GENES and HUMAN DISEASES
Vittorio Enrico Avvedimento

Abstract

The publication of the first draft of the human genome in 2001 has stirred curiosity and attention, because for the first time we can approach some basic questions, which have so far been in the realm of philosophy. For example, we are learning some astonishing facts on the birth and death of genes, new functions appearing or lost during the evolution of our species (see 1). Essentially, we can now ask: 1. from where and how the building blocks of our body and mind were derived and assembled; 2. why we live longer than mice; and 3. why and how we age and die.

The basic structure of DNA is rather simple: two intertwined strands composed by 4 letters: A,T,G,C. These letters (bases attached to a sugar, ribose, and a phosphate) when facing each other in the DNA strands are chemically complementary; i.e., T pairs only with A, G only with C. When a string of a DNA sequence, for example ATCC, is present in one strand, on the other, the complementary sequence assembles TAGG. This is the physical base of inheritance, since each strand can replicate itself chemically, providing a mixture of bases (ATGC) is present in the environment. Evolution has provided an enzyme to make a better and faster synthesis, DNA polymerase, but the basic plan of inheritance and transmission resides in the complementarity of the bases.

The chromosomes are bundles of DNA and proteins (histones) that protect the DNA strings. All together, the chromosomes are 50 cm long and comprise 3.5 billion bases (AGTC). These bundles have been unwound and read. The 4letter codes can fill 350.000 pages of a magazine with no pictures included.
The chromosomes (in humans, 22 pairs plus X and Y as sex chromosomes) are transferred through generations and are formally speaking, immortal. Moreover, they define a species. In fact, a species is genetically isolated: i.e., the genes and chromosomes can be transferred only within the same group of individuals. This is driven by the sex chromosomes, which represent a sort of ship carrying the other chromosomes (autosomes). Sometimes, changes in the code (mutations) in sex chromosomes allow the breeding and transfer of the chromosomes with and to a new group (a new species). This is the basis of **GENETICS,** i.e., a set of general rules governing DNA transfer to the next generation.

The information stored in the 3 billion DNA bases to be expressed has to be transcribed in one strand as RNA. RNA, ribonucleic acid, is very flexible, but very unstable, because the presence of an oxygen in the sugar-base skeleton (in the DNA in this position there is no oxygen, hence the name as 2-**d**e-oxyribo**n**ucleic **a**cid) renders RNA sensitive to the attack and cleavage by the hydroxyl group of water. This is why RNA is not the basic block that forms the chromosomes. It can be lost and degraded easily. The inheritable material has to be stable to be transferred to the next generation. Since our life is embedded in water, DNA has been selected as the best material to keep the code, since it is very stable in water due to the absence of this oxygen.
Once transcribed from DNA, RNA (called a messenger, since it transports information) moves from the nucleus, where the chromosomes are located, to the cytoplasm of the cell where it is translated into proteins. The proteins form the actual blocks of our body (muscle, eyes, hair, skin, organs, brain); they are the receptors that sense the environment, the motor that allows movement, *et cetera*…….

The central dogma in biology states that information flows from DNA to proteins in all living creatures.
The information in DNA, that is transferred into RNA and proteins, is codified by small stretches of DNA called **genes** embedded in long non-coding DNA segments, which are apparently useless. In fact, the genes are coded (inscribed) only by 5% of 3 billions of bases of DNA. The rest is evolutionary "junk" or structural segments stabilizing the chromosomes, such as the centromere, a site that controls division of the bundles or the telomere that stabilizes the chromosomal ends.

All the cells of the body (billions) contain the same DNA and the same instructions since they are derived from a single cell, composed half from the father and half from the mother. Why, then, are the brothers and sisters or cells from different organs (brain versus skin, for example) different? Because the expression of the genetic information varies. Some genes are there in the DNA, but are not expressed.

EPIGENETICS is another level of information layered on the primary code. This epigenetic code specifies heritable changes of gene expression, without altering the primary DNA sequence. Epigenetic modifications of DNA control the expression of the genes. Some of these epigenetics rules are sex specific and differentiate brothers and sisters and cells from different organs (called imprinting). The mechanism?

DNA-proteins bundles, the chromosomes, constitute **chromatin**. **Chromatin** is organized in small particles called nucleosomes. The nucleosomes, formed by a short DNA string and histones, represent the compacted version of DNA. Each chromosome, which are ca. 2.5- 5 cm long, is reduced to a few microns. The packing index is 10,000 fold.

The epigenetic modifications are essential for packing the chromosomes. The condensed chromatin cannot be transcribed, and genes contained in it are silenced. If many segments of the chromatin are silenced, the cell, the organ, and the individual express few genes and lose their functions. Diseases and aging derive from hyper-compacted silenced chromatin.

The mechanisms that compact and silence DNA-chromatin are**:**

1. **methylation of DNA** (methylation means to introduce a CH3 chemical group stably in the DNA, methylated DNA condense the chromatin). DNA methylation is inherited when cells divide.
2. **methylation of the proteins (histones)** that cover the DNA in the chromosomes and pack the chromatin.

QUESTIONS

We have been studying the mechanism of epigenetics - changes that silence and modify the genes in cells derived from different organs (thyroid, skin, liver, heart, and neurons) and in several human diseases, including cancer and autoimmune diseases, and we have asked **2 questions**:

1. **How** can the information in the packed DNA and chromatin be retrieved, unwound, and transformed into RNA (transcription) and once the transfer is completed, how can it be packed back?
2. **What** directs the DNA methylation that silences the genes?

RESULTS

Question 1. To approach these complex issues in a satisfactory way, we reduce or break them down into a series of simple questions that can be addressed by experiments. These experiments include negative and positive controls and produce results that can be disproven or verified to dispel or accept the hypothesis.

To this end, we have used a well-defined model system, in which a protein receptor that binds the hormone estrogen senses the presence of the hormone and upon binding it, signals to DNA to express several genes. Estrogen is a hormone that activates the expression of many genes (that are translated into proteins, which allow duplication and survival of cells in the breast, uterus, liver and brain). The hormone bound to the receptor penetrates into chromatin-DNA and binds a specific DNA sequence present in the genome in several places. We have monitored the initiation of the transcription process of 2 prototypic genes, whose expression is stimulated by estrogens and is translated into mRNA and protein. We have found that the receptor bound to the hormone (estrogen) induces relaxation of DNA,

where the receptor binds, looping out the chromatin, 10.000 bases away from the receptor bound region. Swinging of relaxed DNA promotes the association of the receptor-hormone and the enzyme that synthesizes the RNA (RNA polymerase). This enzyme normally rests at the start site of the gene. As soon as the RNA polymerase touches the receptor, the enzyme gets directions and starts transcribing.

During these experiments, we found that DNA untwisting, which renders the strands flexible, is caused by nicks in the DNA induced by the receptor. These "holes" open an entry site for RNA polymerase, which now can find the bound receptor. These nicks are sealed rapidly by several repair (sewing) enzymes. At the end of transcription, the DNA is wound back into packed chromatin.

This mechanism was unexpected, since cleavage or "nicking" of DNA is rather a dangerous event and the cells try to avoid it or, if it happens, quickly repair the lesion. However, the price to pay for estrogen, sex differentiation, and maintenance may be high, because these repeated nicks and sealings may result in imperfect repair and possible change of the primary code. This mechanism may explain the occurrence of breast cancer, which is considered strictly dependent on estrogens.

Also, it suggests that transcription (central dogma) is a costly process that may work to deteriorate the whole machine in the long run.

(Perillo et al, Science, 2008 Jan 11, 319 (5860):202-6).

Question 2. Genomic DNA can be modified by methylation of cytosine (1 of the 4 bases that make up DNA). This is one important mechanism that silences genes and is exploited by evolution to stabilize DNA, to generate "*imprinting*" and to modify the expression of genes. For example, in tumors, silencing by methylation of genes that inhibit growth ("brakes") can give a powerful selective advantage to a cell. A single cell without these "brakes" becomes immortal, grows and proliferates indefinitely (cancer cell).

To date, the primary cause and the mechanism leading to methylation are not known.

We have formulated a simple hypothesis based on the fact that DNA methylation in somatic (body) cells (not germ cells) occurs randomly, i.e., occurs in segments of DNA, differing from cell to cell. Since DNA damage is also distributed randomly, we have speculated that DNA methylation marks damaged DNA.

By using a sophisticated genetic system, we have induced a break in the double helix of DNA at a single site of the genomes of mouse or human cells. This rupture (a major damaging event) is repaired by a very precise mechanism: the damaged chromosome pairs and retrieves genetic information from an undamaged and homologous chromosome partner. In fact, we have 23 pairs of chromosomes (one set from the mother, the other from the father) with similar information. This diploidy is the advantage of sexual reproduction.

The repair of the double strand break left a SCAR (methylation) in a fraction of the repaired molecules flanking the cut. As a direct consequence, the gene in these SCARRED molecules was silenced.

We conclude that the scar marks damaged and repaired genes. (Cuozzo et al., PLOS Genetics, 2007 3, 1144-1162). This appears to be a strong evolutionary response used by the cell to protect the genome. Packed chromatin DNA is more resistant to damage, and silencing protects the cell from expressing a damaged gene, since the repair may introduce some changes in the code.

Perspectives

The experiments and data indicated above raise many questions. The first and most important is the following: are these peculiar examples or do they represent basic mechanisms underlying transcription and DNA methylation? Can we extend them to all genes? Specifically,

1. Does transcription of all genes start by"nicking" the DNA? If so, what type of protection do transcribed genes use to reduce long term damage?
2. Does DNA methylation mark damaged segments of the genome. How does it happen?

We are trying to set up experiments to address these questions, and this is part of this 2009 semester's activity with Max Gottesman, at Columbia Medical Center's Institute for. Cancer Research. In the meantime, here are some matters that have arisen:

a. We found in an autoimmune disease, systemic sclerosis (a deadly disease in which the skin hardens and becomes thinner), that the cells (fibroblasts) of these patients were over-stimulated by a particular set of antibodies present in the serum. The DNA in these cells was heavily damaged and methylated. In a few generations these cells age and stop growing. These cells exposed to a chromatin de-compacting drug resume transcription and start to duplicate. (Svegliati et al., N.Engl.J.Med. 2006, 354(25):2667-76; Gabrielli et al., 2009, N.Engl.J.Med., 2009 in the press).

b. A set of genes is induced by vitamin A. Retinoic acid (Vit.A) binds a specific receptor and stimulates, like the estrogens, the transcription of several genes. We have preliminary data indicating that the vitamin A receptor starts transcription by inducing nicks in the DNA.

c. The link between double strand break repair, DNA methylation, and gene silencing has been independently confirmed by another laboratory: O'Hagan HM, Mohammad HP, Baylin SB (2008) *Double Strand Breaks Can Initiate Gene Silencing and SIRT1-Dependent Onset of DNA Methylation in an Exogenous Promoter CpG Island*. PLoS Genet 4(8): e1000155. doi:10.1371/journal.pgen.1000155

References

1. See the movie at http://www.youtube.com/watch?v=x0mFEq6aOCM&feature=related
2. Nature 409 pp745-964 (15 February 2001); Science 291, 1145-1434,
(16 February 2001) human genome
3. Nature 420, 515-516 (5 December 2002) Mouse genome