

Genomics and medicine: an anticipation.

From Boolean Mendelian genetics to multifactorial molecular medicine

Jean-Claude Kaplan^{a*}, Claudine Junien^b

^a Inserm UR129, CHU Cochin, université Paris-V, 24, rue du Faubourg-Saint-Jacques, 75014 Paris, France

^b Inserm UR383, CHU Necker-Enfants malades, université Paris-V, 49, rue de Sèvres, 75015 Paris, France

Abstract – The major impact of the completion of the human genome sequence will be the understanding of diseases, with deduced therapy. In the field of genetic disorders, we will complete the catalogue of monogenic diseases, also called Mendelian diseases because they obey the Boolean logic of Mendel's laws. The major challenge now is to decipher the polygenic and multifactorial etiology of common diseases, such as cancer, cardio-vascular, nutritional, allergic, auto-immune and degenerative diseases. In fact, every gene, when mutated, is a potential disease gene, and we end up with the new concept of 'reverse medicine'; i.e., deriving new diseases or pathogenic pathways from the knowledge of the structure and function of every gene. By going from sequence to function (functional genomics and proteomics) we will gain insight into basic mechanisms of major functions such as cell proliferation, differentiation and development, which are perturbed in many pathological processes. By learning the meaning of some non-coding and of regulatory sequences our understanding will gain in complexity, generating a molecular and supramolecular integrated physiology, helping to build a molecular patho-physiology of the different syndromes. Besides those cognitive advances, there are also other issues at stake, such as: progress in diagnostic and prediction (predictive medicine); progress in therapy (pharmacogenomics and gene-based therapy); ethical issues; impact on business. © 2000 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

genomics / medicine

Résumé – Génomique et médecine : réflexions prospectives. De la génétique booléenne des lois de Mendel à une médecine moléculaire multifactorielle. La médecine sera l'une des grandes bénéficiaires de l'établissement de la séquence du génome humain. Cette connaissance permettra de comprendre les processus pathologiques et d'élaborer des traitements étiologiques. Grâce à la génomique, on pourra compléter le catalogue des maladies monogéniques et accélérer les progrès dans la recherche des gènes impliqués dans les maladies polygéniques, et multifactorielles qui constituent l'essentiel de la pathologie de l'adulte (cancer, maladies cardiovasculaires, maladies de la nutrition, maladies auto-immunes, maladies neurodégénératives). Il est légitime de penser que chaque gène a sa pathologie propre, et que lorsque tous les gènes, et les protéines qui en dérivent, seront connus il s'agira de connaître les maladies dans lesquels ils sont impliqués. C'est ce qu'on appelle parfois la « médecine inverse ». Il faut comprendre à présent la fonctionnalité des gènes et des protéines (génomique et protéomique fonctionnelles), élucider la syntaxe génomique et élaborer une physiologie moléculaire intégrée. À côté de ces enjeux cognitifs, on doit envisager les retombées en matière de diagnostic, permettant notamment de prévoir le

* Correspondence and reprints: Fax : 01 44 41 15 22.
E-mail address: kaplan@icgm.cochin.inserm.fr (Jean-Claude Kaplan).

développement de certaines pathologies avant l'apparition des symptômes (médecine prédictive). Des avancées majeures dans le domaine thérapeutique découleront de la pharmacogénomique et des thérapies fondées sur la connaissance des gènes. Tous ces progrès, accomplis grâce à la génomique, comportent aussi de très importants enjeux d'ordre éthique et économique. © 2000 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

génomique / médecine

Version abrégée

Le projet prométhéen de déchiffrement du génome humain « Human Genome Project », lancé au début des années 1990, est pratiquement achevé avec l'annonce, fin juin 2000 de la première version brute de la séquence de l'ADN génomique de *Homo sapiens* (trois milliards de paires de bases). Le but aura été atteint en un temps record, ce que même les plus optimistes n'avaient pas imaginé.

Cependant, même si l'on prévoit que toutes les séquences codantes seront identifiées rapidement, permettant un inventaire de tous les gènes humains, dont le nombre varie entre 30 000 et 100 000, selon les estimations, il reste encore à savoir comment ces gènes sont régulés, et à connaître les protéines qui en dérivent. Cela revient à comprendre la logique de l'organisation du génome, de sa plasticité, de son fonctionnement dans son environnement cellulaire, et au cours du temps. D'une génomique unidimensionnelle (information brute sur les séquences codantes) on passe à une génomique tridimensionnelle. On peut en dire autant de la nécessité de comprendre la forme et la fonction des quelques 80 à 100 000 protéines qui constituent la panoplie des protéines humaines, ainsi que les multiples réseaux d'interactions qui expliquent la naissance, la vie et la mort des cellules, des tissus, de l'organisme entier.

Notre réflexion porte sur les retombées médicales de la connaissance des génomes. Elle est nécessairement transversale car les enjeux médicaux concernent tous les aspects de la génomique.

Nous avons regroupé ces enjeux en cinq champs de réflexion correspondant chacun à un impact significatif de la génomique : 1° cognitifs ; 2° biotechnologiques ; 3° diagnostiques ; 4° thérapeutiques ; 5° économiques ; 6° éthiques.

1. Impact médical d'ordre cognitif : la compréhension des maladies

Il faut apprendre la syntaxe génomique (logique de sa régulation, de sa dynamique), et dresser l'inventaire des variations non pathologiques, c'est-à-dire des polymorphismes, avec le déchiffrement des millions de SNP (*single nucleotide polymorphisms*), déchiffrer les fac-

teurs modulant l'expression du génome (épistatiques et épigénétiques).

La génétique comparative est un outil d'une puissance considérable, car l'analyse et la compréhension des génomes d'organismes plus simples permet déjà de mieux comprendre la génomique des organismes complexes. Il est probable que, à côté de la souris, la drosophile et même la levure vont bientôt constituer des modèles intéressants pour la compréhension de la physiopathologie des maladies humaines.

L'inventaire des gènes de protéines (qui occupent seulement 10 % du génome humain) permettra de dresser le catalogue complet des protéines humaines constituant ce qu'on appelle désormais « le protéome ». La protéomique va devenir un champ d'investigation majeur, tant au niveau moléculaire (déchiffrement des partenaires dans les interactions) qu'au niveau cellulaire.

Les maladies monogéniques, avec leur hérédité simple parfaitement conforme aux lois de Mendel, obéissant à une logique comparable à celle de Boole (pour chaque gène c'est la loi du tout ou rien, avec une logique de type 1/0), ont été les premières à bénéficier des progrès de la génomique balbutiante. Environ un millier de gènes de maladies ont déjà été identifiés par les procédés regroupés sous le vocable de « clonage positionnel ». Le déchiffrement complet du génome facilitera grandement l'identification des milliers d'autres gènes impliqués dans des maladies monogéniques et restant à découvrir. Cela est d'autant plus important qu'ils concernent des affections très rares, donc inaccessibles à l'analyse de liaison pour obtenir une première localisation. Le travail de clonage aura de plus en plus lieu devant des écrans d'ordinateurs, ce qu'on a appelé plaisamment « clonage in silico ». Les modèles animaux, induits par invalidation génique (ou *knock-out*) ou par induction de mutations spécifiques seront de puissants outils dans cette démarche. On peut extrapoler et imaginer que tout gène est susceptible d'entraîner une pathologie s'il est muté, ce qui revient à associer une, ou plusieurs pathologies à chaque protéine d'où la notion de « médecine inverse » (on a le gène, reste à trouver la pathologie correspondante).

Un des défis restant à résoudre est celui de la compréhension des corrélations entre génotype et phénotype, ce qui amène à la création d'une physiopathologie moléculaire, source d'ouvertures thérapeu-

tiques nouvelles et étiologiques. La découverte du réseau des relations fonctionnelles entre gènes, dont aucun ne « parle tout seul », débouche directement sur la notion de terrain, dont les bases moléculaires vont enfin pouvoir être définies. Dans ce nouvel éclairage qui devrait permettre d'expliquer les susceptibilités individuelles, vis-à-vis des pathogènes, des parasites, des médicaments, l'établissement d'un riche catalogue de SNP jouera un rôle décisif.

Un autre grand défi est celui des maladies polygéniques, qui représentent la pathologie courante et la principale cause de mortalité chez l'adulte (maladies cardio-vasculaires, cancers, maladies de la nutrition, maladies auto-immunes, maladies dégénératives). L'identification de ces gènes sera très difficile, et leur validation encore plus, mais c'est un des objectifs prioritaires sur l'agenda de la génétique moléculaire.

2. Biotechnologies

Pour pouvoir gérer l'énorme masse de données informatisées (bio-informatique), et pour refonder une nouvelle chimie des protéines de nouveaux outils sont nécessaires. Ceux-ci sont activement recherchés et mis au point. Citons deux exemples a) les bancs automatisés pour la création et l'exploitation de « bio-puces » permettant l'analyse simultanée de milliers de gènes et de leurs transcrits ; b) l'analyse comparative de la panoplie protéique des cellules normales et pathologiques par des stations automatisées d'électrophorèse bi-dimensionnelle couplée à la spectrographie de masse, avec en fin de chaîne l'identification des produits par référence aux banques de données informatisées.

3. Diagnostic

Les progrès accomplis dans l'identification des gènes de maladies, ainsi que dans l'exploration systématique et simultanée d'un très grand nombre de gènes par les « bio-puces » permettront de prévoir longtemps à l'avance le développement de certaines pathologies de l'adulte, c'est à dire de diagnostiquer les prédispositions. Cette nouvelle forme de médecine constitue la « médecine prédictive », qui ne va pas sans causer de graves problèmes éthiques lorsque, et c'est encore très souvent le cas, il n'existe aucun moyen de prévention ou de traitement de la maladie annoncée.

4. Enjeux thérapeutiques

Après la moisson des gènes impliqués dans les maladies monogéniques les plus fréquentes et les plus

graves, le temps est venu d'exploiter les connaissances acquises pour traiter, voire guérir ces affections le plus souvent très graves. On a cru un peu vite que la thérapie génique, c'est-à-dire la cure par l'introduction dans l'organisme du malade de la version normale du gène défectueux, allait représenter une sorte de panacée universelle. Un recul de près de dix ans nous a appris à prendre conscience des difficultés. À l'heure actuelle, à côté de cette vision un peu simpliste qui ne tenait pas compte des problèmes d'adressage, et en somme de « galénique moléculaire », on envisage surtout une thérapie fondée sur la connaissance des gènes, c'est à dire exploitant la cascade physiopathologique que la découverte de la mutation causale aura permis de mettre à jour dans chaque maladie. Une autre stratégie, en prise directe avec la génomique, consiste à cribler dans des expériences à haut débit des milliers de gènes pour trouver des molécules susceptibles de modifier dans le sens recherché (inhibition ou stimulation) l'expression des gènes-cibles. Cette nouvelle branche de la pharmacologie est la « pharmacogénomique » (qu'il ne faut pas confondre avec la pharmacogénétique qui est le domaine de la susceptibilité individuelle aux médicaments). Les grandes firmes pharmaceutiques regroupées en un petit nombre de géants industriels l'ont bien compris, puisqu'elles investissent des sommes énormes, considérant sans doute à juste titre qu'ils trouveront ainsi les médicaments des grandes pathologies communes. Dans la même veine, la protéomique, fille de la génomique, permet d'envisager une « pharmacoprotéomique ».

5. Enjeux de société

La génomique fait irruption dans la société pour de multiples raisons. En dehors des spectaculaires effets d'annonce des découvertes bio-médicales dans la grande presse avant la presse scientifique, l'impact de la génomique touche essentiellement deux domaines : la vie privée des individus et de leurs familles, et l'argent, posant ainsi des problèmes d'ordre éthique et d'ordre économique.

L'éthique est en jeu par le simple fait que l'on a désormais accès à l'intimité du programme génétique de chacun, voire de son avenir morbide, et, si on pousse à l'extrême, de sa longévité tout court. Le fait que le destin génétique de chacun puisse à présent être dévoilé, d'abord à l'individu lui-même (patient présent ou futur), puis à la famille, enfin à la société toute entière (employeurs, assurances et, surtout, systèmes de santé), pose de nombreux et sérieux problèmes éthiques. Ils sont d'un ordre très différent de ceux qui se posaient jusqu'alors en médecine.

Quant à l'économie elle escompte de nombreuses retombées de la génomique, et le marché de la santé

connaît un véritable *boom*. Les grandes firmes, comme les « start-ups » misent sur cet Eldorado. Cette irruption de la Bourse dans les laboratoires, même publics, est

aussi une des conséquences inattendues de la génomique.

1. Introduction

The project of sequencing the human genome (three billion base pairs per haploid genome) was launched in 1990, under the banner of the Human Genome Project (HGP), and was originally expected to be completed within about 20 years. In fact, the first draft of the whole human genomic sequence is expected to be obtained in the course of 2000, thanks to a fierce battle between the public non-profit international consortium of the HGP, led by Francis Collins, and the private for-profit Celera company, led by Craig Venter [1].

Our purpose is to give a general survey of the medical impacts of this feat, long considered as a Holy Grail.

It should be emphasized that the major incentive in this enterprise has been medical. Even those who were reluctant to pour so much money into the sequencing of the human genome quickly became permeable to the medical importance of the project. Deciphering the genome (which is defined by the sum genes + non-coding sequences) should provide the complete catalogue of genes and proteins, among which many, if not all, are susceptible to be involved to some extent in human disease whether inherited or acquired.

Genomics is a new sub-speciality dealing with the study of genomes of any species, from the most modest prokaryotes to that of *Homo sapiens*. Since it is a novel branch of biology, we shall try in this article to anticipate the likely consequences on medicine.

The idea of reading the whole human genome sequence emerged when it became obvious that hunting for monofactorial disease genes, obtained by linkage analysis followed by physical mapping (formerly called reverse genetics, and now coined 'positional cloning') was extremely rewarding in solving some major medical mysteries (Duchenne muscular dystrophy, cystic fibrosis, Huntington disease, etc.). But it was also clear that this indirect strategy to reach the morbid genes was painstaking and cumbersome. As thousands of sequenced sections of DNA were poured into the databases, it became clear that the morbid genes could be more expeditiously found by computer searching after preliminary regional assignments ('in silico' strategy). Ultimately, it was evident that the HGP was the only way to establish without delay the complete catalogue of the 30 000 to 100 000 human genes, interval between the minimalist [2, 3] and the maximalist [4] estimations.

2. The issues at stake

Genomics will have multiple impacts on medicine, among which we identify five main categories: a) the

cognitive issues; b) the consequences on biotechnologies; c) the diagnostic issues; d) the consequences on therapies; and d) the societal issues.

2.1. Cognitive issues

First of all we hope to learn the 'syntax', i.e. the logics of genomes (regulations), dynamics (recombination, transposition); variations (polymorphisms, with special emphasis on SNPs or single nucleotide polymorphisms, which are expected to occur every 300–100 nucleotides); modulations, whether epistatic (modifying genes) or epigenetic (methylation and imprinting); and interactions with environmental factors. The overall expectation is to pass from the dull unidimensional sequence to a complex tridimensional structure tightly integrated in the cellular architectonics and function.

Comparative genomics has shown more similarities than differences in the structure and function of genomes throughout evolution. Hence, the genome of a worm such as *Caenorhabditis elegans*, a fly such as *Drosophila melanogaster*, even a yeast such as *Saccharomyces cerevisiae*, may represent useful models for the understanding of human disease.

2.1.1. From genomics to proteomics

From the coding sequences of genomes (only 10 % of the genome length in *H. sapiens*) the whole catalogue of proteins will be deduced. The total number of proteins remains conjectural, somewhere between 50 000 and 100 000, not counting multiple isoforms. To study in every entity the structure, the function, the expression in space (differentiation) and time (development), and the multiple interactions involved in architectural and functional cellular networks represents a formidable challenge, defining a global novel conception of proteinology coined 'Proteomics'.

2.1.2. 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 patho-physiology of all syndromes.

2.1.3. 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 Genomics. 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). Approximately 8 000 different monogenic diseases have been itemized in the McKusick catalogue, among which 1 000 distinct causative genes have already been characterized by positional cloning as of February 2000 [5]. Most are rare 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 so-called candidate gene strategy, in which there are two sorts of candidates. i) Regional candidates, if linkage analysis has provided conclusive results as to the chromosomal assignment of the morbid locus. In this case all the genes present in the suspected region are found by simple computer interrogation (*in silico* cloning), and can be screened for mutations in the patient's genome. ii) Functional candidates, which are useful whether or not there is a prior indication about the regional assignment of the morbid locus. They are based on the knowledge of protein function correlated to the hypothesized patho-physiology. For instance, in a muscular disorder, one would first search among proteins that play a role in the architecture or in the function of muscle cells. Regardless of the strategy used, the ultimate validation relies on the finding of pathogenic mutations in a given 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 'post-genomic' area. The strategy of gene inactivation or induced mutations in orthologous genes in animal models (mouse, *Drosophila*, *C. elegans*) will be of great help. Finding the disease(s) that are related to each protein, the opposite of reverse genetics, is sometimes coined 'reverse medicine'.

2.1.4. From genotype to phenotype and vice-versa

Once a morbid gene is characterized, besides the immediate diagnostic benefit, the new information gathered is exploited to understand the patho-physiology. This implies the investigation of all possible pathogenic mutations in the newly discovered morbid gene (single identical mutation disorders such as sickle cell anaemia are the exception rather than the rule), and their correlation to the phenotype, i.e. to the clinical manifestations and severity. From the experience acquired over the past 10 years, it is clear that there are no simple 'automatic' correlations between genotype and phenotype. A given mutation may often give rise to variable degrees of severity. Moreover, a given morbid gene may cause, when mutated, several distinct diseases (usually but not always explained by different mutations). 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' correlation, i.e. to define the additional factors involved. Not speaking of the environmental influence,

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 are capable of modulating the expression of a key protein. In this line, it is now clear that monogenic diseases, i.e. the disorders where only one gene is affected by an etiological mutation, can no longer be considered as mono-factorial disorders. Finding the additional modulating factors will be just as difficult as deciphering the genes affected in polygenic diseases (see below).

Conversely, it is not possible to proceed automatically from phenotype to genotype, because there is an increasing number of examples, such as myopathies, retinopathies and deafness, in which a given disease with a stereotyped phenotype may be caused by several alternative genes (genetic heterogeneity).

2.1.5. The problem of polygenic (complex) diseases

These diseases are caused by mutations 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 causes of morbidity and lethality, such as cardio-vascular disease, cancers, nutritional disorders (diabetes, obesity), auto-immune diseases (such as multiple sclerosis) and degenerative disorders (Parkinson disease, Alzheimer 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 co-participating, to the complex cases where many genes are simultaneously contributing with an equivalent weight.

The identification of these 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 been deluded, as it is now clear that a variety of parametric and non-parametric statistical approaches has to be applied on extensive (thousands) numbers of DNA samples. So far there has not been much of a breakthrough, even for widely studied illnesses such 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 sample large family cohorts. Hopefully the knowledge of the full catalogue of genes, markers and proteins should be of great help in this matter with the support of molecular epidemiology and bioinformatics.

2.1.6. *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 one kb, and the possibility of simultaneously investigating thousands of these markers by using biochips (see below) will represent a powerful and novel tool to delineate the molecular basis of individual idiosyncrasy.

2.1.7. *Towards molecular medicine*

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

2.2. **Biotechnologies**

With the advent of genomics we have to face two major theoretical challenges. One is qualitative, and concerns the comeback of biochemistry of low abundance proteins, and the uncovering of their intricate interactions. The other is quantitative: how will we be able to deal with the enormous mass of data suddenly arising (sequences, structures)? New tools will have to be devised at the different levels of complexity: i) at the molecular level: mass screening of genome and transcriptome by micro-arrays on chips (biochips), of proteome by mass spectrometry; ii) at the cellular level: by using appropriate models, particularly using the emerging possibility of obtaining and manipulating pluripotent stem cells; and iii) in vivo: by the manipulation of the genome of an increasing number of organisms serving as models. To exploit the comprehensive databases hosting all the information on genomes, mutations, non-pathogenic variations, phenotypes, bio-informatics, still in its infancy, is going to become a major speciality.

Genomics requires high speed, standardization and mass screening that can only be achieved by full-automation at all levels to obtain high-throughput. This implies a trend towards an industry-like type of research, leaving behind the obsolescent cottage-type organization of traditional academic research.

2.3. **Diagnostic issues**

Genotyping is already currently performed as a diagnostic tool in an increasing number of monogenic diseases.

There is no problem other than technical in already declared disorders, such as major genetic disorders with early onset (infancy or childhood), where it has an enormous value to ascertain the diagnosis at the molecular level, and to provide genetic counselling.

The problem is completely different when the test is performed at a pre-symptomatic 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 disease and/or cure it. Unfortunately such cases are rare (haemochromatosis, Mediterranean fever, chronic glaucoma), and the majority of late-onset monogenic diseases cannot be prevented or even cured, e.g. Huntington disease, familial Alzheimer disease, in which healthy carriers of the gene defect are doomed to a 100 % chance of developing a lethal disease. Announcing to an individual the bad news that he has the mutation can be compared to a verdict of the 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.

2.4. **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, i.e. be transcribed and end up with the production of the desired protein. Hence the idea of the introduction in patients of the normal version of a defective gene, tinkered with 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: i) to compensate a genetically defective function; ii) to obtain a therapeutic effect (DNA as a drug), for instance in cancer; and iii) to promote vaccination. The initial expectations concerning the efficacy 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 [6]. For the time being the real impact of genomics on gene therapy is difficult to foresee.

In contrast, two other forms of therapeutic exploitation of genome deciphering are very promising.

2.4.1. 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. This gene-based drug therapy approach is very promising. It has been recently illustrated by a success obtained in Friedreich ataxia. In this disease the progression of heart deterioration could be slowed down by a free-radical scavenger. This drug was tried after it was shown that the defect in frataxin impairs mitochondrial iron transport, hence generating high amounts of free-radicals [7]. The general rationale of such an approach is: a) to find the causative role of a gene (defective, or down-regulated or up-regulated) in a given disease or syndrome; b) to discover the role of the corresponding protein in normal and pathological conditions; c) to understand the pathogenesis at the biochemical level; and d) to try to find drugs acting at any step of the pathogenic pathways, by interfering positively or negatively on the proteic factors or metabolites involved. This strategy represents genuine molecular etiological pharmacology, and it should greatly benefit from the progress of genomics.

2.4.2. Genome-based therapy

This approach can be subdivided into three subcategories.

- Pharmacogenetics, already mentioned, dealing with individual 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 set up on a large scale by major pharmaceutical companies.
- Pharmacoproteomics, where proteins themselves are used as drugs. Progress in proteomics will greatly help this strategy. Recognition at the structural and functional levels of the recurrent protein domains should be extremely useful in protein design and engineering.

2.5. 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. A thorough treatment of the societal impact of genomics is beyond our scope.

Briefly, it involves mainly economy and ethics.

2.5.1. 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. Large pharmaceutical firms, as well as many new small start-ups are betting on the rich harvest reasonably expected from genomics-derived discoveries, and are investing tremendous amounts of money in programmes involved in the medical applications that we described above. There is also a great debate on gene patenting. The war between supporters and opponents of patenting is crystallized in the battle opposing the private for-profit initiatives of the Celera company (Craig Venter) and the public non-profit programmes of the HGP (Francis Collins).

2.5.2. Ethics

The families of patients, and more generally the entire public, are becoming increasingly 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 to the swift accomplishments in molecular genetics, may be a source of delusion, generating a crisis of confidence in medicine. Even more important is the fact that it is now possible through genome analysis to detect mutations and polymorphisms, to ultimately unveil one's genetic fate. There is a real risk of intrusion into individual privacy if the genetic information is not kept scrupulously confidential. There is also a risk of discrimination when a third party (insurance companies, employers, etc.) claims, sometimes legitimately, the right to have access to this information.

Answering these problems is not simple, and there is now a world-wide debate on these difficult issues. We 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.

It is also important to emphasize the societal relevance of numerous non-profit Patients' Associations. They represent a new partnership with significant lobbying power to protect the rights and interests of patients. Some are very powerful, raising important funds by charity campaigns, to help research and care, and they are playing an increasingly important role in society.

3. 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 and biotechnologies. Also, it is mandatory to help the public understand what is the present state of the art, through better and 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 may remain the creed of physicians in the 21st century.

4. Bibliography

Nature insight: Functional genomics, *Nature* 405 (2000) 819–865.

Beaudet A., Making genomic medicine a reality, *Am. J. Hum. Genet.* 64 (1999) 1–13.

Cassier M., Relations entre secteurs public et privé dans la recherche sur le génome, *Méd. Sci.* 16 (2000) 26–30.

Collins F., Positional cloning: let's not call it reverse anymore, *Nat. Genet.* 1 (1992) 3–6.

Collins F., Medical and societal consequences of the Human Genome Project, *N. Engl. J. Med.* 341 (1999) 28–37.

Jordan B., *Génétique et Génome. La fin de l'innocence*, 1 vol., Flammarion, 1996.

Jordan B., *Années génome: les dix glorieuses*, *Med. Sci.* 16 (2000) 5–9.

Kaplan J.-C., Junien C., *Génomique et médecine. Une gestion de la complexité*, in: *Développement et applications de la génomique- L'après-génome*, Académie des Sciences: Rapport sur la technologie n°1 Editions Tec & Doc, Paris, 1999, pp.167–185.

Lander E., Weinberg R., *Genomics: journey to the center of biology*, *Science* 287 (2000) 1777–1782.

Rosa J., *La révolution du génome*, *Med. Sci.* 16 (2000) 5–9.

Weatherall D., *Science and the quiet art. Medical research and patient care*, 1 vol., Oxford University Press, 1995.

References

[1] Weissenbach J., Salanoubat M., *Séquence des génomes : le feu d'artifice*, *Médecine/Sciences* 16 (2000) 10–16.

[2] Ewing B., Green P., *Analysis of expressed sequence tags indicates 35,000 human genes*, *Nat. Genet.* 25 (2000) 232–234.

[3] Roest Crollius H., Jaillon O., Bernot A., Dasilva C., Bouneau L., Fischer C., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W., Weissenbach J., *Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence*, *Nat. Genet.* 25 (2000) 235–238.

[4] Liang F., Holt I., Pertea G., Karamycheva S., Salzberg S.L., Quackenbush J., *Gene index analysis of the human genome estimates approximately 120,000*, *Nat. Genet.* 25 (2000) 239–240.

[5] Antonarakis S., McKusick V., *OMIM passes the 1,000-disease-gene mark*, *Nat. Genet.* 25 (2000) 11.

[6] Cavazzana-Calvo M., Hacein-Bey S., de Saint Basile G., Gross F., Yvon E., Nusbaum P., Selz F., Hue C., Certain S., Casanova J.L., Bousso P., Deist F.L., Fischer A., *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease*, *Science* 288 (2000) 669–672.

[7] Rustin P., von Kleist-Retzow J.C., Chantrel-Groussard K., Sidi D., Munnich A., Rotig A., *Effect of idebenone on cardiomyopathy in Friedreich's ataxia: a preliminary study*, *Lancet* 354 (1999) 477–479.