

1

DEVELOPMENTS AND EXPANSIONS IN BIOLOGY AND GENETICS

Ernesto Burgio ISDE Scientific Office

FROM GENETICS to EPIGENETICS "The Times They Are a-Changin'"

In the collective consciousness the word *Genomics* essentially means *Human Genome Project* (HGP). Yet while many scientists continue to speak of *HGP* as the most important tool proposed by molecular biologists, geneticists and... mathematicians to successfully complete the *biomedical revolution* of the twentieth century, there are many researchers who consider *HGP* a failure, a gigantic project which will not pay back the enormous investment (of economic, financial and human resources in general) that has allowed its implementation.

The truth, as often happens, lies somewhere in between.

It is true that *HGP* did not yield the fruit that many expected in the biomedical field, disappointing (at least so far) the expectations of researchers and "Big Pharma".

It is certainly true that bioinformaticians, geneticists, molecular oncologists and pharmacologists have found themselves overwhelmed by an avalanche of sequences, in large part difficult to decipher on the basis of current patterns of genome structure and functioning..

It 's especially true that the results in the bio-medical field - concerning progress in diagnosis, understanding of the genetic basis of diseases and discovery of new therapies, more specific and targeted - were below expectation.

It is also true, however, that the knowledge acquired in the field of molecular biology, genomics (including comparative genomics), developmental and evolutionary biology, genetics of population and the systemic understanding of genome as a molecular network (DNA + epigenome) interacting with the information coming from the internal and external environment are literally **transforming our representation of the genome and its way of "working" and evolving**.

In fact it has become evident that the genome is a complex molecular system (*network*) made up not only by DNA sequence, but also by a dynamic and responsive structure of histones and an "epigenetic" cloud of molecules (methyl and acetyl groups, enzymes, transcription factors, microRNAs ..) revolving around it and playing an important role in transferring information from outside to DNA and in modulating the response, to the extent that some scientists have used the term ***natural genetic engineering***.

To summarize:

It is not true that the *Human Genome Project* (*HGP*) was a failure.

We would rather say that *HGP* has provided us with great scientific advancements..

even if these **findings were not the ones expected** by those who had invested heavily in it (even in economic terms) hoping to quickly find new diagnostic and therapeutic tools in medicine

On the contrary, in the almost 25 years of its history, *HGP* produced a great progress in understanding the genome and its way of "working and changing" in response to information and stimuli from the environment, enabling us to understand:

-that cells and organisms function as (*eco*)-systems. The "**systemic**" **representation of life** being one of the basic ideas in modern biology;

- that **all permanent changes in our phenotype (both physiological and pathological) are the product of complex interactions between environment, epigenome and genome**;

- that **trans-generational changes in epigenetic setting** of tissues and organs could (help us to) explain the dramatic increase in chronic diseases, in all countries characterized by industrial development and western lifestyles... A transformation we may call "***Epidemic Revolution of the twentieth century***" consisting of a significant increase in endocrine and metabolic disorders (*obesity, metabolic syndrome, diabetes II*); neurodegenerative diseases and disorders of neuropsychological development (*ADHD, autism, Alzheimer and Parkinson diseases*); cardiovascular (*atherosclerosis, hypertension, vasculitides*) and immune-mediated diseases (*allergies, autoimmune diseases*); *cancer*..

The lowest common denominator in all of these findings concerns the **central role of epigenetics** which proved the **key discipline** for understanding the **ongoing dialogue between environment and genome**, and the way **information** (coming from the environment) may affect:

- in the short term ***gene expression***,

- in the long run the ***phenotypes*** and even the ***genomes*** (potentially interfering on the **evolution** of living organisms)

FROM GENES to GENOMES

From genomes sequencing we realized that.....

Percentage of **non-coding DNA** increases during evolution

... AS EUKARYOTIC **COMPLEXITY** GROWS SO DOES **NON-CODING DNA**

→ GREATER THAN **95%** OF our DNA

→ LESS THAN 1.5% OF HUMAN GENOME ENCODES PROTEINS, BUT **ALL DNA IS TRANSCRIBED**
40% OF HUMAN GENOME IS **TRANSPOSONS & REPEAT GENETIC ELEMENTS**.

The **C-value enigma** or **C-value paradox** is a term used to describe the complex puzzle surrounding the extensive variation in nuclear genome size among eukaryotic species.

At the center of the C-value enigma is the observation that **genome size does not correlate with organismal complexity; for example, some single-celled protists have genomes much larger than that of humans**.

Prior to the human genome sequence, the expected gene number most commonly cited was 100,000, even though lower estimates were becoming increasingly common ...

As a result, **the finding of 20,000-25,000 genes in the human genome has inspired extensive commentary**. Some authors even characterized this as a new "**G-value paradox**" or "**N-value paradox**", in reference to the "**C-value paradox**"

If the **Central Dogma** of molecular biology depicted **one direction-flow of genetic information**

Now we can argue that things are much different...

....(that) only a **small fraction** of the chromosomal DNA **codes for proteins**

...(that) ... about 45% of the human genome is **derived from transposon DNA**

... (that).. the genome of primates and, in particular, of man is made up

by **8% of HERVs** (endogenous retroviruses) and

more than **half of mobile retrosequences** (characterized by retroviral-like structure) ...

playing a continuing role (reactive-adaptive) in processing / engineering

the entire genome (→ **Natural Genetic Engineering**)

... (that).. **Transposable elements** can be seen as a **natural genetic engineering system**

capable of acting not just on one location at a time but **on the genome as a whole** ..

This dynamic view of the genome has been illustrated most impressively by **Shapiro** who stated that the **genome is composed of modular units arranged in a "Lego-like" manner** that can be altered under certain circumstances

... (that).. the whole Genome is a Complex and **highly dynamic** molecular **Network** of **interacting Genes** and **non-codifying sequences**.. and **proteins**

configuring a new model of molecular-systemic-network genome

composed by *DNA*, *epigenome*, *proteome* and of *sequences communicating in cis and trans*

and, above all, **reactive /interactive (in a continuous dialogue with the environment)**

FIELD INTERACTIONS AND BIOLOGICAL REGULATION

Ernesto Burgio ISDE Scientific Office

As we have already seen the *Genome Project* and the subsequent and increasingly sophisticated, research programs have provided their more valuable results in **basic** rather than in **applied research**, revealing a **genomic organization very different from the expected one**

In the almost 25 years of its history, HGP produced a great progress in understanding the genome and its way of "working and changing" in response to information and *stimuli* from the environment... enabling us to argue:

- that cells and organisms function as (*eco*)-systems. The "**systemic**" **representation of life** being one of the basic ideas in modern biology;

- that **all permanent changes in our phenotype (both physiological and pathological) are the product of complex interactions between environment, epigenome and genome;**

- that **transgenerational changes in epigenetic setting** of tissues and organs could (help us to) explain the **dramatic increase in chronic diseases** in all countries characterized by an industrial development model and western lifestyles... a transformation we may call "**Epidemic Revolution of the twentieth century**": a striking epidemiologic change consisting of an impressive increase in **endocrine and metabolic disorders** (*obesity, metabolic syndrome, diabetes II*); **neurodegenerative** diseases and disorders of neuropsychological development (*ADHD, autism, Alzheimer and Parkinson diseases*); **cardiovascular** (*atherosclerosis, hypertension, vasculitides*) and **immune-mediated** diseases (*allergies, autoimmune diseases*); **cancer**..

The key term in this context is certainly **DOHAD (Developmental Origins of Health and Diseases)**.

According to this **pathogenic model**, initially applied to some specific areas (*cardiovascular, endocrine and metabolic diseases*), that we should instead transform in a sort of **universal pathogenic paradigm**, (many adult diseases would be the product of an altered tissue programming *in utero (fetal programming)*, responsive/adaptive to environmental/micro-environmental conditions/information perceived as potentially **dangerous** for the single organism or simply different from those for which it was genetically **programmed**.

As a matter of fact we can nowadays demonstrate that not only the near-totality of our diseases are directly or indirectly connected with environmental changes/inputs, but our own physiological and pathological phenotypes are largely shaped by environmental factors. Chemical, physical, biological agents continuously induce modifications in the **epigenetic setting** of our cells and tissues (the "*soft wired*" memory of our bodies), above all, in embryo's and fetus' cells and tissues, during the very first stages of development, conditioning in this way the *fetal programming*.

We can also argue that such **trans-generational changes in the epigenetic setting** of tissues and organs could explain the epidemic change we are witnessing all over the world, and especially in all countries characterized by industrial development and western lifestyle, from a prevalence of **acute exogenous diseases** (infectious and parasitic) to a predominance of **chronic endogenous** ones.

We want to remind that if we expose primates (and other mammals) in the very first stages of their development to some **xeno-biotics** (endocrine disruptors etc) or to **heavy metals** (mercury, lead, cadmium) their **fetal programming** (the epigenetic setting of their endocrine and immunological tissues) will change, conditioning the whole life of their cells and tissues and opening the way to many diseases we have already mentioned (breast cancer, Alzheimer disease etc).

In this context we want also to mention the vehement appeal launched some years ago on the pages of *The Lancet* by researchers at the *Harvard School of Public Health*, (Grandjean P. and Landrigan P. *Developmental Neurotoxicity of Industrial Chemicals*. *Lancet*. 2006; 368 (9553): 2167-78) about a **silent pandemic** of autism, dyslexia, ADHD, learning disabilities, afflicting more than 10% of the industrialized world's children, possibly linked to the ubiquitous spread of pollutants - *heavy metals (mercury, lead, arsenic), dioxin-like molecules, PCBs, PAHs, toluene* and hundreds of other toxic molecules, only a fraction of which are known and monitored - in the atmosphere and food chains... inhaled and / or ingested by millions of women.. slowly bio-accumulated in their organs and tissues (especially adipose tissue)... and eventually mobilized and transmitted to embryo and foetus.

We should also try to understand this dramatic change in an **evolutionary context**: we emphasize that all these diseases appear to be both a late result, in adults, of a developmental **process gone awry**, and the product of a **gap between a molecular memory fixed in our cells in millions of years of biological evolution, and an environment changing very quickly**.

2B

CANCER: from SMT to TOFT (Tissue Organization Field Theory of Cancer) ?

A - Is Cancer a Genetic Disease (Weinberg/Vogelstein..) or a Tissue (field), environmental/epigenetic one?

The somatic mutation theory

[Vaux DL *In defense of the somatic mutation theory of cancer* Bioessays. 2011 May;33(5):341-3]

According to the **somatic mutation theory (SMT)**, cancer begins with a genetic change in a single cell that passes it on to its progeny, thereby generating a clone of malignant cells.

(In this MODEL) a (stochastic) change in the DNA of a somatic cell alters its characteristics so that it undergoes clonal expansion. Cells within this population acquire further mutations so that eventually a sub-clone emerges that is able to grow or metastasize sufficiently to cause death of the host.

Since Boveri's initial statement of the theory in 1914 [Boveri T. 1914. *Zur Frage der Entstehung maligner Tumoren*, Vol. 4. Jena, Germany: Gustav Fischer; p. 64.], the SMT has been expanded and refined, to now be the generally accepted model for cancer development [Hanahan D, Weinberg RA. 2000. *The hallmarks of cancer*. Cell 100: 57–70]

B- The somatic mutation theory under attack

[Thomas D, Moore A. *Counterpoints in cancer: the somatic mutation theory under attack*. Bioessays. 2011 May;33(5):313-4]

(Ana Soto and Carlos Sonnenschein, among others..) propose an alternative model to the SMT, in which the role of the **tissue microenvironment** is dominant: the tissue organisation field theory (TOFT)

[Soto AM, Sonnenschein C. *The Tissue Organization Field Theory of cancer: A testable replacement for the Somatic Mutation Theory*. 2011. BioEssays 33: 332–40]

For those who favor a **reductionist view in biology, the cell is the “unit” of the organism**, and hence, explanations of observations gathered at the tissue level of organization should necessarily be found at the cellular level.

However, **the vast majority of phenomena observed during embryonic development are seldom explained when focusing efforts solely at the cell level**. Take for example the development of the kidney; interactions between the ureteric bud and the metanephrogenic mesenchyme results in the reciprocal induction of the collecting system, derived from the ureteric bud, and of the nephron, derived from the metanephrogenic mesenchyme. **A single cell isolated from either one of these tissues, in the absence of the other tissue, fails to originate the tissues** that would result from their reciprocal interactions. Moreover, the shape of the nephron, like that of other anatomical structures, requires **biomechanical forces, which are generated in and by tissues** .. From this evidence, we conclude that interactions among different components of a tissue cannot be reduced to cellular events. **The above-mentioned mechanical forces are emergent phenomena and exert downward causation (from the tissue level to the cell level)**. Thus, **these tissue-based phenomena are inextricably linked to the three dimensions of space (topology) and by their developmental history (time)** ... The first premise of the tissue organization field theory (**TOFT**) **states that carcinogenesis takes place at the tissue level of biological organization, as does normal morphogenesis** (see example above referring to kidney development). **Chronic abnormal interactions between the mesenchyme/stroma and the parenchyma of a given morphogenic field would be responsible for the appearance of a tumor**

[Bizzarri M, Cucina A, Conti F, D'Anselmi F.. *Beyond the oncogene paradigm: understanding complexity in cancerogenesis*. Acta Biotheor 2008; 56: 173–96; Soto AM, Sonnenschein C. 2004. *The somatic mutation theory of cancer: growing problems with the paradigm?* BioEssays 26: 1097–107]

A corollary of the TOFT is that carcinogenesis is a reversible process, whereby normal tissues (or their components) in contact with neoplastic tissues may normalize the latter [Sonnenschein C, Soto AM.

Theories of carcinogenesis: an emerging perspective. Semin Cancer Biol 2008. 18: 372–7]

Disturbed stroma/epithelium interactions are at the core of the **TOFT**. .. the **first report proposing the stroma as the target of a carcinogen based on experimental evidence was published by Orr** ...

Further, Orr concluded that the outcome of his experiments **invalidated the SMT**

[Orr JW, Spencer AT. *Transplantation studies of the role of the stroma in epidermal carcinogenesis*. In *Tarin D*, eds; Tissue Interactions in Carcinogenesis. London, UK: Academic Press 1972.. pp. 291–304]

Soto and Sonnenschein are not alone. For example, in a recent issue of *BioEssays*, Harry Rubin reviewed the critical role of non-cell autonomous changes in carcinogenesis, a process he refers to as **“field cancerization”** [Rubin H. 2011. *Fields and field cancerization: The preneoplastic origins of cancer*. BioEssays 33: 224–31.2].

Most basic research on cancer concerns genetic changes in benign and malignant tumors. Yet evidence indicates that the majority of the mutations in tumors occur in the **preneoplastic field stage of their**

development. That early stage is represented by grossly invisible, broad regions of "field cancerization" which have not, heretofore, been operationally analyzed in cell culture. Conditions are described for quantitating preneoplasia by increased saturation density followed by progression to transformation. These parameters are ***driven by Darwinian selection of spontaneously occurring, cumulative mutations***, in accordance with recent genomic analyses of human cancer, just as it is in the evolution of species. The cell culture model will allow correlation of the preneoplastic increases in saturation density with genetic changes, and development of methods for demarcating fields during surgery so that they can be excised along with the tumor, thereby reducing the possibility of recurrence at the site.

C- We should try to avoid the radical antithesis between A (cancer = genetic disease/SMT) versus B (cancer = tissue disease-TOFT)

it is certainly correct to assume that everything that happens in the heart of the cell (at the genetic + epigenetic level) necessarily **follows** what happens at the ***tissue*** or even at the ***systemic*** level what is **less acceptable is to think that the only significant phenomena concern the tissue level ..** while what happens **at the molecular and in particular at the (epi)genetic level** is something **epi-phenomenal and.. negligible.** Certainly we must **stress** the limits of the **Stochastic (Somatic) Mutation Theory ..** and the importance of moving **from a gene-centric paradigm** to an **epigenetic one (fluid.. adaptive** to the information coming from the environment) ... but it would be meaningless to deny that **a developmental process gone awry** should also be defined by alterations in the (epi)genetic program... in fact, the other important dimension neglected in our discussion, is the **temporal and evolutionary one:** to be clear, **the need (perceived and declared) of representing cancer as a distorted process both ontogenetic & phylogenetic.** Using these coordinates we should reflect more and better.. **if we want to go beyond a simple opposition between models A (genetic/epigenetic/molecular) versus B (systemic)**

D -Prolegomena to a new paradigm in carcinogenesis

Over the past ten years **knowledge in the field of molecular biology, genomics, evolutionary biology have greatly increased and is emerging a whole new model of genome: dynamic and interactive with the environment.** If for almost half a century it was thought to **DNA as a simple reservoir of information (the result of millions of years of molecular evolution and almost unchanging over time)** and the other components of chromatin and in particular the **histones** (the proteins around which DNA is supercoiled, in order to fit into a nucleus of a few microns in diameter) as a simple structure even more stable (**evolutionarily preserved for hundreds of millions of years**) and simply appointed to provide the best means of exposure of DNA (i.e. of 'genes'), in recent years we have realized that **the entire genome should be represented as a molecular network, complex and dynamic, in constant interaction with the environment; while the environment should be seen as a source of information - chemical molecules, metal ions, radiations - interacting with the fluid component of the same genome, the epigenome, pushing it to continually transform itself to respond more effectively to stress.** In such a **dynamic and systemic representation, the three-dimensional structure of chromatin would constitute a molecular complex intimately responsive:** the same genomic and chromosomal changes, should be interpreted in this light **and mutations, traditionally interpreted as stochastic, should be interpreted as active / defensive changes, first epigenomic** (concerning chromatin structure in its three-dimensional frame) and **only later genetic (concerning the same DNA sequence).** **Cancer** should be interpreted as (the product of) a **long and adaptive-reactive process, started in the womb or even in germ cells.**

Big science: The cancer genome challenge

In a recent editorial of *Nature* Heidi Ledford stated that **the millions of genetic sequences and SNPs accumulated** in an attempt to decipher the genetics of cancer have built **giant haystacks in which researchers have gone lost ...**

Many scientists have **looked for mutations that occur repeatedly in a given type of tumour.**

"If there are lots and lots of abnormalities of a particular gene, the most likely explanation is often that those mutations have been selected for by the cancers and therefore they are cancer-causing.."

But.. "It's very clear, now that all the genes have been sequenced in many tumours,

you have drivers that are mutated at very low frequency, in less than 1% of the cancers" says Vogelstein.

To find these low-frequency drivers, researchers are sampling heavily — sequencing 500 samples per cancer should reveal mutations that are **present in as few as 3% of the tumours.**

Separating drivers from passengers will become even more difficult as researchers move towards sequencing entire tumour genomes.

To date, only a fraction of the existing cancer genomes are complete sequences.

To keep costs low, **most have covered only the exome, the 1.5% of the genome that directly codes for protein and is therefore the easiest to interpret..**

Assigning importance to a mutation found in the murky non-protein-coding depths of the genome will be more challenging, especially given that scientists don't yet know what function — if any — most of these regions usually serve. The vast majority of mutations fall here.

The full genome sequence of a lung cancer cell line, for example, yielded 22,910 point mutations, only 134 of which were in protein-coding regions

CONCLUSIONS

Finally it is important to emphasize that all the current - epidemiological and toxicological - methods of assessing the risk associated with environmental pollution are proving to be totally inadequate.

§ The results of such an increasingly massive/collective exposure could become evident after decades, partly because of the common latency between exposure and pathological effects and partly because of the possible **transgenerational amplification** of harm related to the bioaccumulation of pollutants, the "silent transmission" of epigenetic changes and the resulting genetic and epigenetic instability (as seen above the consequences of epigenetic changes directly induced by toxicants may lead to long-term and possibly transgenerational changes in **fetal programming**).

§ For all these reasons we emphasize the **inefficiency of our common methods - epidemiological and toxicological - for evaluating the risks** for our health and the risks for the new generations' health directly connected to **environmental pollution**.

§ **Epidemiologists** generally evaluate the diseases' burden directly connected with environmental pollution by comparing two populations - the one more directly exposed to a known source of pollution (a factory/industrial implant or incinerator etc) or a highway (with high traffic rate), the other supposed to be less exposed - systematically "*forgetting*" that nowadays **we are all exposed** (through the nutritional chains and through direct transgenerational transmission of pollutants from our mothers) to a **constantly growing burden of xenobiotics** (more than 100.000 synthetic molecules) that cannot be recognized by our cellular and nuclear receptors and that may interact in a wrong way with our biochemical pathways and with the genetic expression of our cells and tissues.

§ With regard to **toxicology**, we want to remind that toxicological studies usually analyze the harmful effects of exposure to a **few pollutants for short periods of time**, while it is increasingly evident that the real problem is the continuous exposure to small amounts of thousands of different substances (many of which are poorly studied) for extended periods, and especially the exposure of the embryo and fetus during specific periods of plasticity and vulnerability (**windows of exposure**).

§ The continuous progress in our knowledge in many connected fields (*molecular biology, systems biology, developmental biology*) has pushed some scientists to reevaluate the **neo-lamarckian paradigm** of bio-evolution, in which environmental inputs directly induce and modulate our individual development (*ontogenesis*) and even species evolution (*philogenesis*) (--> **evo-devolutionism**) through a direct induction of epigenetic changes and genetic (reactive/defensive) *shuffling* (**natural genetic engineering**).

§ It is very important to understand what all this means: the **chemical fall-out** we are witnessing all over the world is a real threat for the new generations and the wildlife.

It should be clear by now that the main problem to consider is the **dramatic transformation** produced by man in a few decades. What we're witnessing is a global/molecular alteration that may deeply interfere with the **epigenetic setting** of organs and tissues and, in the long run, even with the **genetic setting of our and other species**: an immense problem that does not seem addressed to date with sufficient attention.