



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

The Human Genome and Advances in Medicine: Limits and Future Prospects

A. Pardo

Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico.

Introduction

On April 14, 2003, the International Human Genome Sequencing Consortium announced the successful completion of its task. The correct sequence of the bases cytosine (C), thymine (T), adenine (A), and guanine (G) in the gene-containing regions of DNA had been elucidated with an accuracy of 99.99% for 99% of the euchromatin. This is considered to be the most that can be achieved with current technology, and all that now remains is to sequence the remaining regions, which are more difficult because they include almost 400 highly repetitive DNA fragments in addition to the centromeres, the structures that divide chromosomes.

The Consortium of which the Human Genome Project (HGP) formed a part included 20 centers in 6 countries (China, France, Germany, Great Britain, Japan, and the United States of America). This international group chose to announce the completion of the task in April 2003 in order to coincide with the 50th anniversary of the publication, in April 1953, of the paper by Watson and Crick¹ that first described DNA's double helix structure.

The HGP's initial objectives were fulfilled 2 years ahead of schedule, and, in addition to compiling a highly accurate sequence of the human genome which has been made freely available and accessible to everyone, the Consortium has developed a set of new technologies and has constructed genetic maps of the genomes of various organisms. Moreover, this program of scientific investigation is linked to a parallel bioethics program. It is also interesting to note that, thanks to advances in technology, this result was achieved for a cost lower than the initial budget, which estimated that 500 Mb would be sequenced annually at a cost of 0.25 dollars per finished base. The final figure was 1400 Mb per year at a cost of 0.09 dollars per base.

The size and scope of the HGP has also provided valuable lessons about the organization and management

of large projects involving international collaboration, and those lessons will no doubt prove useful in the administration of other large scale projects.

What was the genesis of this project? What general lessons has it taught us so far? How will it influence medicine? What future prospects, hopes, and fears has it given rise to? What ethical problems does it pose? These are just some of the general questions that this article will attempt to analyze.

What is the Human Genome and When Did the Idea of Sequencing It Arise?

The genome is the total set of genes carried by an organism, and each gene is a segment of DNA's double helix structure containing the recipe for making a polypeptide chain in a protein. A protein may contain a single polypeptide chain, as in the case of insulin, and therefore a single gene will code for this protein, or it may contain more than one chain, as in the case of hemoglobin, so that this protein is encoded by more than one gene. There are around 100 billion (US 100 trillion) cells in the human organism, and each one of these contains a complete genome. This genome is found on the 23 pairs of chromosomes in the cell nucleus. Around 1.8 meters of DNA containing approximately 3000 million (3 billion US) base pairs is packed into the nucleus of each cell. The genetic code uses groups of 3 DNA bases to specify the amino acids that make up the polypeptide chains of proteins, the principal actors in life's drama.

One of the first genomes to be completely sequenced was that of simian virus 40 (SV40), which contains 5226 nucleotides.² By the beginning of the 1980s, viral genomes containing over 100 000 bases had been sequenced, making it possible for scientists to envisage the possibility of sequencing bacterial genomes containing over 1 000 000 bases.

When the idea of sequencing the human genome was first proposed during the mid-1980s, the undertaking seemed hardly feasible using the technology available at that time. However, after various preparatory meetings, the National Institutes of Health and the Department of Energy of the USA officially announced on October 1, 1990 the launch of a program to sequence the human genome, and James Watson (of Watson and Crick) was

Correspondence: A. Pardo.
Facultad de Ciencias, Universidad Nacional Autónoma de México,
Ciudad Universitaria, s/n, 04510 Del Coyoacán, México DF, México.
E-mail: aps@hp.fiencias.unam

Manuscript received August 26, 2003.
Accepted for publication September 9, 2003.

appointed director of the recently created National Center for Human Genome Research. Around the same time, the public consortium known as the Human Genome Project was formed, and this organization announced a 15-year plan (from 1990 to 2005) with the following objectives: *a*) to determine the complete nucleotide sequence of human DNA and identify all the genes in human DNA (estimated to number between 50 000 and 100 000); *b*) to build physical and genetic maps; *c*) to analyze the genomes of selected organisms used in research as model systems (eg, the mouse); *d*) to develop new technologies; and *e*) to analyze and debate the ethical and legal implications for individuals and for society as a whole.

Public Interest Versus Private Interests

One of the difficulties that had to be overcome in the task of accurately sequencing the bases that make up the human genome was that approximately 50% of DNA is highly repetitive. The strategy adopted by the HGP was to sequence the DNA whose location on the chromosomes was already known.³ However, this strategy was challenged in 1998 by J. Craig Venter and his team, who had just set up a private company called Celera Genomics. Taking advantage of recent advances in technology, this team proposed an alternative strategy based on cutting the genome into small segments and using a computer to reassemble the sequences by matching the overlapping ends of each fragment. With these innovations, this private consortium announced that they would sequence the human genome in 3 years, in other words, that they would complete the task by 2001. This undoubtedly brought immense pressure to bear on the public group, the HGP, headed since 1992 by Francis S. Collins, and also gave rise to fears that a private company might control a large part of the human genome through patents. After several unsuccessful attempts to get the private and public sector groups to collaborate, an agreement was reached to simultaneously publish a first draft of the human genome in February, 2001. This draft did not, however, have the degree of precision of the current one. Consequently, the HGP Consortium published its results in *Nature*⁴ in February 2001, and Celera did likewise in *Science*.⁵

The sequences were subsequently corroborated with a greater degree of reliability, and in April 2003, with the sequence practically complete, the HGP Consortium declared the task to be completed.^{6,7}

Discoveries and Surprises

The Number of Genes in the Human Genome

One of the surprising facts thrown up by the sequencing of the human genome was that it only contains approximately 30 000 genes. Owing to its size, it had been estimated that the genome would contain between 50 000 and 100 000 genes.

In simple organisms, such as yeasts, the number of genes directly correlates with the size of the genome because most of the information in the genome clearly codes for proteins, and the individual genes have a well-defined beginning and a clear stop point and exit for the messenger RNA. It had seemed logical, therefore, that the greater the complexity of the organism, the larger would be the number of genes. However, the sequencing of the genomes of other organisms has yielded unexpected results. For example, the common fruit fly, *Drosophila melanogaster*, has approximately 13 500 genes, fewer than other simpler organisms, such as the earth worm, *Caenorhabditis elegans*, with 18 500 genes, and the mustard plant, *Arabidopsis thaliana*, with around 28 000.⁸⁻¹⁰ Therefore, the human genome only has around 2 000 more genes than *Arabidopsis* despite its obviously greater biological complexity.

So we have learned that the human genome has fewer genes than expected and also that the distance separating them is considerable. It has been calculated that the gene density in the human genome is around 12 per 100 000 bases, while in *Drosophila* this figure is 117, and in *Arabidopsis*, 221.

It is important to understand that the genes in human DNA, as in most eukaryotes, are highly fragmented; in other words, not all of the bases from the beginning to the end of the gene are read to make a protein. The DNA in the genes has coding regions, called exons, interrupted by long noncoding sequences, called introns (intergenic regions). These noncoding regions are removed by the process of splicing in the formation of messenger RNA, so that the resulting messenger RNA is much shorter than the original DNA from which it was produced. For example, it has been reported that around 54% to 59% of genes in human chromosomes 14 and 22 undergo alternative splicing—the exons combine in different ways and produce various different proteins.^{4,5} This means that the number and variety of proteins in an organism does not depend solely on the number of genes in the genome, but rather on the way these genes are used.

Another important question thrown up by the results of the HGP was the following: If only 1% to 2% of the bases in the human genome code for proteins, then what do the rest do? An equivalent part of the noncoding portion of the genome probably contains most of the sequences that regulate the expression of genes, such as the promoters, regions that occur before the beginning of the gene. There are many other elements in the genome that affect the behavior of other components, such as the centromeres and telomeres. Finally, a large part of the genome is made up of highly repetitive DNA sequences, the function of which is little understood.

Why are there so many repetitive sequences in the human genome not found in the genomes of invertebrates? Many DNA sequences seem to have originated as a result of the movement of genetic elements called transposons, segments of DNA that can

move from one site to another within the genome. It has been postulated that many of the changes that have occurred during the evolution of vertebrates may have been triggered by the action of transposons which jumped to regulating regions and modified the expression pattern of the genes.

Genome sequencing is a tool that allows us to reconstruct the history of hundreds of millions of years of evolution marked by mutation, that is, the process of exchange and rearrangement of the sequences that has contributed to the formation of new species or has given rise to new genes. The task of solving these puzzles and fitting each piece into its place still presents a huge challenge because clues to our history still lie undiscovered in the noncoding sequences found in each chromosome, the sequences previously considered to be "junk DNA." For example, the complete sequencing of the sex-determining Y chromosome has revealed some very intriguing facts that have aroused great interest among geneticists and biologists who study evolution. These will be described in general terms in the following section.¹¹

The Y Chromosome

The 2 human sex chromosomes, X and Y, both had their origin in the same ancestral autosome several hundred million years ago, but their sequences diverged through evolution. As a result, sequences identical to those of the X chromosome that permit recombination between the two chromosomes in those regions only exist today in the terminal regions of the Y chromosome. However, over 95% of the modern Y chromosome has specific regions with no equivalents on another chromosome that would enable recombination during sperm production, and this is a rare example of persistence in the absence of sexual recombination. These regions contain genes that specifically code for testicular proteins as well as highly repetitive sequences which—probably because they are not understood—were previously considered to be nonfunctional "junk" DNA. With the complete sequencing of these regions, it has been found that some of these sequences are palindromic (as in the phrase *Anita lava la tina*); that is, they read the same from left to right as from right to left, on both strands of the double helix. This fact has led to the hypothesis that X-Y recombination has been replaced by recombination between the arms of the Y chromosome in the regions where the palindromic sequences are located.¹² In this context, the Y chromosome reveals great powers of self-preservation, using evolutionary strategies to survive in the absence of recombination with another homologous chromosome.

Genomics and Health

Probably one of the greatest expectations generated by the sequencing of the human genome has been the hope that this knowledge might benefit humans through

its medical applications. The understanding of the role played by genetic factors in human health and disease will make it possible for us to discover better ways to approach the prevention, diagnosis, and treatment of pathological processes. It is thought that the science of genomics will soon explain the mysteries of the hereditary factors associated with heart disease, cancer, diabetes, schizophrenia, and many other chronic degenerative processes. It is also hoped that it will give us a better understanding of the genetic factors that influence our susceptibility and/or response to various infectious diseases. Genomics holds the promise of individualized medicine that can be tailored to each patient's genetic profile.

One of the challenging aspects of any analysis of the influence of an individual's genes on the development of certain diseases is ascertaining whether a particular disease is caused by a single gene or the interaction between several genes. It is also essential to understand how the environment influences the expression of such interactions.

There are relatively few known diseases that are associated with mutations in a single gene. They include sickle cell anemia and cystic fibrosis. In the case of the gene that causes cystic fibrosis, over 900 different mutations have been identified that affect the function of the protein it encodes. In normal cells, the protein produced by this gene acts as a channel that allows cells to release chloride and other ions. In people with cystic fibrosis, however, this gene has a mutated sequence, and the protein produced is defective so that the cells do not release chloride. The result is an improper salt balance. This gives rise to the production of an abnormally thick mucus which, among other things, obstructs the airways and leads to infections.¹³

However, the origin of most human diseases and of the variations in individual responses to drugs is more complex and involves the interrelation between multiple genetic factors, such as genes and the proteins they produce, and nongenetic factors, such as the influence of the environment.

Although all individuals share DNA sequences that are 99.9% the same, each person has a unique genome. The remaining 0.1% is responsible for the genetic diversity between individuals. Many differences are due to a variation in a single base pair in a gene. Single nucleotide polymorphisms (SNPs) are variations of a gene that occur because of a change in a single letter (nucleotide) in the DNA sequence, for example, the substitution of "CTA" for "CCA." SNPs contribute to the differences between individuals. While most of these polymorphisms have no effect, others cause slight differences in certain characteristics that do not affect health, such as physical appearance. Others, however, may increase or decrease the individual's risk of developing certain diseases.

This happens, for example, in the case of acquired immune deficiency syndrome (AIDS). We now know that not all individuals exposed to the type 1 human

immunodeficiency virus (HIV) become infected, and that the progression period from infection to AIDS is highly variable among infected individuals. Some patients may develop the disease in 3 years, while others remain asymptomatic for more than 15 years. Although the reasons for these differences are not entirely understood, it has recently been discovered that genetic factors play a very important role in the transmission of the virus and progression to disease. There must be 2 co-receptors on the surface of a cell in order for the virus to attach itself effectively and later infect the host cell. The first of these is CD4, the key receptor for T lymphocyte facilitators, and the second is one of the members of the chemokine family of receptors. C-C chemokine receptor 5 (CCR5) is one of the main co-receptors used by the virus to penetrate macrophages and T lymphocytes, so that it plays a critical role in the pathogenic process of AIDS. Several studies have demonstrated that the polymorphic allele CCR5-Delta32 (which contains a 32 base pair deletion) has a powerful protective effect in the progression of the HIV infection.¹⁴

Similar findings will probably emerge in relation to other diseases, so that in the future we will understand such enigmas as why not all smokers develop chronic obstructive pulmonary disease or lung cancer, or why not everyone who is exposed to avian antigens develops hypersensitivity pneumonitis.

Scientists have started to compile a catalogue of the common variations in the human population, which includes SNPs, small deletions and insertions in the coding DNA, and other structural differences. Part of this database is already available to the public.¹⁵

Another important point is that sets of nearby SNPs on the same chromosome are inherited in blocks. These patterns of SNPs on a block are known as haplotypes, and certain SNPs can be used as tags to identify the haplotypes in a block.

The elucidation of the complete human genome has given rise to a new project the aim of which is to develop a haplotype map of the human genome called the HapMap.¹⁶ The HapMap locates blocks of haplotypes, and the specific SNPs that identify them are called SNP tags. The International HapMap Project was started in 2002 and will be of fundamental importance in examining the genome in relation to phenotypes. It will also be a tool that will enable researchers to identify the genes and genetic variations that affect health and illness.

In addition to its use in analyzing the relationship between genes and disease, the HapMap will be a powerful resource for studying the genetic factors that contribute to individual variations in our response to environmental factors, susceptibility to infection, adverse reactions, and response to drugs and vaccines. Using only the SNP tags, researchers will be able to identify regions on the chromosomes with different distributions of haplotypes in two groups of people, for example, those who suffer from a disease and those

who do not. This will also facilitate the development of tests that can predict which medicines and vaccines might be more effective in individuals with particular genotypes for the genes that affect the metabolism of these drugs.

Functional Analysis of the Genome

The complete sequencing of the genome of an organism is only the first step in the quest to understand its biology. It is still necessary to identify all the genes and ascertain the function of the products expressed by these genes, that is, functional RNA and proteins. Functional genomics is based on the key premise of the central dogma of molecular genetics, which states that DNA sequences are used as templates for the synthesis of RNA, and this RNA is subsequently used as a template for the synthesis of proteins.¹⁷ Moreover, scientists still have to analyze and understand the noncoding regulatory regions and other functional elements of the human genome and of the genomes of other organisms. This has led to the creation of a project called the ENCyclopedia Of DNA Elements—or ENCODE. The goals of this new project are to identify and map the exact location of all the protein-encoding and non-protein-encoding genes, and to identify other functional elements encoded in the DNA sequences, such as promoters and other transcriptional regulatory sequences, as well as determinants of chromosome structure and function, such as origins of replication. The aim is to provide a comprehensive encyclopedia of all these elements in order to help researchers better understand human biology and predict potential disease risks, and to stimulate the development of new therapies for the prevention and treatment of disease.

It has been said that the basis for understanding the genome of a mammal is the characterization of the part that is transcribed (ie, the transcriptome) and the identification of the proteins it produces (ie, the proteome).

Many technologies have been developed to study functional genomics, and foremost among these are the cDNA microarrays or DNA chips, which have been widely used to explore the expression profiles of thousands of genes simultaneously.^{18,19}

This technology has been used to gain a greater understanding of the molecular mechanisms of various diseases, such as, for example, pulmonary fibrosis. Idiopathic pulmonary fibrosis belongs to the category of idiopathic interstitial pneumonias and is characterized by the relatively rapid destruction of the lung parenchyma. As a result, some 50% of patients die within 3 years.²⁰ In a recent study, lung biopsy samples from patients with idiopathic pulmonary fibrosis and other patients with normal lungs were analyzed using this technique of oligonucleotide microarrays.²¹ The results showed that gene expression patterns clearly distinguished normal from fibrotic lungs, and that many of the genes that were significantly increased in fibrotic

lungs encoded proteins associated with the extracellular matrix and enzymes responsible for its replacement. This study, and others that have investigated various pathological processes,²² illustrates the analytical power of gene expression in the identification of the molecular pathways involved in disease.

The identification of the different groups of genes involved in the pathogenic processes of human disease will also facilitate the discovery of new molecular targets that can eventually be used in the treatment of such diseases.

For example, we have recently found in hypersensitivity pneumonitis, an inflammatory lung disease characterized by lymphocytic alveolitis, the exaggerated expression of a chemokine derived from dendritic cells known as CCL18. This chemokine is a powerful attractor of T lymphocytes and, at least theoretically, blocking it for therapeutic reasons could reduce the lymphocyte infiltration that characterizes this disease.²³

Other new genomic technologies include: *a)* toxicogenomics, which studies the genetic basis of an individual's response to environmental factors, such as drugs and contaminants; and *b)* pharmacogenomics, which deals with the development of drugs designed for specific pathogenic processes that will target specific metabolic pathways. In general terms, the genomic sciences have been defined as those which study genes, their products, and their interactions.

Bioethics

One of the earliest objectives of the HGP was to set up a program, called ELSI, to analyze the ethical, legal and social implications of genomic sciences. In this context, UNESCO created the International Bioethics Committee, and in 1997 published a declaration that states, "Recognizing that research on the human genome and the resulting applications open up vast prospects for progress in improving the health of individuals and of humankind as a whole, but emphasizing that such research should fully respect human dignity, freedom and human rights, as well as the prohibition of all forms of discrimination based on genetic characteristics, proclaims the principles that follow and adopts the present Declaration."

The articles of this declaration²⁴ deal with the following topics: *a)* human dignity and the human genome; *b)* rights of individuals; *c)* research on the human genome; *d)* conditions for the exercise of scientific activity; *e)* solidarity and international cooperation, and *f)* the promotion of the principles set out in the declaration. Article 1 of this universal declaration on the human genome and human rights states: "The human genome underlies the fundamental unity of all members of the human family, as well as the recognition of their inherent dignity and diversity. In a symbolic sense, it is the heritage of humanity."

The medical application of the information generated by genetics must be consistent with the general

principals of medical ethics: *a)* beneficence, or acting for the good of individuals and their families; *b)* doing no harm; *c)* respecting the autonomy of the individual, that is, allowing individuals to make independent decisions after providing them with information; and *d)* individual and social justice.

Genetic information is confidential, and it is the responsibility of institutions and authorities not to interfere without prior consent. However, there are certain circumstances that could justify the intervention of the state, such as those related to public health issues, or the well-founded request of an authority in connection with a judicial investigation.

How can we define the limits between what is permitted and what is prohibited, or between privacy and responsibility towards third parties? These are the kind of topics that must be discussed and analyzed by the ethics committees in each country, which should then inform their respective legislators on these issues. Other aspects that need to be reported and considered include: privacy and justice in the use of genetic interpretation, nondiscrimination, and the need to distinguish between information that we individually prefer not to know and facts that must be revealed for family or social reasons.

Closely related to these ethical considerations is the problem of the privatization of knowledge and the granting of patents. For example, the last nucleotide in the genetic code of the coronavirus responsible for severe acute respiratory syndrome had hardly been read when the race had already begun to take control of the intellectual rights to the sequence. In private hands, a patent on a viral sequence could delay or increase the cost of developing a treatment or diagnostic tests for a particular disease. This question has caused concern among biomedical researchers, who are afraid that broad patents on genetic sequences will affect research work in universities and public institutions and will have a detrimental effect on future public health strategies. An example of this is the case of the predictive test for breast cancer, which uses the genes *BRCA1* and *BRCA2*. The Curie Institute in Paris has been struggling for the right to continue analyzing these genes at a third of the price currently charged by the genome company Myriad Genetics (Utah, USA), which was granted a European patent for these genes in 2001.

Future Prospects

Molecular biology has implicitly promised to transform medicine by elucidating the smallest details of the mechanisms of life. To the extent that the molecular processes of diseases are revealed, we will, in many cases, be able to prevent them or to design effective cures or individualized treatments. Genetic tests will be able to predict an individual's susceptibility to a disease, and the diagnosis of many pathological processes will be much more detailed and specific than it is today. New drugs will be designed based on an understanding of the

molecular mechanisms of common diseases, such as diabetes and systemic arterial hypertension, and it will be possible to treat these diseases by focusing on specific molecular targets. In the case of diseases such as cancer, for example, drugs can be adapted to the specific response of the patient and, within a few decades, it will be possible to cure many potential diseases at a molecular level before they develop.

Most probably these changes will not all occur in the immediate future. It will take us a long time to understand the human genome, the book of our species, with its 23 chapters called chromosomes, each containing thousands of stories known as genes, composed of paragraphs called exons, interrupted by as yet indecipherable messages called introns, written in words called codons, made up of letters called bases.

No doubt, access to the exact sequence of the genome will gradually modify, with increasingly greater impact, the practice of medicine in the coming decades, and in this context it is essential that this knowledge and these technologies be immediately incorporated into public and professional education; this is a priority and the task must begin today.

Prometheus stole fire from the gods for the benefit of mankind; it is up to us to ensure that our new Promethean knowledge be used to throw light on many of the mysteries of biology.

REFERENCES

1. Watson JD, Crick FHC. Molecular structure of nucleic acids. *Nature* 1953;171:737-8.
2. Reddy VB, Thimmappaya B, Dhar R, Subramanian KN, Zain BS, Pan J, et al. The genome of simian virus 40. *Science* 1978;200: 494-502.
3. Watson JD. The Human Genome Project: past, present, and future. *Science* 1990;248:44-9.
4. The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409: 860-921.
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
6. Coolins FS, Green ED, Guttmacher AE, Guter MS. A vision for the future of genomics research. A blueprint for the genomic era. *Nature* 2003;422:835-47.
7. National Center for Biotechnology Information (NCBI). Human genome sequencing Available at: <http://www.ncbi.nlm.nih.gov/genome/seq/>
8. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000;287:2185-95.
9. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science* 1998;282:2012-8.
10. Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, et al. Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* 2000;408:823-6.
11. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003;423:825-37.
12. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 2003; 423:873-6.
13. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989;245:1066-73.
14. Ioannidis JP, Rosenberg PS, Goedert JJ, Ashton LJ, Benfield TL, Buchbinder SP, et al. International meta-analysis of HIV host genetics. Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3' A alleles on HIV-1 disease progression: an international meta-analysis of individual-patient data. *Ann Intern Med* 2001;135:782-95.
15. NCBI. Single nucleotide polymorphism. Available at: <http://www.ncbi.nlm.nih.gov/SNP>
16. Available at: <http://genome.gov/Pages/Research/HapMap>
17. National Human Genome Research Institute. Crick F. Central dogma of molecular biology. *Nature* 1970;227:561-3.
18. Petricoin EF III, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, et al. Medical applications of microarray technologies: a regulatory science perspective. *Nat Genet* 2002;32(Suppl):474-9.
19. Busquets X, Agustí AGN. Chip genético (ADN array): el futuro ya está aquí. *Arch Bronconeumol* 2001;37:394-6.
20. Selman M. Clasificación actual de las neumonías intersticiales idiopáticas [editorial]. *Arch Bronconeumol* 2000;36:543-4.
21. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, et al. Gene expression analysis reveals matrixin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci USA* 2002;99:6292-7.
22. Sheppard D. Uses of expression microarrays in studies of pulmonary fibrosis, asthma, acute lung injury, and emphysema. *Chest* 2002;121(3 Suppl):21S-5S.
23. Pardo A, Smith KM, Abrams J, Coffman R, Bustos M, McClanahan TK, et al. CCL18/DC-CK-1/PARC up-regulation in hypersensitivity pneumonitis. *J Leukoc Biol* 2001;70:610-6.
24. UNESCO. Universal Declaration of the Human Genome and Human Rights. Available at: http://www.unesco.org/human_rights/hrbc.htm