Committee 2
Genetic Knowledge, Human Values
and Human Responbility

DRAFT--7/1/95
For Conference Distribution Only

THE HUMAN GENOME: FACTS, ENIGMAS AND COMPLEXITIES

by

Guido Pincheira
Professor of Genetics
University of Chile
Santiago, CHILE

The Twentieth International Conference on the Unity of the Sciences
Seoul, Korea August 21-26, 1995

ICUS XX
Committee II.
"Genetic knowledge, human values
and human responsibility"

The human genome: facts, enigmas and complexities

Dr. Guido Pincheira V.
Universidad de Chile

Organisms may be considered, primarily, as genetic systems in a very dynamic interaction with environmental factors.

The genome or genotype is the total genetic information of an organism. The phenotype is the appearance of such organism at different stages of development.

During the last few decades, an impressive growth in our knowledge of Genetics at the molecular level has occurred. In fact, only 50 years ago the chemical nature of the genetic material, the nucleic acids, was known.

Forty-two years ago, in 1953; the english physicist Francis Crick and the american biologist, James Watson, proposed the molecular structural organization

of deoxyribonucleic acid (DNA), a double helix formed by two nucleotide chains held together by hydrogen bonds between two pairs of nitrogen bases:

Adenine (A)=(T) Thymine

Guanine (G)=(C) Cytosine

In 1961, the genetic code for 20 aminoacids was uncovered, indicating how DNA controls the composition of proteins. Living systems share the basic characteristics of such genetic code determining a degree of relatedness in the living world that has a very important meaning for the unified understanding of life.

In 1977, DNA sequencing became possible with the use of restriction enzymes (endonucleases) and the application of recombinant DNA techniques.

In 1981, the first transgenic animal was obtained by incorporation of a human gene into a mouse.

Most of these advances were achieved by studying the genomes of some viruses and bacteria, because they are less complex and offer qualitative and quantitative advantages for research. For example, the genome of E. coli is a DNA molecule made of 4.2 millions nucleotide pairs, organized in one chromosome that host 1500 genes, approximately. Comparatively, the amount of DNA in the genome of the human species and other higher organisms, is much higher and, it is present in association with many different proteins, forming the nucleoproteins, also called chromatin. Structurally, the genetic material of these organisms is found forming spools, coils and supercoils in order to architect the chromosomes with a solenoid type of structure made of nucleoproteins.

The human genome, has a DNA content of 3.3 billion nucleotides pairs, distributed and packed in one set of 23 different chromosomes. Each chromosome

contains only one DNA molecule, which is made of hundreds of millions of nucleotide pairs.

Hence, the human genome must not be viewed as one extended linear DNA double helix; but as a group of topological structures of nucleoproteins (the chromosomes) that must function in many different ways.

Physiology

The genome is, generally, in a very dynamic state and should not be considered only as a fixed stable body of information coded by sequences of nucleotides.

The genome physiology includes the following processes:

- replication in a semiconservative way to create similar genomes;
- mutation to produce new forms of genes;
- repair activities to correct its own defects or errors;
- recombination and reorganization to generate new combinations of genes,
- expression and regulation of the coded information through transcription and translation to produce RNAs and proteins; in order to control metabolism and other cellular processes.
- interactions with many other molecules to make possible its own structural transitions during the cell cycle.

After intensive studies in different organisms, the physiological activities of the genome are only partially characterized. Although genetic research has provided many important concepts and notions that apply to human beings, it has also unveiled important differences in the organization and function of the genome, mainly derived of more complex structural characteristics.

Since, the chromatin has most of human DNA structurally inaccessible and functionally relatively inactive, these genome activities require processes of unwinding and rewinding or coiling of the nucleoproteins and the opening of the DNA double helix with a high degree of precision, generating recurrent and reversible changes of chromatin conformation.

What we really know is that the concept of specific pairing between nitrogen bases is essential for the nucleic acid physiology and that the helical structure of DNA is very flexible, having the ability to change its organization in order to function.

It has also been discovered that DNA may exist in more than one type of double helical structure, depending of the characteristics of the environment in which the DNA is present.

It has also been possible to prove the role that elements of symmetry and symmetry operations play in processes of recognition and interaction between nucleic acids and proteins involved in the genome physiology.

Nevertheless, there are still many questions about why and how these genome functions are performed and regulated, specially in higher organisms. An example of this situation is that, although we know how the E. coli chromosome replicates, we still do not understand the precise mechanism of chromosome duplication or chromosome pairing in higher organisms, because chromatin activation or inactivation are still not well understood.

Genetic goals

One of the primary aims of Genetics is to determine the function and localization of the genes in the genome.

Genes are very versatile structures. There are genes that are permanently active and are called constitutive genes. Other genes are active for a certain and

very precise time, during which the product of their activities creates the conditions suitable for other genes to function.

In this respect, the most fundamental problems to be solved are related to the coordination of gene activities and how specific genes are turned on or off at particular times and at a particular stage of development of the organism.

In order to have a gene active, it is necessary to provide the right conditions in the environment in which the gene is present. This is specially true and relatively simple in bacteria, but in animal cells, due to a more stable environment, provided by other tissues or body fluids, the mechanisms of gene action may be very sophisticated.

The characterization of genes present in the human genome is an extraordinary complex task, due to the diversity of their functions, the interactions between parts of the same gene, between different genes localized in different chromosomes and between genes and environmental factors.

Functionally, the genes are usually identified by mutations that are manifested by biochemical or cytogenetic correlates. In human beings, often the gene in which a mutation occurs has no detectable biochemical or cytogenetic correlation and it may only be recognized by its effect in the phenotype of persons inheriting the mutation. In such cases the localization of the gene is usually done by classical linkage analysis in families showing genetic segregation of the defect. This analysis may be intricate, because it is necessary to have other well characterized genes serving as genetic markers in specific chromosomes.

Modern Genetics, through a combination of techniques at the molecular level and the analysis of data obtained from observations of hundreds of millions of individuals, makes possible to detect small subtle variations in the phenotype of human beings that facilitate the identification of genes and genetic defects.

An example of such strategy was the pinpoint of the cystic fibrosis gene in the human genome. Lap-Chee Tsui and col. studied, first, members of the families with the disease and by linkage analysis located the gene on chromosome 7. Later, by using a gene probe, they were able to detect mutations of the gene in patients with different racial or ethnic background.

A similar methodology was followed to identify other genes responsible for leukemia, vascular diseases, etc.

Genome organization

Concerning the genome organization, it is clear that genes are in a linear disposition in the chromosomes, with relatively precise limits and intervals between them.

Nevertheless, it is also known that parts of certain genes may overlap and a stretch of DNA may code for more than one protein, influencing more than one character.

The genome organization is, usually, represented by genetic maps, which may be considered as relatively stable; because different portions of the genome may undergo reorganizations of the gene disposition. These processes may produce variations in studies of genetic linkage. and are normally involved in the regulation of gene expression or the creation of new genes. An example of this situation is the reorganization of genes that control the immune system in human beings in which the construction of a functional gene for immunoglobulin takes place in response to the presence of an antigen.

An additional source of complexity in the organization of the genome is the existence of the "insertion sequences" or "transposons" that move through the genome like " jumping genes", taking positions in different chromosomes. In this way, portions of DNA are able to insert themselves at places where they may

produce disruption or inactivation of neighboring genes. An example of these genes in the human genome is a transposon that has been localized in the middle of the gene for blood clothing protein factor 8.

It has been calculated that, approximately, 100.000 genes may be responsible for the different characteristics of a human being. There are genes present in a single copy and genes present in more than one copy (2 to several thousands). The majority of these genes code for the structure of different RNAs or proteins.

Others are regulatory genes of different types, responsible for the switching on or off and the modulation of the expression of structural genes. The mechanisms for these activities are based, as already mentioned, in the interaction between DNA, proteins and other molecules ruled by a sort of a recognition code.

Some regulatory genes may enhance or attenuate the expression of other genes that may be situated at a considerable distance (up to 3000 nucleotides) in the genome. They do this through structural topographic transitions or foldings of DNA or the chromatin which may vary in different phyla.

There are also genes with a type of function still not known. Apparently, they are only concerned with its own survival and are called "selfish genes".

Only 5000 genes have been plotted in the DNA geography of the 23 human chromosomes with different degrees of precision. The majority of these genes code for the structure of different classes of RNA and proteins. Very few are regulatory genes, although they must exist in large numbers. As a consequence, our knowledge of the regions which control the expression of genes in higher organisms is still insufficient.

Another important enigma in understanding the human genome is that 80% of its DNA appears without a clear functional purpose and is usually called "junk DNA". It is possible that this DNA has non coding functional sequences for proteins, as is the case of DNA of telomeres. The function of this DNA needs to be uncovered before we attempt to go further in genome manipulations.

Similarly, it is also known that the coding sequences of many genes are interrupted by non-coding sequences. called introns. Although these introns are transcribed into RNA, afterwards they are removed through different mechanisms in order to provide a precise message for the synthesis of a protein, or to have a fully functional RNA molecule.

Sometimes, the "splicing" operation is incorrect and the resulting gene product is deficient, as occurs in some types of thalassemias, in which the defect produces an abnormal globin chain or a failure to produce it.

Many traits are also multigenic, involving the cooperation of many genes and environmental factors. Among the genes involved in the control of a certain character. there may be differential contributions to express it, etc.

Due to the interrelation of metabolic pathways genes may have multiple effects. Simple one to one relationship between genes and phenotypic characters are rare in the natural world.

In brief, the present status of genetic knowledge demonstrates that we are far away of the concept of a gene as an isolated and fixed locus coding only for a single protein.

There are many other areas of Genetics in which the words of literature Nobel laureate Isaac Basheris Singer "our knowledge is a little island in a great ocean of nonknowledge" have a clear validity.

The human genome project

A great deal of expectation has been created with the human genome project, oriented to uncover the nucleotide sequences of DNA in the 23 different chromosomes: It is important to mention that an average of 99.9% of all nucleotide sequences in the human genome are common to all members of our species. Only, about, 3 million nucleotide pairs are responsible for the biological individuality of each person, but that is a considerable amount of genetic information, magnified by many possibilities of getting new combinations of genes.

Knowing the nucleotide sequences in the human DNA may help to know the structure of many genes, making possible to isolate them and to characterize the physiology of normal and defective genes; specially how the expression of them is regulated and controlled.

Such knowledge may possibilitate for example: the "switch back" of a silent gene, or how to correct defective genes modifying its nucleotide sequence or how to place genes in the precise location in the chromatin present in the cells of a particular tissue or organ for a correct expression.

Gene therapy requires that new genes must be inserted correctly and expressed accordingly in the host genome. If this is achieved, we are in solid ground to repair monogenetic defects, such as hemophilia, Lech-Nyhan syndrome, phenilketonuria, dwarfism, etc.

Multigenic defects will require not only that; but the correct coordination and regulation of expression of the genes involved; a real complex task that has to cope with the hierarchies of organization of the human genome.

Our knowledge of mechanisms and regions controlling gene expression is still very limited and we can not ensure that genes or portions of genes inserted in different regions of a genome or in other genomes will perform under normal

cellular control; which may vary within the same organism. In fact, this situation determines that only a minority of DNA sequences are normally transcribed in any particular cell at a certain time.

At present, it may appear too optimistic to think that we will have the solutions for diverse problems of human gene physiology if we only know the one dimensional nucleotide pairs script of DNA, but it may be a very important step to begin with.

This is clearly demonstrated by the situation that after 13 years of having important basic knowledge to manipulate genes, we still do not find the expected gene therapy for many diseases.

Instead, we continue discussing the same problems with new difficulties derived from our lack of knowledge about expression and regulation of genes in higher living forms.

In the case of the human genome, a preliminary challenge is to know better how chromatin operates with different genetic backgrounds and environmental conditions.

My feeling is that it is very necessary to avoid a certain degree of scientific arrogance that could make us forget the problems that emerge day after day in our efforts to manipulate the genetic material in very complex genomes. In this respect, it is very necessary to try to balance scientific realism with personal optimism in our efforts to apply genetic knowledge.

Although the knowledge that genetic research has provided is tremendously important, it is possible to consider it as working concepts or tentative truths subjected to permanent criticisms, revisions and modifications by more recent research results.

Moreover, scientific explanations of living phenomena usually tend to be fragmentary and may not satisfy the different degrees of complexity of living process as a whole.

The methods of science have built in limitations that we usually tend to ignore when try to apply new information to solve some of the pressing problems of the society without having a clear assessment of the possible consequences in long term scales.

Nevertheless, geneticists, nourished by their natural curiosity, will continue exploring the essence of living phenomena. The results of this continuos effort in science must find the space within a moral framework that ensures the application of such new knowledge without unnecessary risks for the human population and their surroundings.

In 1989, two authors, David Suzuky and Peter Knudtson, published a book with the title "Genethics" (Harvard University Press, Cambridge, Mass.). The epilogue of such book contains a proposition of 10 moral principles, which I subscribe and recommend to adopt in order to cope with the difficult ethical dilemmas originated between the application of contributions of modern genetics and human values.

These ten moral principles are:

1. *To grasp many of the difficult ethical issues arising from modern genetics, one must first understand the nature of genes - their origins, their role in the hereditary processes of cells and the possibilities for controlling them.*
2. *The vast majority of human hereditary differences are polygenic, or involve the interplay of many genes; therefore, it is a dangerous simplification to proclaim a casual relationship between human behaviors and so-called "defects" in human DNA.*

3. *Information about an individual's genetic constitution ought to be used to inform his or her other personal decisions rather than to impose them.*
4. *While genetic manipulation of human somatic cells may lie in the realm of personal choice, tinkering with human germ cells does not. Germ-cell therapy, without the consent of all members of society, ought to be explicitly forbidden.*
5. *The development of biological weapons is a misapplication of genetics that is morally unacceptable - as is the air of secrecy that often surrounds it.*
6. *The information contained in genetic molecules is vulnerable to loss through mutations caused by sunlight, radioactivity, chemicals and other external mutagenic forces, Each of us has a responsibility to develop an awareness of potential mutagens in our immediate surroundings and to seek to minimize environmentally induced damage to our DNA.*
7. *Until we have a better understanding of the extent of genetic exchange between distantly related species in nature, we ought to consider evolutionary "boundaries" - areas of relatively limited genetic exchange - as at least provisional warning signs of potential danger zones for the casual transfer of recombinant genes between species.*
8. *Genetic diversity, in both human and nonhuman species, is a precious planetary resource, and it is in our best interests to monitor and preserve that diversity.*
9. *The accumulation of the genetic knowledge alone - however precious in its own right - does not guarantee wisdom in our decisions regarding human heredity; if such knowledge breeds a false sense of human mastery over genes, it can even lead to folly.*

10. *The research for meaningful ethical principles to help guide us through difficult personal and collective decisions arising from the applications of modern genetics will be an endless process. To succeed, it must lead us beyond the rigid boundaries of Western science and even Western philosophical thought to rich, cross-cultural realms that embrace other ways of knowing.*