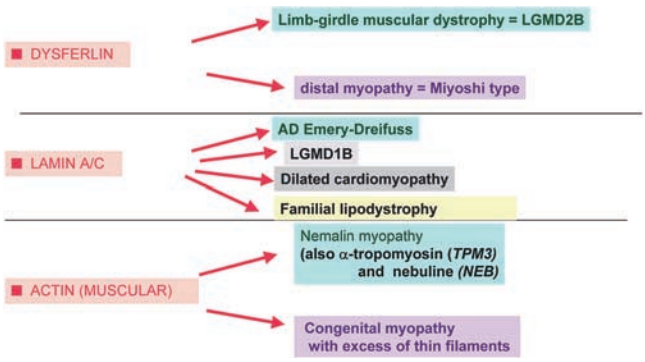

Figure 1 *Three examples of disease genes with phenotypic heterogeneity (one gene → several diseases). Note that the same homozygous mutation in the dysferlin gene was observed in two siblings, one exhibiting a proximal muscle dystrophy (LGMD2B), the other one the distal Myoshi-type dystrophy.*

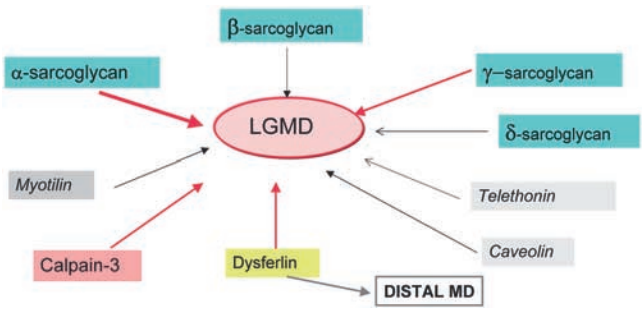

Figure 2 *Example of genetic heterogeneity (one disease ← several alternative causative genes). Note that a defect in the dysferlin gene may also produce a different phenotype.*

been disproven, as it is now clear that a variety of parametric and non-parametric statistical approaches have to be applied on extensive (thousands) numbers of DNA samples. So far there has not been much of a breakthrough, even for such widely studied illnesses as late-onset diabetes or multiple sclerosis. It is necessary to overcome an accumulation of hurdles: the involvement of many minor low penetrant genes; the difficult validation of the 'culprit' genes; the necessity to gather and investigate an extremely high number of family cohorts. Hopefully the knowledge of the full catalog of genes, markers and proteins should be of great help in this matter with the support of molecular epidemiology, bioinformatics and robotized massive gene analysis.

The molecular basis of individual susceptibility

Individuals do not react equally to environmental, infectious and iatrogenic challenges. These differences are due to disparities in the genetic make-up of individuals. With remarkable shrewdness, Sir Archibald Garrod wrote in 1931, 'Diathesis is nothing else but chemical individuality. The factors which confer upon us predispositions to and immunities from the various mishaps which are spoken of as diseases, are inherent in our very chemical structures, and even in the molecular grouping which confer upon us our individualities, and which went to the making of the chromosomes from which we sprang'. The characterization of SNPs, which occur on average every 1 kb, and the possibility to investigate simultaneously thousands of these markers by using microchips will represent a powerful and novel tool to delineate the molecular basis of individual idiosyncrasy.

Towards molecular medicine

Ultimately from the advances of genomics we will gain insight on the molecular mechanisms underlying all kinds of syndromes: infections (with understanding of virulence, susceptibility/resistance to microbes, resistance to antibiotics, etc), malignancies, neuropsychiatric illnesses, degenerative disorders, developmental diseases, senescence. The classical nosology (anatomoclinical) will have to be revisited in the light of molecular etiology.

Diagnostic issues

Genotyping is already currently performed as a diagnostic tool in an increasing number of monogenic diseases. In already declared disorders, such as major genetic disorders with early onset (infancy or childhood), there is no problem other than technical. Here gene testing has enormous value to ascertain the diagnosis at the molecular level, and to provide genetic counseling.

The problem is completely different when the test is performed at a presymptomatic stage, in an apparently normal individual. This is the case in illnesses with adulthood onset. This predictive medicine is highly beneficial whenever it is possible to prevent the appearance of the disase, and/or to cure it. Unfortunately such cases are rare (hemochromatosis, Mediterranean fever, chronic glaucoma), and the majority of late-onset monogenic diseases cannot be prevented or even cured, eg Huntington's disease, familial Alzheimer's disease, in which healthy carriers of the gene defect are doomed to a 100% chance of developing a lethal disease. Table 1 illustrates the various situations. Announcing to an individual the bad news that he has the mutation can be compared to a verdict of death penalty. Thus the practice of this type of medicine must be strictly controlled, and the future patient needs specific psychological assistance.

Even in familial cancers, such as familial breast cancer, because of variable age-dependent penetrance, the discovery of a mutation in the BRCA1 or the BRCA2 genes is a difficult issue, yielding a statistical risk which is difficult to manage, in the absence of standardized preventive measures. Thus predictive medicine is a double-edged sword: beneficial if prevention or cure is possible, detrimental if no action at all can be taken.

Therapeutic issues

Contrasting with the spectacular progress in the identification of many morbid genes, therapeutic progress is still lagging behind.

Ten years ago, it was believed that gene therapy would be the general panacea, because *in vitro*, or in simple cellular models (*ex vivo*), it was found that isolated genes may function, ie be transcribed and end up with the production of the desired protein. Hence the idea that the introduction in patients of the normal version of a defective gene, tinkered in order to reduce useless length (use of cDNA rather than the native genomic gene sequence) and to provide easy access to the nucleus and maximal level of expression (by placing the grafted sequence under the control of the best promoter, and using a variety of vectors, viral or non-viral to facilitate cell penetration). Three different therapeutic applications were sought: (1) to compensate a genetically defective function; (2) to obtain a therapeutic effect (DNA as a drug), for instance in cancer; (3) to promote vaccination. The initial expectations concerning the effectiveness of gene therapy in the two first applications were not fulfilled, since no proven cure has been obtained so far, with the exception of two recently published cases of apparent cure of infants suffering from a form of X-linked severe combined immunodeficiency, SCID-XI. In the latter cases the successful stable correction of a deficit in a cytokin receptor is explained by the selective advantage conferred on the transfected cells. For the time being, the real impact of genomics on gene therapy is difficult to foresee.

Two other forms of therapeutic exploitation of genome deciphering are now emerging. (1) Gene-based therapy: this novel strategy consists in devising diffusible drugs after elucidating the pathogenesis, once the causative role of a gene in the determinism of the disease (inherited or acquired) has been obtained. The rationale of gene-based drug therapy is : (i) to find the causative role of a gene (defective, or down-regulated or up-regulated) in a given disease or syndrome; (ii) to discover the role of the corresponding protein in normal and pathological conditions; (iii) to understand the pathogenesis at the biochemical level; (iv) to try to find drugs acting on appropriate targets of the pathogenic pathways, by interfering positiveley or negatively with the proteic factors or metabolites involved. This strategy represents genuine molecular etiological pharmacology, and it should greatly benefit from the progress of genomics.

There are still few examples where this approach was successful, for example Freidreich ataxia and chronic

Table 1 Examples of adult-onset diseases that can be diagnosed by gene testing at a presymptomatic stage

Disease	*Gene*	*Prevention or therapy in 2001*
Huntington disease	*huntingtin*	None
Alzheimer disease (familial)	*Presenilin 1 or 2 or several other genes*	None
Polycystic kidney disease	*PKD1 or PKD2*	Prevention of renal failure (?)
Breast/ovary cancer (familial)	*BRCA1 or BRCA2*	Preventive bilateral mastectomy + ovariectomy
Familial colic polyposis	*FAP*	Preventive total colectomy
Thyroid medullary cancer	*MEN2*	Thyroidectomy
Hemochromatosis	*HLAH*	Iron chronic depletion
Mediterranean fever	*Marenostrin/Pyrin*	Prevention on amyloidosis by colchicin
Chronic glaucoma	*MYOC*	Lowering intra-ocular pressure (pharmacology or surgery)

myeloid leukemia (CML). In the former the administration of a free radical-scavenger slows down the progression of heart deterioration. It was tried on the basis of the recently acquired evidence that the protein defect impairs mitochondrial iron transport, hence generating high amounts of free-radicals. In the case of CML the potent new drug STI 571 with anti-protein kinase activity was designed to inhibit the uncontrolled protein kinase hyperactivity of the fused bcr-abl protein produced by the 9/22 chromosome translocation of the Ph1 chromosome. (2) Genome-based therapy: this approach comprises pharmacogenetics dealing with individual differences in reactivity to drugs in terms of efficacy or tolerance; pharmacogenomics, in which high throughput screening of many molecules, either already existing or obtained by random combinatorial chemistry, is performed on arrays of thousands of genes, looking for specific genomic targets. This novel approach is currently being massively set up by major pharmaceutical companies; pharmacoproteomics: where proteins themselves are used as drugs. Progress in proteomics will greatly help this strategy, allowing protein design and engineering.

Finally the emerging and promising stem cell therapy (regenerative medicine) will benefit from basic progress in cell biology deriving from genome deciphering.

Societal issues

From the very beginning it was recognized that genomics would have a tremendous societal impact. Hence the creation in the USA of the ELSI program (ethical, legal and social implications) as a part of the HGP project, receiving 5% of its annual budget. An extensive treatment of the societal impact of genomics is beyond the scope of this review. Briefly, it involves economical and ethical issues.

Economy

Genomics is not only promising in medical and scientific terms, but also in financial terms. The market is huge, and billions of dollars are at stake. Consequently life sciences are now dominating the stock market and vice versa. Big pharmaceutical firms, as well as many new small start-ups, are betting on the rich harvest reasonably expected from genomics-derived discoveries, and invest tremendous amounts of money in programs involved in the medical applications that we described above. There is also a hot debate about gene patenting.

Ethics

Families with patients, and more generally the entire public, are becoming more and more aware of the medical relevance of morbid gene identification. Legitimately they are now expecting rapid therapeutical benefits. The fact that these are appearing too slowly, as compared with the swift accomplishments in genome sequencing and morbid gene identification may be a source of delusion, and generate loss of confidence in medicine. Another matter of concern is the fact that it is becoming possible through extensive genome analysis to gather information on individual sequence variations, and ultimately unveil one's genetic fate. This represents a risk of intrusion into individual privacy if the genetic information is not scrupulously kept confidential. There is also a risk of discrimination when a third party (insurance companies, employers) claims the right to have access to this information.

Coping with these problems is not simple, and in each country these difficult issues are publicly discussed. I also see a real threat in the unrestricted dissemination in the population of useless genetic kits. Patients are not mere health consumers and should be protected from the pressure of the market.

Finally, I would like to emphasize the increasing societal role of non-profit patients' associations. They represent a new partnership with significant lobbying power to protect the rights and interests of patients.

Conclusion

The deciphering of genomes will generate great benefits in medicine, if the societal impact is well geared. To enhance the pace of progress derived from genomics, it is necessary to boost some areas such as functional genomics, proteinology, bio-informatics, biotechnologies. Also it is mandatory to help the public to understand what is the present state of the art, through accurate information and education.

In 1902 Sir William Osler wrote, 'To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge that may be quickly available for the prevention and cure of disease, these are our ambitions'. Nowadays, this pledge still applies perfectly to molecular medicine, and should remain the creed of physicians in the 21st century.