Editorials

The Human Genome: Implications for Medicine and Society

Maria - Cristina Kotta, Aris Anastasakis, Pavlos Toutouzas

Pheidippides Project, Department of Cardiology, University of Athens, Hippokration Hospital

"Intelligent life on a planet matures from the moment it discovers its origin."
Richard Dawkins, 1976[1].

Key words:
Genetics, genomics, polymorphisms.

Manuscript received:
April 24, 2001;
Accepted:
December 11, 2001.

Corresponding Author:
M.C. Kotta

5, Lambrou Porfira Str.
111 41 Athens
e-mail:
mckotta@hotmail.com

On June 26, 2000, at a press conference in the White House, President Bill Clinton announced the completion of the working draft for the decoding of human genome. Among the participants were Craig Venter, head of Celera Genomics (Rockville, Maryland) corporation, and Francis Collins, director of the National Human Genome Research Institute (NHGRI). Research teams of the above mentioned organizations, representing different bodies of the private and public sector respectively, had independently defined the greater part of the human genome nucleotide sequence. Similar announcements were also made in Europe, especially in Great Britain, a country participating to a great extent in this research[2].

Approximately seven months later, on February 15 and 16, 2001, these two bodies (public and private) published the results of human genome decoding along with the relevant methodology that has been utilized, the results analysis and the corresponding conclusions, in the internationally acknowledged scientific journals Nature[3] and Science[4], respectively [3. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921, 4. Venter, J.C. et al. (2001) The sequence of the human genome. Science 291: 1304-1351].

The Human Genome Project (HGP), as this project is internationally known, does not constitute an "invention" but a "discovery", which resulted from an international and coordinated effort, the systematic implementation of know-how and, also, from a great number of scientific innovations. The first human genome map already developed, as well as its complete and 99.99% accurate publication expected by 2003 or earlier[5], constitute the starting point for the oncoming biological revolution of the 21st century. It must be noted that this revolution relates not only to sciences relevant to biology, like medicine or pharmacology, but also to other sciences like psychology, sociology, or law.

The Human Genome Project chronicle

The history of the HGP starts on December 1984, at a small conference in Utah, USA. At that time, the US Department of Energy (DOE) was interested in the development of methods for the diagnosis of genetic material alterations (mutations), due to radiation exposure and in understanding their respective consequences on the human body. Al-

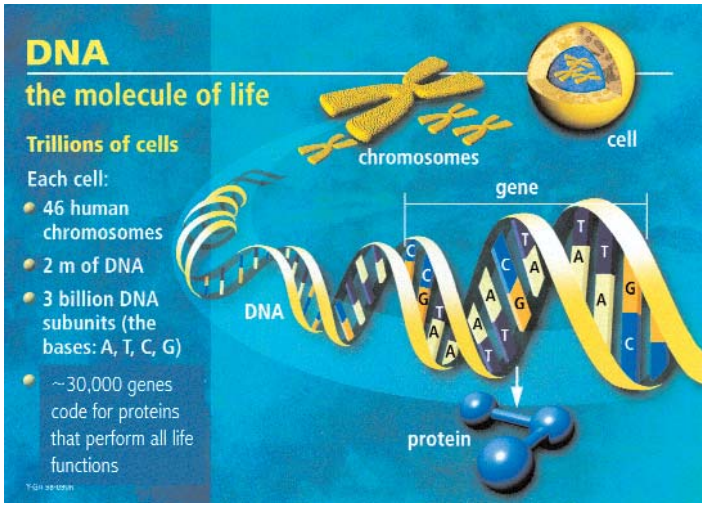


Figure 1. Genome is defined as the whole set of genetic instructions for the creation and function of an organism from conception until death. The human genome is distributed in 23 pairs of chromosomes, structures that consist of DNA (the genetic material) and proteins, and are present in the nucleus of every somatic cell. DNA (deoxyribonucleic acid) is a polymer of simpler chemical compounds, the nucleotides, that consist of a nitrogenous base, a sugar (deoxyribose) and phosphoric acid. The nitrogenous bases are adenine (A), thymine (T), guanine (G) and cytosine (C). Genes are segments of DNA whose particular order of bases defines a particular genetic information. Our genetic alphabet consists of approximately 3 billion bases. The sequence of 10% of these bases constitutes the approximately 30,000 human organism. The rest of our genetic material is, yet, of unknown function.

The above graphic has been created by the U.S. Department of Energy, Human Genome Project (http://www.ornl.gov/hgmis.html).

though DOE had, at that time, developed tests for tracing mutations, more sensitive tracing methods were needed to be developed. At the Utah conference, which was organized on the initiative of the then DOE Director, David Smith, it was made evident that the development of a reference field, which would constitute the basis for any kind of research, was necessary in order to further implement and develop the genetic know-how, as well as to identify large areas of genetic material. And thus, the attempt for decoding the human genome was proposed[6,7] (Figure 1).

During the 1984-1988 period, the HGP was still under discussion. As happens in every similar great effort, DOE's announcement to initiate this project encountered many reactions, friendly or not. A large part of the scientific community was very skeptical as to whether mapping of the human genome was feasible and, if so, at what financial cost[7].

In 1988, the American Congress ratified the commencement of the project, under the name of Human Genome Project, and appointed the National Institute of Health (NIH) as coordinator. One year later, in 1989, the National Center for Human Genome Research (NCHGR) was established within NIH, under the direction of James Watson, and it was later renamed to its present name of National Human Genome Research Institute (NHGRI). Watson remained in charge of NHGRI up until 1992, when he was forced to resign and was succeeded by Francis Collins, who holds this position to date[8].

Due to the range of the project, the HGP acquired, an international profile in 1995 and several European research centers were included in the program; the newly established Sanger Center, Cambridge, Great Britain, was the main participant and was financially supported by Wellcome Trust[9].

The role of Celera Genomics

As above mentioned, the HGP started as a non profit program, mainly financed by the American government and the British corporation Wellcome Trust. In 1998, however, the public HGP faced competition following the establishment of the private corporation Celera Genomics, with Craig Venter as its chairman. Celera has been established as a subsidiary of Perkin-Elmer Biosystems corporation (Norwalk, Connecticut), as the latter was interested in funding the decoding of the human genome aiming at faster acquisition of results, which would in turn be at the disposal of its clients. Some of its clients are Amgen Inc. (Thousand Oaks, California), Novartis Pharma (Basel, Switzerland) and Pharmacia & Upjohn (Bridgewater, New Jersey)[10].

The establishment of Celera corporation, in 1998, was the starting point of a conflict with the public HGP, a conflict which still continues. The differences of these two organizations can be summarized along two main axes. The first of these axes relates to the experimental methods used for the purpose of genome decoding, and the second to the release and utilization of results[11].

More specifically, Celera used the shotgun method for the definition of the nucleotide sequence of the human genome (human whole genome shotgun sequencing)[12]. This method relates to the accidental fragmentation of DNA into several fragments, the automated determination of their sequence and their classification in a sequence, with the aid of powerful computers and algorithms. It is worth noting that Celera has been equipped with state-of-the-art automated analyzers (sequencers) for the determination of the nucleotide sequence, the Perkin-Elmer PRISM 3700, that operate on a network of capillary tubes and need minimum technical support. Moreover, Celera was equipped with a powerful computer from Compaq Corporation (Houston) for the analysis of results. Its value is estimated at about 80 million dollars ("ASCI Red" military computer of the US government is the next more powerful computer in the world and it is used for planning nuclear bomb testing)[10].

In addition, the release and utilization of results constituted another point of contention between Celera and public HGP. In 1998, when Celera announced the commencement of decoding the human genome, at the same time stated its intention to acquire the "rights" (patents) for several genes[10]. This announcement forced the public HGP to increase the pace of results production and to somehow lower its standards. While the initial intention was to complete 99.99% of the human genome map by 2003 with an experimental repeatability that would minimize errors, finally and for competitive reasons, they settled for a 90% complete version with lower repeatability[13,14]. There is a special importance attributed to repeatability, since automated determination of the genome's nucleotide sequence (as every other DNA sequence) often contains errors (sequencing errors), which can be identified and eliminated only through the repetition of the method, so as to finally obtain an accurate and reliable sequencing[15].

The policy of the announcement of results has been internationally shaped by the Bermuda Convention of 1996, which dictates the announcement of the results in Internet databases within 24 hours from the time of their production[16]. Of course, Celera's initial stance was to permit public access to its results on a subscription fee[17], but this position created, as expected, such a tumult, that the corporation revised its attitude[18]. The joint declaration of the American President Bill Clinton and the British Prime Minister Tony Blair, that they shall encourage and support immediate announcement of the results, so that they may be at the immediate disposal of any interested party[16], played an important role in this matter.

Public HGP's approach and its importance

Contrary to Celera's experimental approach, public HGP approach did not, for the most part, relate to the large scale definition of the genome's nucleotide sequence (large-scale sequencing) but to the genome's mapping and then to the definition of its sequencing[19,20]. With the term mapping we refer to the definition of the location of a genetic area (be it a gene or not) on a particular chromosome[21]. Looking, for example, for a particular gene within the whole of the human genome is like looking for a particular house within a country. In order to locate the street, we need maps for guidance. We have to start with a map of the country, then the map of the city, the map of the neighborhood and, finally, we locate the particular address.

Genetic maps, like their street counterparts, are designed at different scales and at differing high or low resolution levels (high/low resolution map). Genetic linkage maps are included among the low resolution maps. These maps are used to indicate the relative location of genetic markers on a chromosome. Genetic markers are DNA segments, that are either expressed (i.e., are genes) or not, and their inheritance can be confirmed in the laboratory (for instance, using molecular hybridization techniques such as Southern and FISH, PCR techniques, DNA sequencing, restrictive endonucleases analysis, etc.).

On a genetic linkage map, the gene (or other markers) locus is given in terms of genetic distances in centimorgans (cM), a unit measuring genetic recombination frequency. Recombination is a phenomenon observed during the gamete formation, where parts of genetic material are exchanged between homologous chromosomes. Consequently, the closer two markers are located on a chromosome, the fewer possibilities exist to be separated due to recombination. When the separation possibility of two markers is 1 in 100 (recombination frequency 1%), we consider that the distance between them in 1 cM[21].

With the aid of family trees and linkage maps we can trace the heredity of a disease or some other characteristic within a family. The identification of

specific forms of genetic markers (polymorphic markers) in the affected members of a family and their absence from the non affected members of the same family, may, for a example, indicate the possible locus of the pathogenic gene, even if the molecular basis of the disease and the responsible gene are not known to us. Since this method is based on the identification of the polymorphic markers heredity, the possibility of accidental joint inheritance of the said markers (i.e., marker linkage), is taken into consideration. This can be statistically calculated from the decimal logarithm of the ratio of the pro and con probabilities for the two markers to be in some sort of linkage. This calculation is known as LOD score (logarithm of the odds score); values equal to or greater than 3 are indicative of the linkage between the two markers. These mathematical calculations are now being conducted with the aid of computer statistical software[22]. This method has been successfully applied in the detection of genes responsible for many hereditary diseases, such as Huntington disease[23], hypertrophic cardiomyopathy[24], cystic fibrosis[25] and many others.

HGP's aim was the development of a genetic linkage map, where the presence of a great number of markers in close genetic proximity should facilitate, in an initial phase, the establishment of new pathogenic genes. Such a map was published in 1994, comprising 5840 markers, with an average distance among them of 0.7 cM. From these markers, 427 were genes involved in several diseases, such as cystic fibrosis, myotonic dystrophy and others[26]. Within the next four years, two more genetic maps have been published (Généthon genetic map[27] and Marshfield genetic map[28], which constituted more analytical reference fields for genetic linkage studies), which were later used for the development of higher resolution maps.

Apart from genetic linkage maps, physical maps also exist, the name of which results from the fact that the markers loci are given in terms of physical and not genetic distance. The development of such maps was achieved either with the use of overlapping DNA clones, or with the use of radiation hybrids, where distances are expressed in nucleotide bases (base pairs, bp) or centirays (cR, a unit measuring the amount of x radiation), respectively. So long as the bp-cR correspondence is known, both types of maps constitute physical maps[22, 29].

High resolution physical maps provide information related to adjacent, overlapping and arrayed cloned DNA fragments, representing a part of a chromosome or a complete chromosome[30]. With the aid of such maps, on which genetic markers are indicate the region where the gene which interests us is located, the region containing the said gene can be isolated through cloned fragments that represent the particular region and are available.

Due to the fact that mapping and definition of the nucleotide sequence of the human genome are being conducted in large scale and with the participation of many research organizations, a common language for the interpretation and classification of the results was needed. This need was satisfied with the identification of genetic markers with the name sequence-tagged sites (STSs). STSs are unique sequences of a particular location with a length of 200 to 300 bases, which can be "drawn up" from the whole of the genome with the use of a form of the polymerase chain reaction technique (electronic polymerase chain reaction, ePCR)[31]. Given the fact that oligonucleotide primers, necessary for PCR implementation, are unique for each STS and are published in databases on the Internet[32], each research team is able to use particular STSs in order to classify in a row a number of cloned DNA segments, which shall entail common STSs sequences. In this way, gradual re-composition of a chromosome, or a part of it, is possible[22].

Beyond STSs, another group of sequences, the ESTs (expressed sequence tags), have acquired particular importance. ESTs are genetic markers, which however are expressed and constitute part of a gene. Their identification and placement on a physical map provide us with information regarding the locus of the candidate genes in the particular genetic region. In 1998, one of the most important objectives of the HGP was achieved through the publication of a physical map, which contained 41,664 STSs, 30,181 of which were ESTs sequences and represented unique genes[33]. Moreover, following the recent announcement of the human genome research results by the public HGP, the most complete - to date - physical map of the genome has also been published[34,35], and, although, it entails some shortages, it covers 96% of the euchromatic region (region entailing active genes) of the human chromosomes[36].

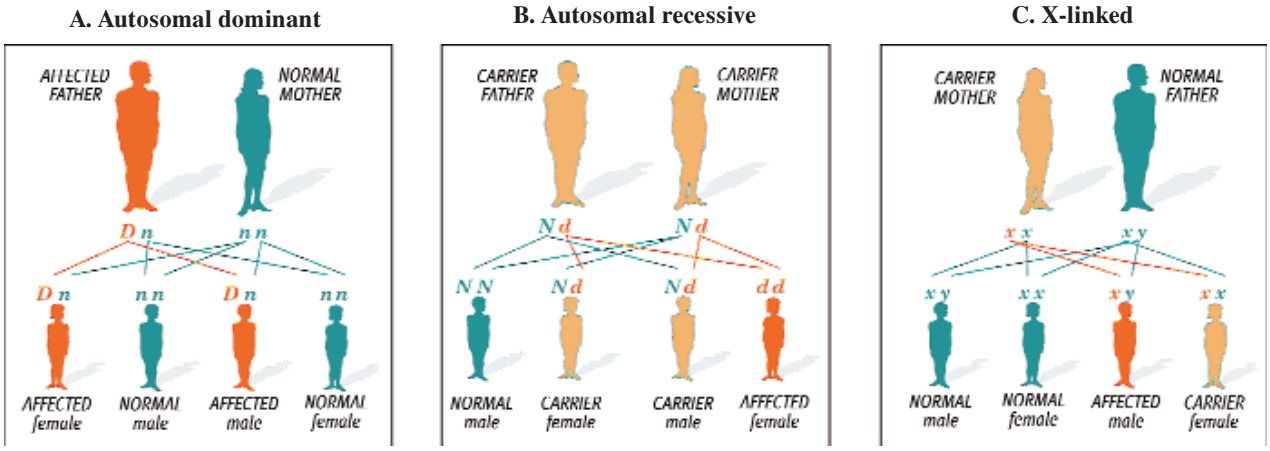
Consequences of genome mapping

The higher resolution human genome physical map contains the complete nucleotide sequence of the

genome. The lower resolution maps, developed in between, acquire special importance and are used for important applications, since they play an ancillary role in the development of the final map and constitute reference fields for the isolation of new genes responsible for several diseases.

Genetic diseases, for which a particular gene is responsible, show simple mendelian inheritance of a phenotype and are divided in autosomal dominant (for example hypertrophic cardiomyopathy), autosomal recessive (e.g. Naxos disease) and sex-linked diseases (e.g. types of familial dilated cardiomyopathy)[37] (Figure 2). For many recessive diseases, we know the respective defective protein and, consequently, we have the ability to intervene and control the course of the disease, because we know its biochemical basis. For example, administration of blood coagulation agent VIII to patients suffering from classical hemophilia or, simply, dietary avoidance of phenylalanine and other substances metabolized by it, in patients suffering from phenylketonuria. In cases where the protein is known while the respective gene is unknown, this very fact can provide us with information in order to recognize and isolate the gene, by utilizing the knowledge of protein sequencing, which indicates and refers us to gene sequencing[38]. As an example we could mention the fibrillin gene. Fibrillin is a glycoprotein constituting a component of the micro-fibrils that are contained in abundance in the connective tissue[39]. Gene isolation has been achieved on the knowledge of the protein sequence of fibrillin, which has been utilized for designing special oligonucleotide combinations (degenerate primers) for the application of PCR[40].

Unfortunately, in most diseases, we do not possess data related to their molecular basis. In such cases, we use other methods for the identification of the responsible genes. One of these methods is positional cloning, as well as positional candidate cloning[41]. Positional cloning is a technique permitting the confirmation of the genetic basis of a disease and the detection of the responsible gene, even when the function of the said gene is not known. Initially, this technique uses data originating from genetic linkage studies, which, as previously mentioned, indicate the region where the gene we are looking for may be located. As soon as this region is identified, its "exploration" begins (for example, using chromosome walking/jumping techniques) and, finally, the gene is identified with the aid of several methods (i.e. exon trapping, zoo blots) or by "reading" coder sequences, i.e. expressed sequences, which are identified by the presence of several characteristics (intron-exon boundaries, splice donor/acceptor sites, CpG islands, trans-



The above graphic has been created by the U.S. Department of Energy, Human Genome Project (http://www.ornl.gov/hgmis.html).

Figure 2. Mendelian inheritance.
A. Individuals that carry the mutant gene (***D***) are affected. Their offspring have 50% chance of inheriting the mutant gene and be affected, irrespective of sex.
B. Individuals that carry one copy of the mutant gene (***d***) are carriers of the disease (heterozygotes), whilst those that carry two copies (homozygotes) are affected.
The offspring of two individuals that are both carriers have 25% chance of being healthy, 50% chance of inheriting one copy of the mutant gene and be carriers, and 25% chance of inheriting both copies and be affected, irrespective of sex.
C. Female individuals that carry one copy of the mutant gene (***x***) are carriers of the disease. Their offspring have 50% chance of inheriting the mutant gene. Their female offspring will be carriers of the disease, whilst their male offspring will be affected.
(Affected individuals are shown in red, carriers in brown, and healthy individuals in blue).

lation initiation/termination codons, open-reading frames)[21].

A variation of positional cloning is positional candidate cloning. Both methods are based on and use data extracted from genetic linkage studies in the same way; the only difference is that in positional candidate cloning the final recognition of the gene is facilitated by a general idea as to which genes could be considered as candidates for being responsible for the disease under inquiry. The principle governing this method is that the candidate gene shall be situated in the same genetic region indicated by the linkage studies as the possible region of the gene we are looking for[22].

In the field of cardiology this method has been applied with particular success, like for example in the implication of the plakoglobin gene for the autosomal residual form of the ARVC disease (Naxos disease). This form of the said disease shows typical cardiological ARVC findings, albeit accompanied by characteristics of diffused nonepidermolytic palmoplantar keratoderma (NEPPK) and woolly hair (WH)[42]. Due to the fact that several types of NEPPK (diffused, focal and striated) are found to be related with gene mutations of the keratin family, these genes have been examined as candidates responsible also for the Naxos disease[43]. Moreover, chromosome regions 14q23-24, 1q42 and 14q12-22, that are found to be involved in the autosomal dominant form of the ARVC disease, have also been examined as responsible candidates.

The construction and utilization of family trees of the island's inhabitants indicated, initially, that the disease exhibited an autosomal recessive heredity. Later on, particular genetic markers known (from genetic maps available on Internet databases) to be connected with candidate genes and selected candidate chromosome regions were selected. These markers were used for analytical linkage studies. The studies provided statistically important indications (LOD score=3.62) for the linkage of the responsible hereditary factor of the disease with marker KRT9 (part of keratin 9 gene) in a distance of 7 cM in chromosome region 17q21.

Two years later, further linkage studies limited the area in a distance of 0.7 cM, which has been found to contain, inter alia, the gene of plakoglobin[44], which is a protein expressed in desmosomal cellular ligaments. Studies on mice with mutations of the plakoglobin gene showed that these mice exhibited cardiological and dermatological anomalies[45] similar with those of the Naxos disease. This information has been utilized in the establishment of the said gene as responsible candidate for the disease. The definition of the nucleotide sequence of a gene of a patient suffering from the Naxos disease, showed the existence of a mutation related to the obliteration of two bases of the plakoglobin gene.

The positional candidate cloning strategy has also been proven successful in the implication of the fibrillin gene in the Marfan syndrome[40], one of the most common genetic diseases of the connective tissue. Immunohistochemical studies showed that this protein, which is normally abundant in the connective tissue, is drastically reduced in the connective tissue of patients with Marfan syndrome. Studies that started with the isolation of this protein and the definition of its sequence, resulted to the isolation of the gene, which was located in chromosome region 15q15-21, a region implicated for Marfan syndrome by genetic linkage studies[46]. The definition of the nucleotide sequence of the gene led to the identification of a missense mutation of the gene in patients with Marfan syndrome[47].

The first gene ever isolated by the positional cloning method was the gene responsible for cystic fibrosis, by J.M. Rommens et al. in 1989[25]. Cystic fibrosis is one of the most common genetic diseases of the white race with autosomal recessive character. Based on the know-how and the research data available in late '80s, it took four years to isolate the particular gene.

In the 80's, although the relevant methodology was known, there were not enough tools available for such kind of studies. The gradual accumulation and recording of genetic markers, through the development of genetic and physical maps of an increasingly higher density and resolution, within the framework of HGP, led to an increase in the identification and isolation of new genes. It is worth noting that after the publication of the analytical genetic map in 1996[27], 16 reports were published within the following year related to the isolation of new genes with the aid of the positional cloning method[33]. In total, more than 30 genes involved in several diseases, like breast cancer, achromatopsy and others, have been isolated and identified within the framework of researches based on the on-going HGP research[3].

From macrocosm to microcosm

Perhaps, the most important contribution of biology to the new perception of medicine is the framework of disease classification and understanding on a

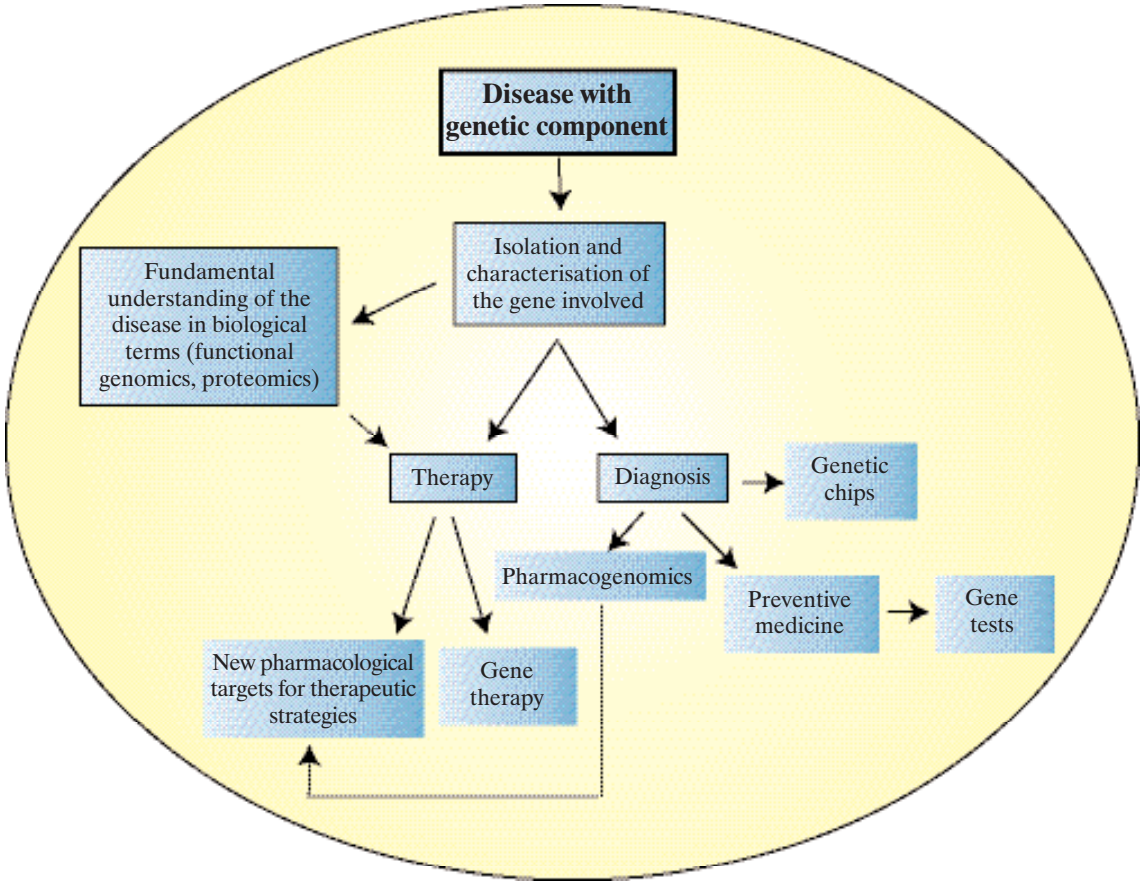


Figure 3. Genetic diagnosis, prevention and therapy.

biological basis. The genetic information that is accumulated at an amazing pace, will gradually stress the need to redefine diseases, based on the biochemical processes that compose them and not on the phenotype they present[48]. In the long course towards a profound understanding and treatment of diseases, the isolation and characterization of the involved genes may not constitute a panacea, but it certainly is a starting point (Figure 3).

Studies conducted in relation to Parkinson's disease led to the identification of some of the responsible genes[19]. Among these, there was a gene that encodes α-synuclein, a presynaptic neural protein[49]. On the occasion of the identification of this gene, further research was incited that, finally, led to the identification of another responsible gene, which is involved in the proteolytic decomposition of α-synuclein and is also responsible for the disease. Given the fact that a-synuclein is a protein found to accumulate in brain cells of patients with Parkinson's disease, and Alzheimer's disease, it is concluded that the characterization of just this gene gives new perspective to the research of the cerebral proteolytic processes related to these diseases and also to other neurodegenerating diseases, like Huntington disease and spinocerebellar ataxia[19]. Conclusively, if every disease can be compared to a big puzzle, every discovery constitutes a piece contributing to its completion, providing an increasingly explicit image.

From genetics to genomics

In the course of the research conducted for the characterization of the a-synuclein gene it has been found that in the families under study, the cause of the disease was a mutation of the gene, resulting in

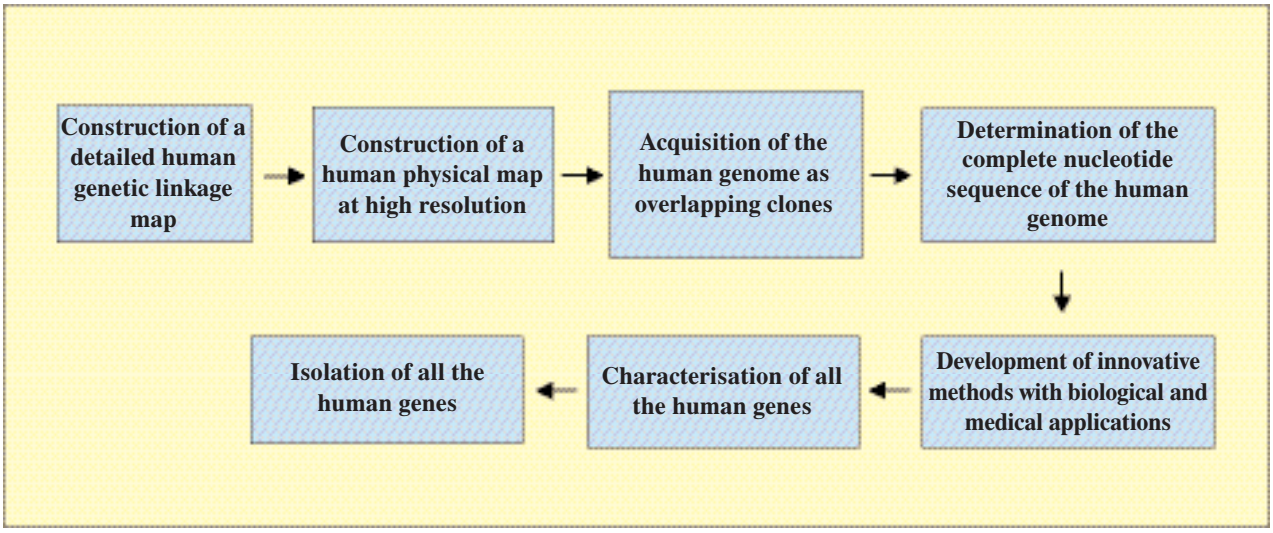


Figure 4. The goals of the human genome project.

the substitution of the amino acid alanine by threonine at position 53 of the protein sequence (Ala53Thr). It was then considered that this substitution at the particular position of the protein chain disturbs its structure in such a way that it enhances its incorporation, a phenomenon considered as characteristic of the amyloid structures present in these diseases.

The main reason for the importance of a gene isolation is due to the fact that a gene leads us to the correspondent protein; and a "defective" gene can lead us to a "defective" protein. The nature of the defect may suggest why this protein is not functional. And if we gain access to this shortcoming, we shall then be in the stage of genetic therapy[49].

The basic doctrine of biology that describes the flow of genetic information, is the replication, transcription and expression of the genetic material. DNA is replicated, transcribed to RNA, and RNA is translated into proteins. Following scientific developments of recent years, we tend to adopt a new triptych as the central axis of biological research; sequence implies structure and structure implies function (the sequence-structure-function paradigm)[7]. A new genetic field, functional genomics, arose for the organization of information flow from sequence to function, having as ultimate aim to utilize function research as a means leading to therapeutic strategies[50].

The term "function", in relation to protein action, refers to a variety of biological fields. Thus, regarding a protein with enzymatic action, biochemistry is interested in its substratum, genetics in its encoding gene, structural biology in its stereochemical layout and physiology in the affected tissues and organs. The sum of this information contributes to the understanding of its function[50]. Due to the fact that proteins direct life mechanisms in every living creature, the identification of all proteins and protein complexes related to biological processes is of critical importance to the understanding of human biology.

The new field of proteomics, considered as an extension of functional genomics, deals with the study of protein expression patterns in biological systems, since every cell type expresses a different set of proteins[51]. Proteomics promises to contribute to the understanding of diseases, by establishing a framework in which the comparison of protein expression of healthy and pathogenic tissues (for example, absence of some proteins or detection of a drastic decrease / increase in their levels in pathogenic tissues) shall direct the research related to the causes of pathogenesis[52]. This is already applied in breast cancer, through the comparison of expression patterns of normal and neoplastic mammary cells[53].

The human genome genes

HGP's aim in this long course for the exploration of the human genome was to define, through successive steps, all its genes, a knowledge that will lead to the development of methods which could have many biological and medical applications (Figure 4). Since the completion of the first mapping (the working draft), a new race begun for the definition of genes, the number of which, at that time, was estimated bet-

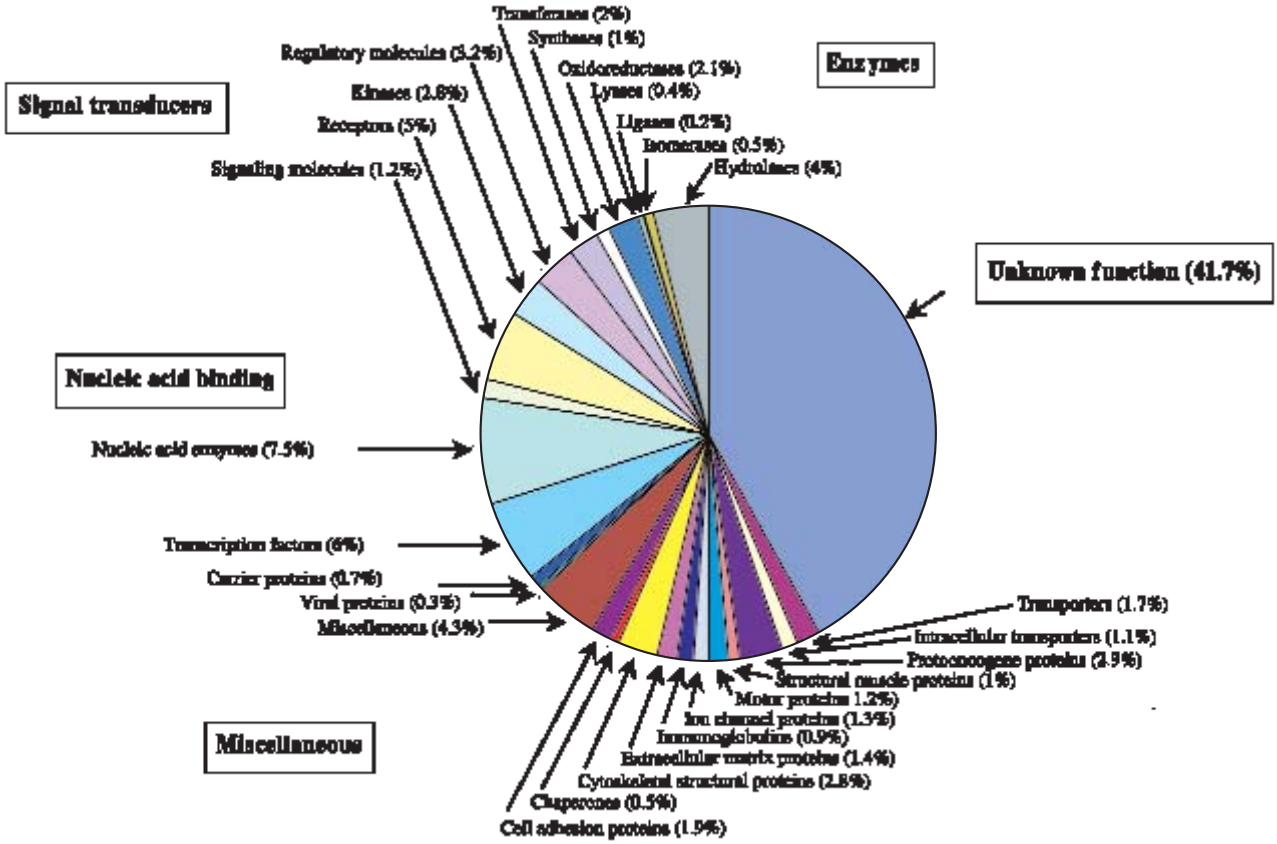


Figure 5. Distribution of the molecular functions of human proteins.

ween 35,000 and 150,000[54]. Recent announcements of Celera Genomics and public HGP estimate that the number of genes varies between 26,000 and 35,000[4]. Out of the total human genes, 10,000 have been fully characterized to date[55] and they are published on the Internet OMIM (Online Mendelian Inheritance in Man) database[56].

Some of the first estimations calculate that 40[4] to 60[3] per cent (40-60%) of our total genes, after their transcription to RNA and before their translation into proteins, follow a process of alternative splicing, a phenomenon we knew it was happening but we didn't know its extent. That means that a gene can produce different forms of mRNA (messenger RNA, the RNA that is being translated into protein), and this complexity of the gene expression implies a variety of proteins.

But how can we count the number of our genes? The sequences we meet in our genetic map are (i) already characterized genes, (ii) segments that apply to ESTs sequences, and we can thus recognize the respective genes, (iii) regions with as yet unknown genes, (iv) regions which are not expressed. The sum total of the last two categories supersedes to a great extent the two first.

The *de novo* characterization of genes is based, to a great extent, on homology rules. A gene can be identified through the relation (homology) it may have with the respective gene of another species[55]. This results from mathematical calculations carried out by powerful electronic computers and from searches in databases containing nucleic and protein sequences, structures and protein patterns. Such an analysis has been conducted by Celera Genomics[4], in order to form a first estimate of the distribution of the molecular function of the approximately 26,000 human genes, according to the calculations of the said organization. The analysis is presented in Figure 5 (Data are from Venter J.C. et al: The sequence of the Human Genome. Science 2001; 291: 1304-1351).

Comparative genomics

In our effort to understand human biology, we have to ask ourselves: which are the particular and exclusive characteristics of our species, i.e., the ones that differentiate us from every other species. Some-

Table 1.

Species whose genome has been sequenced	Year	Importance
Bacterium *Mycoplasma genitalium*	1995	The smallest bacterium to be identified. It represents a model of the minimal gene complement required for an independent existence.
Yeast *Saccharomyces cerevisiae*	1996	The first eukaryote whose genome was sequenced.
Bacterium *Escherichia coli*	1997	The most widely used prokaryote in world-wide research.
Nematode worm *Caenorhabditis elegans*	1998	The first multicellular eukaryote whose genome was sequenced.
Fly *Drosophila melanogaster*	03/2000	The largest organism whose genome had been sequenced at the time.
Human *Homo sapiens*	06/2000	The most complex life form.
Mouse *Mus musculus*	(pending)	Research model of human disease due to its significant genetic homology with humans.

thing that becomes increasingly obvious in our exploration of our genetic "universe" is that our differences become evident through our similarities, certifying in this way the indisputable unity of life.

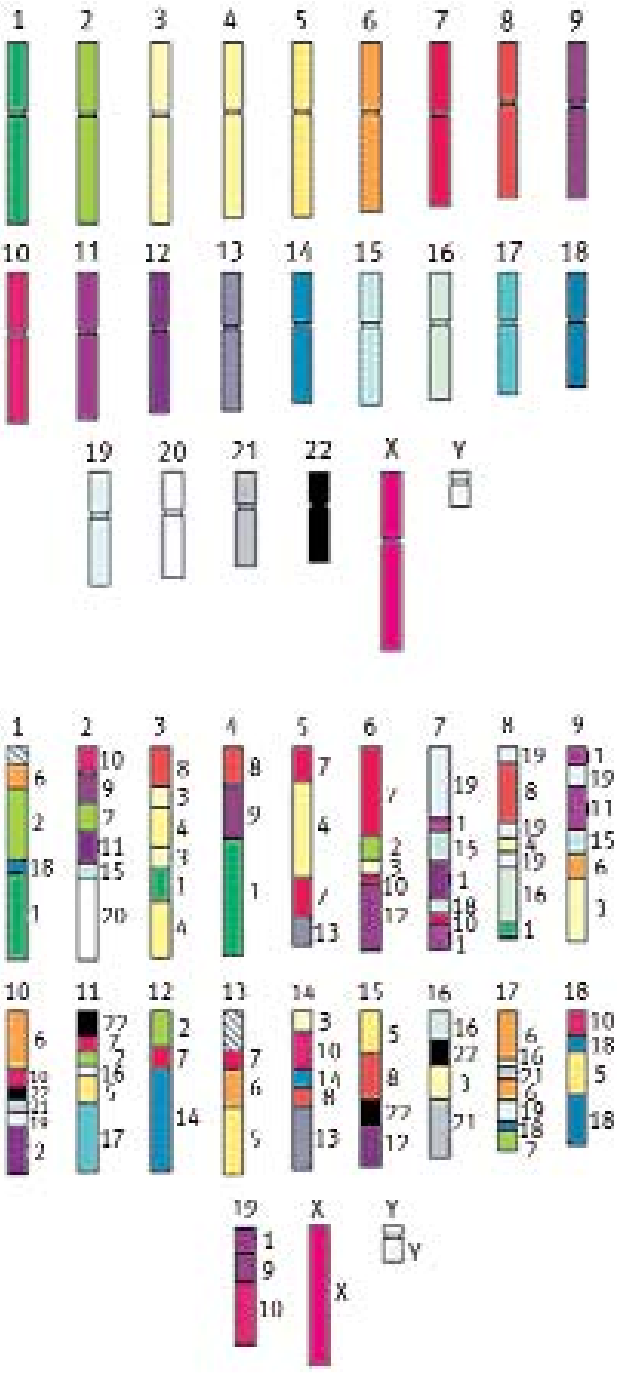
The utilization of other species as models for a more detailed study of the human kind is not a new idea and, for that reason, during the recent years, part of the scientific research has been dedicated to decoding the genome of several organisms-models, like the mycoplasma *M. genitalium*[57], the bacterium *E. coli*[58], the nematode *C. elegans*[59] and the fly *D. melanogaster*[60] (Table 1). As a result, we have extracted a great number of genetic information related to normal gene expression and regulation, genetic diseases, and to our evolutionary course from the decoding and analysis of these organisms genomes[61], in the context of the newly constituted faculty of comparative genomics.

The basis of comparative genomics is the doctrine of evolutionary biology stating that all living beings on earth originate from a common ancestor and, consequently, share the same basic set of genes. The different species resulted from a process of vast genetic expansion, and the new genes that arise, and the proteins that they encode, emanate from older genes which have been duplicated and modified (gene duplication/modification). Due to the fact that these phenomena develop gradually over time, the kinship of these genes can be traced through the similarities that have been preserved in their sequences[62]. These genes are defined as "orthologs" (gene orthologs) or homologous, and compose the so-called gene families. The same is valid for the respective proteins (protein orthologs, protein families).

Before the completion of the human genome decoding, comparative studies were performed on the genome of *Drosophila melanogaster* and 289 characterized human genes, mutations of which have been implicated for several human genetic diseases[63]. From these genes, 112 (39%) seem to have no homologous (ortholog) genes in *Drosophila*. Some of these are genes of a cancerous predisposition, genes with proteins that participate in biochemical paths related to insulin, thyroidal hormones and others. The absence of these genes reflects the differences in the biology and physiology of the two organisms.

With the human genome sequence at our disposal, we shall have the possibility to exploit such differences and similarities. For example, a known gene in *Drosophila*, with known correspondent protein and function, shall be used as "matrix" for the detection of the homologous gene and the homologous protein in humans. And then, the research related to the importance and contribution of that genetic information in human biology will be conducted faster and easier.

In spite of that, and in order to define the unique human physiological, anatomical and other differentia, we need comparative data with closer related species. The mouse genome decoding, that is under development, seems be of critical importance to the understanding of the human genes function[64]. The



The above graphic has been created by the U.S. Department of Energy, Human Genome Project (http://www.ornl.gov/hgmis.html).

Figure 6. Color representation of the genetic homology of human and mouse chromosomes.
Color representation of the genetic homology of human and mouse chromosomes. The degree of genetic similarity (homology) between the two species is estimated to be approximately 85%. Parts of the 23 pairs of the human chromosomes (each shown in a different color) have homologous segments in the 20 pairs of mouse chromosomes. The coloring and numbering alongside the mouse chromosomes indicate the homologous segments of the human chromosomes. The significant genetic similarity between the two species makes the mouse a powerful research tool for the understanding of human genetics.

mouse is quite "familiar" to us, due to the "preference" we have developed for it as a "model" during the last 15 years, in the frame of human genetics research. We thus know that there is a percentage of homology, at a genetic level, in the order of 85% between the two species[65] (Figure 6), a fact that we have already successfully exploited in recognizing the genetic basis of some human diseases, like in the case of the dermal anomaly known as piebaldism[20].

Moreover, a very promising program concerns the chimpanzee genome[66]. It is believed that the study of primates, with which we share the most similar evolutionary course, shall provide data concerning our more subtle differences with them. Many diseases common in both species, like AIDS, Alzheimer's disease and cancer, occur, however, in different frequency and clinical severity. Another relevant section concerns similarities and differences in the reproductive biology of the two species and, finally, the differences of perception. These comparative data may suggest the genetic factors that affect perception and behavior. Such knowledge on the biological functions of the brain may helps us differentiate, at a molecular level, normal from disordered cerebral functions[67].

HGP consequences in multifactorial diseases

Genetic research is often faced with the fact that some of the most common and severe diseases, like cardiovascular diseases, most cancer types, diabetes and psychiatric diseases, do not follow the laws of simple mendelian inheritance. On the contrary, they develop as a result of the interaction of complex hereditary predispositional factors (complex susceptibility traits) with the environment. Thus, the correlation of genotype-phenotype is not clear, since a particular genotype may result to different phenotypes (due to the interaction of many genes and the environmental influences) or different genotypes may develop the same phenotype (like, for example, in the case where different genes encode proteins participating in common biological cycles)[19,41].

The study of such multifactorial diseases at a genetic level with the aid of linkage methods has not been especially successful[41]. For that reason, their study is based to a great extent on population genetic studies of several species (i.e. genetic association studies)[68]. These studies, instead of investigating familial heredity patterns of a disease (as linkage studies do), they usually compare heredity

patterns between healthy and affected subjects, without any kind of kinship, from a statistically meaningful population sample. Such studies have been proven of critical importance for the implication of the angiotensin converting enzyme *(ACE)* gene in myocardial infarction[69,70].

The reason for which the individuality of each person is indisputable is that each one of us is a product of its genes, of the genetic variety that these genes present and of his/her life experiences that have influenced him/her, i.e. of his/her environment. At a genetic level, variety is the factor contributing to our particular phenotype differences, be they favorable (e.g. special dexterities), or not (e.g. predisposition for a disease)[71].

At a molecular level, the greater part of the human genome variety is attributed to a polymorphism known as SNPs (single nucleotide polymorphisms). SNPs are mononuleotide polymorphisms, encountered on average in every 1,000 to 2,000 bases within the human genome and are present both within genes and in genome regions that are not expressed[72]. Despite the fact that the first class concerns a relatively small number (<1%), it is estimated that these polymorphisms contribute to the heterogeneity of thousands of human proteins[4].

The human genome variety analysis constitutes the beginning for the clarification of the genetic basis according to which humans react and respond in a different or individual way to different diseases, environmental factors, drugs and other. The impending genetic studies of multifactorial diseases will refer to the mapping of a great number of SNPs within the total of the human genome, while announcements are already made concerning the methods of SNPs use in the clarification of the genetic basis of multifactorial diseases[73]. This has been achieved, for the most part, within the HGP framework, through the mapping and recognition of 1.42 million SNPs[72]. The sequences have been published and are constantly reviewed in the dbSNP database[74] of the National Center for Biotechnology Information (NCBI)[75], on the Internet.

Diagnosis, prevention and therapeutic strategies

Genetic diagnosis

Our knowledge on gene functions shall increase as research for genetic causes underlying various diseases continues. An important application of genetic research has a diagnostic value and relates to the development of gene tests[76]. Gene tests may be used for several reasons, such as:

- Confirmatory diagnosis for a disease of a symptomatic individual
- Carrier screening for individuals with a disease family history
- Presymptomatic diagnosis
- Prenatal examination and examination of neonates with or without family history
- Preimplantational diagnosis
- Identity/paternity examination

The use of gene tests provides important benefits in diagnosing predisposition of a disease, for which we can take preventive measures, like for example adenomatous polyposis. Mutations of the *APC* (adenomatous polyposis coli) gene have been found to constitute a predisposition factor for the occurrence of the disease[19]. Thus, an individual who knows of this predisposition, may take regular checkups, so that in case adenomatous polyps are detected, they shall be removed before they develop into malignant tumors.

In hypertrophic cardiomyopathy, for which a great number of mutations of several genes has been implicated, some mutations of the cardiac troponin T gene seem to relate to drastically reduced life time and higher frequency of sudden deaths in ages below 30 years[77]. Clinical test on relatives of patients with hypertrophic cardiomyopathy that were carriers of the gene mutations, have demonstrated that many of the said relatives, especially in younger ages, are carriers of these mutations, but clinically asymptomatic. Considering the risk of these patients, genetic diagnosis is necessary, so that these individuals have a 6-month follow-up for symptom manifestation, be encouraged to change their life style (e.g. avoidance of sports), receive preventive antiarrhythmic treatment, or even a defibrillator.

Despite the fact that the number of gene tests that are developing and are gradually becoming commercially available is constantly increasing[76], their use in certain cases involves some risk, because society is not yet accustomed to such practices, neither is it adequately prepared to confront some of the social consequences. As a characteristic example, we could mention Alzheimer's disease, the predisposition of which implicate three isomorphs of apolipoprotein E (*ε2, ε3, ε4*), each of which seems to relate to a low, middle and high predisposition, respectively, for disease occurrence.

Although an appropriate gene test has been developed, the results interpretation is difficult, as it has

been established that many individuals carrying pathogenic genes never exhibit this disease (incomplete penetrance)[41]. Thus, ethical issues arise to the interpretation as well as to the announcement of the relevant results. In addition, the use of gene tests in diseases where prevention and application of therapeutic strategies are not feasible, is under dispute.

A new developing technology, which will contribute dynamically in the field of genetic diagnostics, is gene chip (DNA chips) technology. Gene chips are constructed through the application of techniques similar to those of electronic microchips. Their size is approximately that of a postage stamp and they consist of a small piece of glass on which synthetic DNA sequences (synthetic oligonucleotide arrays) are arrayed. With a small blood sample and a gene chip we can "photograph" our genetic material and detect normal and mutated sequences. This method is expected to prove an important tool in the detection of large gene mutations, for which a great number of different mutations has been reported, like breast cancer predisposition genes *BRCA1* and *BRCA2*[78]. Gene chips are already constructed by several biotechnological companies, like Affymetrix in Silicon Valley, California, and although they are currently being used only at a research level, it is expected to be widely used in the health services sector within the next years[79].

Genetics and cardiology. Molecular diagnosis in clinical practice

During the last years, significant progress has been made in the definition of the molecular basis of many hereditary cardiovascular diseases. The development in the field of molecular biology and genetics has drastically helped in understanding the primary factors and mechanisms responsible for the pathogenesis and expression of these diseases. However, the identification of new groups of "patients" with the aid of molecular diagnostic methods rather than classic clinical diagnostic criteria, led to questions related to the role of genetic control in clinical diagnosis and to its consequences in patient management[80].

When is genetic control feasible? Firstly, when the gene, the mutations of which are responsible for the disease, is known and, secondly, when the disease is monogenic. In which cases can genetic control support clinical diagnosis and patient management? Firstly, when diagnosis is difficult due to "reduced penetrance" and to the "complexity of the clinical expression" and, secondly, when clinical expression varies according to genetic substratum, while the latter directly influences the choice of the therapeutic strategies[81].

The capabilities of genetic diagnosis enforces the role of genetic control in clinical practice. Genetic control can be applied in pre-clinical diagnosis and/or detection of high risk patients, who shall benefit from preventive treatment for sudden death. Such an approach must, of course, satisfy some prerequisites, such as[82]:

i) Availability of appropriate means for performing particular genetic control.

ii) Availability of a sufficiently large database with clinical and genetic information, which shall define the relations of genotype-phenotype, i.e. to correlate the genetic substratum with the clinical expression in order to reach conclusions.

iii) Assessment of the efficacy of the available therapeutic treatment.

iv) Distinct cost-benefit ratio.

In which way does the above apply to hereditary heart diseases? Take for example, the long-QT syndrome (LQTS). LQTS is a cardiovascular disease characterized by a prolonged QT interval in ECG and exhibiting syncope and sudden death. So far, five genes have been identified; their mutations are responsible for LQTS and these genes are being inherited in an autosomal dominant way[83]. The disease exhibits drastically reduced penetrance and varied clinical expression. This has been established, initially, through data according to which 6% of the relatives of patients suffering from LQTS, while having normal QT interval, showed syncope or cardiac death[84]. In addition, it is known that in families with LQTS, parents without LQTS phenotype may have intensely symptomatic descendants, while symptomatic parents may have descendants-carriers of a muted form of the disease. The above support the notion that LQTS diagnosis can no longer be based on exclusively clinical diagnostic criteria[85].

In which cases is genetic control recommended to patients with LQTS? There seem to be three scenarios[82]. The first relates to patients with a clear clinical diagnosis. In that case genetic control is not necessary, but may be useful in order to determine a more specific treatment. More and more data support the notion that the precise clinical expression of LQTS relates to the particular gene, amongst the

genes involved in the syndrome, which is responsible for the disease in that particular patient[86]. The second case concerns patients with suspected LQTS diagnosis or patients with marginal diagnosis based on clinical criteria. In these cases, genetic control is useful in order to confirm LQTS diagnosis. The third case concerns asymptomatic relatives of patients with LQTS. Once we know the genetic substratum of the patient, genetic control may reject or indicate the existence of a pathogenic genetic substratum in his/hers relatives. Similar scenarios are valid for other diseases, like hypertrophic cardiomyopathy.

The basic aim in the research of hereditary cardiovascular diseases like LQTS and hypertrophic cardiomyopathy, was, and still is, the definition of the genotype of a great number of patients and, through clinical-genetic correlations, the development of a database, so that the identification of the genetic substratum would have prognostic value. The achievement of this aim is still under development. This necessary "embracing" of genetics and cardiology may at times clarify or complicate this aim, but it is certain that medicine has taken up the challenge to analyze genetic diseases into their "components" and proceed to the task of a healthy and controlled synthesis. This is the challenge of the new century.

Pharmacogenetics

Pharmacogenetics, known also as pharmacogenomics, is a new field which combines pharmacology and genetics, studying the way in which the genetic profile of an individual defines and influences his/her reaction to drugs. Pharmacogenetics promises that this knowledge will be used in medicine for selecting the best treatment among the available ones, and for designing individualized drugs, which will be more effective and safe since they will be based on the genetic profile of each individual[87]. SNPs, which as already mentioned are responsible for a great part of the genetic variety of individuals, play a leading role in that new approach.

Clinical studies conducted in patients with coronary atherosclerosis showed that the administration of pravastatin slowed down the course of the disease in some patients[88]. These patients were found to have a polymorphism (B1 variation) in the gene of CETP protein (cholesteryl ester transfer protein), a protein with enzymatic action in the metabolism of HDL lipoprotein. On the contrary, those who were not carriers of this polymorphism have not benefited from the administration of that drug. B1 variation of that gene seems to relate to increased CETP concentrations and decreased concentrations of HDL lipoprotein in blood plasma. Conclusively, this observation may be useful for assessing the patients with coronary disease that may benefit from the administration of pravastatin.

The idea of individualized medicine is not new, but it is becoming increasingly feasible thanks to recent innovations. By incorporating pharmacogenetics in clinical practice, physicians will be able to provide treatment according to the genetic profile of the patient, thus knowing his/hers reaction to appropriate drug groups, in the same way drug dosage is currently defined according to age and physical condition of the patient[89].

Ethical, legal and social consequences of HGP

Human genome mapping constitutes a very promising scientific achievement, which, like every other achievement, invention or discovery in the history of science, may not be compatible with society and its law. This is quite normal, since discoveries are *de facto* ahead of their social applications and laws which are introduced to define their proper usage. Fortunately, the social, ethical and legal issues that arise from the research of the human genome have been made evident quite early, and their investigation within the framework of HGP was timely performed. Thus, during the past 8 years, 5% of the financial budget of the public HGP has been dedicated to the ELSI program (Ethical, Legal and Social Issues program), which deals with these issues, guiding US government to taking necessary measures[90].

To date, the ELSI program has indicated several issues concerning gene research[6], such as:

- protection of personal genetic data secrecy of each individual in work, schools, insurance companies,
- use of genetic information as evidence in law cases,
- importance of genetics advisory services in clinical practice and family planning,
- shaping of social consciousness in genetic matters,
- commercialization and copyright of results.

One of the most urgent questions addressed by the ELSI program is who has the right to know what and for whom. With the term secrecy we define the right of a person not to disclose some of his/her

sensitive personal data. Unfortunately, the protection of secrecy of genetic data may prove difficult, as conflict of interest may quite often arise e.g. interests of insured persons and insurance companies. Therefore, as of 1999, more than 16 states in the USA have passed laws prohibiting insurance companies to use genetic data in order to deny insurance coverage or to increase premiums to the genetically "less favored"[91].

Since the commencement of the HGP, fears have been expressed concerning the manifestation of social racism, which shall discriminate against those carrying "bad" genes, like for example, genes related to a strong predisposition for some serious disease. It is quite certain that in the future, as genetic factors involved in human diseases will be traced, it will become increasingly evident that each one of us has a larger or smaller share of genetic imperfections, making the distinction between "good" and "bad" genotype unsubstantial[92].

The concept of "bad" genotype has already been proved popular in US courtrooms, where some defendants, in defending themselves, are trying to raise doubts as to their responsibility in illegal acts, by professing a genetic predisposition for violent behavior[91]. Thus, an effort to inform and educate the judicial corps has begun, in order to get familiar with matters of genetics, which, as it seems, in the following years will be very often presented in the courtrooms and will be much more obscure than today's forensic findings[6].

The knowledge of genetic information may have a proper or not application. In the field of genetic diagnosis, as already mentioned, knowledge of a predisposition may guarantee better diagnosis, taking of preventive measures, like regular check-ups, change in life style, and timely treatment. But there is also another side to this coin, in which case there is no prevention, no therapeutic strategy and the announcement of an obscure predisposition for a certain disease has no meaning, other than confining that person to a morbid perception of himself/ herself, with all the psychological consequences that may entail. For this very reason, genetic advisory services acquire a special value as regards the familiarization of people with genetic matters and the provision of guidance towards more correct choices. One good example refers to family planning, where couples with children suffering from or being carriers of a disease, have a chance of limiting or even eradicating the possibility of this phenomenon to happen again in a future pregnancy. In such cases it is quite difficult, but necessary, to explain to them that "accidental events do not possess memory"[23].

There is, therefore, a need to form and shape social consciousness as regards genetic matters, within which people shall possess a basic perception of genetic concepts, and will thus be able to take proper decisions to matters relating to their genetic data, and be able to grant or not access of these data to third parties.

Finally, the issue of highest priority on the list of the ELSI program is the commercialization and copyright of genetic research results. Is there or is there not an issue of copyright as regards results and applications of gene research like, for example, DNA sequences and transgenic animals. The idea that there should be no copyright to something that constitutes part of our "global heritage", and must therefore be freely available to all mankind, has its supporters as well as its adversaries[93].

The definition of nucleotide sequencing of the human genome has been compared to the periodic table of atomic elements[94], which registers the various chemical elements, but indicates nothing as to their behavior in various combinations. This comparison is well-aimed, since the knowledge of nucleotide sequencing of the human genome provides us with adequate means for a faster and more profound understanding of human biology, but at the same time, this knowledge in itself, does not provide a complete understanding of human functions. Today, we are at a stage where a huge step has been achieved, a step necessary but not sufficient for the accomplishment of our target.

Following this rationale, the intention of various organizations to acquire patent rights on DNA sequences, for which their function is not even partially understood, seems to be in conflict with the meaning of patent and copyright laws. In addition, another matter arises: the agencies consolidating patents in similar cases do not rise claims only for what they know at the present time, but for whatever shall be discovered or invented in the future regarding sequences and corresponding proteins[94].

The relevant laws on patent privileges and copyright differ in Europe and the United States of America, but they are governed by the same principle, i.e. patents may be granted only for inventions fulfilling the following three criteria: innovation, inventiveness and usability[95]. That is, the invention has to constitute an innovation and be announced for the first time, it has to be characterized as inventive because it is not evident to the experts of the field it is addressed to, and has to be usable through practical applications

instead of having only an intellectual or aesthetic value.

Just between the years 1981 and 1995, when genome research had not yet acquired its present prestige, 1,175 patents concerning DNA sequencing at an international level had been granted, 7% of which were granted to individuals, 17% to public organizations and 76% to private organizations, percentages that demonstrate the dominance of the private sector[96]. Many of these patents were related to "incomplete inventions" and for this reason the National Health Institutes in USA, in the frame of HGP, are trying to alert the officers in the Patent Office of the Commerce Department to avoid granting patents for simple gene discoveries, if these are not accompanied by a practical application, like e.g. being the basis for the development of some medicinal product[16].

Despite the above, patents and copyright do have a meaning, in the sense that they serve two basic purposes. Firstly, they motivate inventors and their sponsors in investing in the research and, secondly, they encourage the announcement of information, which, under other circumstances, would have remained confidential. Therefore, patents are an integral part of the commercialization process, because they provide exclusivity, a factor that mobilizes the biotechnological industry to bring an invention, e.g. a drug, from the stage of discovery to that of production[93].

Conclusion

The human genome project may be the most ambitious, international, coordinated effort in the history of science. The endless list of A, T, G and C bases bears the genetic secrets that compose human existence, the mystery we are trying so hard to unveil. The present stage constitutes only the beginning of the quest for understanding genetic factors that underline human diseases and compose our senses, memory and perception. HGP does not promise to provide this understanding straightforwardly but to guide us successful during the search for it. We must therefore try to avoid determinism, i.e., the notion that all our characteristics are necessarily of genetic origin, and abstraction, a perception defining that knowledge of genome sequencing implies immediate understanding of our biological functions[4].

In the 21st century, the century of genetics, the human race shall witness the exciting course of this science and shall benefit from its new accomplishments, and their proper use shall aim at upgrading quality of life. However, due to the ethical, social and legal issues that arise from the knowledge and usage of genetic information, some people wonder whether we really want this information[97]. We may be afraid that this knowledge will confirm what we already knew: the "genetic injustice" of the real world.

But, is it knowledge or ignorance that we must be afraid of? Mankind never suffered from scientific knowledge, but from the improper use of that knowledge. Let us hope that HGP, which shall prove an achievement equal to atom fission[14], will not have similar evil applications. A biological Hiroshima would be much worse than the atomic one. And we cannot just hope to avoid this possibility; it is our responsibility to build a world where genetic injustice shall be relieved by social justice.

References

1. Dawkins R: The Selfish gene. Oxford University Press, 1976.
2. Butcher J: "Working draft" of human genome completed. Lancet 2000; 356: 47.
3. The International Human Genome Mapping Consortium: A physical map of the human genome. Nature 2001; 409: 934-940.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton G, et al: The sequence of the human genome. Science 2001; 291: 1304-1351.
5. Human Genome Project milestones celebrated at White House. Human Genome News 2000; 11(1-2): 4-5.
6. Patrinos A, Drell DW: Introducing the Human Genome Project: Its relevance, triumphs, and challenges. The Judges' Journal of the American Bar Association 1997; 36: 3.
7. Smith D: Evolution of a vision: Genome project origins, present and future challenges, and far-reaching benefits (Part I). Human Genome News 1995; 7(3-4): 2.
8. Rodakis G, Margaritis L: Is human genome decoded? Scientific American (greek edition) Volume B′, 2000; 19: 21-24.
9. The Sanger Centre: Human Genome Project, http://www.sanger.ac.uk/HGP.html.
10. Marshall E: Human Genome: A high-stakes gamble on genome sequencing. Science 1999; 284: 1906-1909.
11. Marshall E: Genome sequencing. Claim and counterclaim on the Human Genome. Science 2000; 288: 242-243.
12. Venter JG, Smith HO, Hood L: A new strategy for genome sequencing. Nature 1996; 381: 364-366.
13. Wadman M: Human genome deadline cut by two years. Nature 1998; 395: 207.
14. Marshall E: Human Genome Project: Sequencers endorse plan for a draft in 1 year. Science 1999; 284: 1439-1441.
15. Felsenfeld A, Peterson J, Schloss J, Guyer M: Assessing the quality of the DNA sequence from the Human Genome Project. Genome Res 1999; 9: 1-4.
16. Marshall E: Genome sequencing. Clinton and Blair back rapid release of data. Science 2000; 287: 1903.
17. Butler D: US/UK statement on genome data prompts debate on "free access". Nature 2000; 404: 324-325.

18. Pennisi E: Human Genome. DOE team sequences three chromosomes. Science 2000; 288: 417-419.
19. Collins, FS: Shattuck Lecture: Medical and societal consequences of the Human Genome Project. N Engl J Med 1999; 341: 28-37.
20. Patrinos, A: To know ourselves, a booklet of the U.S. Department of Energy and the Human Genome Project (Vaughan, D. ed.), 1996; p.p. 1-34.
21. Passarge E: Color atlas of genetics. Thieme Medical Publishers, Inc. New York, 1995.
22. Watson JD, Gilman M, Witkowski J, Zoller M: Recombinant DNA. Scientific American, Inc. W.F. Freeman and Company, New York, 1992.
23. Wexler NS: Clairvoyance and caution: Repercussions from the Human Genome Project, in The Code of Codes: Scientific and social issues in the Human Genome Project (Kevles, D.J. & Hood, L. eds), Harvard University Press, p.p. 211-243, 1992.
24. Coonar AS, McKenna WJ: Molecular genetics of familial cardiomyopathies. Adv Genet 1997; 35: 285-324.
25. Rommens JM, Iannuzzi MC, Kerem B-S, Drumm ML, Melmer G, Dean M, et al: Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989; 245: 1059-1065.
26. Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, et al: A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Centre (CHLC). Science 1994; 265: 2049-2054.
27. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A: et al: A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 1996; 380: 152-154.
28. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 1998; 63: 861-869.
29. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, et al: Comparison of human genetic and sequence-based physical maps. Nature 2001; 409: 951-953.
30. Casey D: Mapping and sequencing the Human Genome, in Primer on Molecular Genetics, (Human Genome Project, U.S. Department of Energy, Washington, D.C. http://www.ornl.gov/hgmis/publicat/primer/prim2.html).
31. Schuler GD: Sequence mapping by electronic PCR. Genome Res 1997; 7: 541-550.
32. The Radiation Hybrid Mapping Consortium: Genome-wide radiation hybrid map, "GeneMap '99", http://www.ncbi.nlm.nih.gov/genemap.html.
33. Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, et al: (1998) A physical map of 30,000 human genes. Science 1998; 282: 744-746.
34. The International Human Genome Mapping Consortium: The FPC database of the human genome BAC fingerprint map, 2001. http://genome.wustl.edu/gsc/human/ Mapping/ index.shtml.
35. The International Human Genome Mapping Consortium: The AceDb version of the whole-genome BAC map. 2001, http://genome.wustl.edu/gsc/Search/db.shtml.
36. The International Human Genome Mapping Consortium: A physical map of the human genome. Nature 2001, 409: 934-940.
37. Roberts R, Schwartz K: Myocardial diseases. Circulation 2000; 102: IV-34-IV-39.
38. Antonarakis SE: Diagnosis of genetic disorders at the DNA level. N Engl J Med 1989; 320: 153-163.
39. Sakai LY, Keene DR, Engvall EJ: Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. J Cell Biol 1986; 103: 2499-2509.
40. Lee B, Godfrey M, Vitale E, Hori H, Mattei M, Sarfarazi C, et al: Linkage of Marfan syndrome and a phenotypically related disorder to two fibrillin genes. Nature 1991; 352: 330-334.
41. Lander ES, & Schork NJ: Genetic dissection of complex traits. Science 1994; 265: 2037-2046.
42. Protonotarios N, Tsatsopoulou A, Patsourakos P, Alexopoulos D, Gezerlis P, Simitsis S, et al: Cardiac abnormalities in familial palmoplantar keratosis. Br Heart J 1986; 56: 321-326.
43. Coonar AS, Protonotarios N, Tsatsopoulou A, Needham EWA, Houlston RS, Cliff S, et al: Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21. Circulation 1998; 97: 2049-2058.
44. McKoy G, Protonotarios N, Crosby A, Tsatsopoulou A, Anastasakis A, Coonar A, et al: Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). Lancet 2000; 355: 2119-2124.
45. Bierkamp C, Mclaughlin KJ, Schwarz H, Huber O, Kemler R: Embryonic heart and skin defects in mice lacking plakoglobin. Dev Biol 1996; 180: 780-785.
46. Tsipouras P, Sarfarazi M, Devi A, Weiffenbach B, Boxer M: Marfan syndrome is closely linked to a marker on chromosome 15q1.5→q2.1. Proc Natl Acad Sci USA 1991; 88: 4486-4488.
47. Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, et al: Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 1991, 352: 337-339.
48. Bell J: The new genetics in clinical practice. BMJ 1998; 316: 618-620.
49. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al: Mutation in the α-synuclein gene identified in families with Parkinson's disease. Science 1997; 276: 2045-2047.
50. Vukmirovic OG, Tilghman SM: Exploring genome space. Nature 2000; 405: 820-822.
51. Giometti CS: From sequence to systems: looking at proteins to understand genome expression. Human Genome News 1999; 10(1-2): 18.
52. Gavaghan H: Companies of all sizes are prospecting for proteins. Nature 2000; 404: 684-686.
53. Bradbury J: Proteomics: the next step after genomics? Lancet 2000; 356: 50.
54. Butler D, Smaglik P: Draft data leave geneticists with a mountain still to climb. Nature 2000; 405: 984-985.
55. Wolfsberg TG, Mc Entyre J, Schuler GD: Guide to the draft human genome. Nature 2001; 409: 824-826.
56. Online Mendelian Inheritance in Man (http:// www.ncbi.nlm.nih.gov/OMIM).
57. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, et al: The minimal gene complement of Mycoplasma genitalium. Science 1995; 270: 397-403.
58. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al: The complete sequence of Escherichia coli K-12. Science 1997; 277: 1453-1474.

59. The C. elegans Sequencing Consortium: Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282: 2012-2018.
60. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al: The genome sequence of Drosophila melanogaster. Science 2000; 287:2185-2195.
61. Casey D: Model organism research, in Primer on Molecular Genetics, (Human Genome Project, U.S. Department of Energy, Washington, D.C. 1992; http://www.ornl.gov/hgmis/publicat/primer/prim3.html).
62. Doolittle RF: Microbial genomes opened up. Nature 1997; 392: 339-342.
63. Rubin GM, Yandell MD, Wortman JR, Gabor-Miklos GL, Nelson CR, Hariharan IK, et al: Comparative genomics of the eukaryotes. Science 2000; 287: 2204-2215.
64. Gura T: Changing of the guard. Nature 2000; 405: 389.
65. Public, private sectors join in mouse consortium. Human Genome News 2000; 11(1-2): 12-13.
66. Baltimore D: Our genome unveiled. Nature 2001; 409: 814-816.
67. McKonkey EH, Varki A: A primate genome project deserves high priority. Science 2000; 289: 1295-1296.
68. Collins A, Morton NE: Mapping a disease locus by allelic association. Proc Natl Acad Sci USA 1998; 95: 1741-1745.
69. Cambien F, Poirier O, Lecerf L, Evans A, Cambou J-P, Arveiler D, et al: Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature 1992; 359: 641-644.
70. Schunkert H, Hans H-W, Holmer SR, Stender M, Perz S, Keil U, et al: Association between a deletion polymorphism of the angiotensin-converting-enzyme gene and left ventricular hypertrophy. N Engl J Med 1994; 330: 1634-1638.
71. Little P: The book of genes. Nature 1999; 402: 467-468.
72. The International SNP Map Working Group: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409: 928-933.
73. Zhao LP, Aragaki C, Hsu L, Quiaoit F: Mapping complex traits by single-nucleotide polymorphisms. Am J Hum Genet 1998; 63: 225-240.
74. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al: dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001; 29: 308-311.
75. Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, et al: Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2001; 29: 11-16.
76. Casey D: What can the new gene tests tell us? The Judges' Journal of the American Bar Association 1997; 36: 3.
77. Marian AJ, Roberts R: Molecular genetic basis of hypertrophic cardiomyopathy: Genetic markers for sudden cardiac death. J Cardiovasc Electrophysiol 1998; 9: 88-99.
78. NHGRI in the news: NHGRI fact sheet on DNA chip technology. (http://www.nhgri.nih.gov/NEWS/human_genome_facts.html)
79. Shreeve J: Gene secrets. National Geographic (greek edition), 1999, 3(4): 94-127.
80. Maron BJ, Moller JH, Seidman CE, Vincent GM, Dietz, HC, Moss AJ, et al: Impact of laboratory molecular diagnosis on contemporary diagnostic criteria for genetically transmitted cardiovascular diseases: Hypertrophic cardiomyopathy, Long-QT syndrome, and Marfan syndrome. Circulation1998; 98: 1460-1471.
81. Vincent G.M: Role of DNA testing for diagnosis, management, and genetic screening in long-QT syndrome, hypertrophic cardiomyopathy, and Marfan syndrome. Heart 2001; 86: 12-14.
82. Priori SG, Barhanin J, Hauer RNW, Haverkamp W, Jongsma HJ, Kleber AG, et al: Genetic and molecular basis of cardiac arrhythmias: Impact on clinical management (Parts I and II). Circulation 1999; 99: 518-528.
83. Splawski I, Shen J, Timothy KW, Lehmann MH, Priori S, Robinson JL, et al: Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. Circulation 2000; 102: 1178-1185.
84. Schwartz PJ, Moss AJ, Locati E, Crampton RS, Tzivoni D, Garson A, et al: The long-QT syndrome international prospective registry. J Am Coll Cardiol 1989; 13(suppl A): 20A.
85. Priori SG, Napolitano C, Schwartz PJ: Low penetrance in the long-QT syndrome. Circulation 1999; 99: 529-533.
86. Wilde AAM, Roden DM: Predicting the long-QT genotype from clinical data. Circulation 2000; 102: 2796-2798.
87. Roses AD: Pharmacogenetics and the practice of medicine. Nature 2000; 405: 857-865.
88. Kuivenhoven JA, Jukema JW, Zwinderman AH, de Knijff P, McPherson R, Bruschke AV, et al: The role of a common variant of cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Group. N Engl J Med 1998; 338: 86-93.
89. Drell D, Adamson A: Fast forward to 2020: What to expect in molecular medicine. Human Genome News 2000; 11 (1-2): 13-14.
90. Marshall E: Human Genome Project: Panel urges cloning ethics board. Science 1997; 275: 22.
91. Rothstein MA: Protecting genetic privacy: why it is so hard to do. Human Genome News 1999; 10(1-2): 14-15.
92. Stewart AC: Ethical and social implications of the Human Genome project: the issues in the UK, in Ethical genetics, Science and Public Policy, 1991; 18(2): 123-129, Beech Tree Publishing, Surrey, England.
93. Healy B: On gene patenting. N Engl J Med 1992; 327:664-668.
94. Alberts B, Sir Klug A: The human genome itself must be freely available to all humankind. Nature 2000: 404: 325.
95. Poste G: The case of genomic patenting. Nature 1995; 378: 534-536.
96. Thomas SM, Davies ARW, Birtwistle NJ, Crowther SM, Burke JF: Ownership of the human genome. Nature 1996; 380: 387-388.
97. Müller-Hill B: The shadow of genetic injustice. Nature1993; 362: 491-492.