

## The Postgenomic Era: Transcriptomic as the New Paradigm of Life

Bruno Giardina<sup>1</sup>, Marcella Renis<sup>2</sup>, Vito Latora<sup>3</sup> and Salvatore Musumeci<sup>4</sup>

<sup>1</sup> Institute of Biochemistry, University "Cattolica del Sacro Cuore", Roma, Italy

<sup>2</sup> Department of Molecular Biology, University of Catania, Catania, Italy

<sup>3</sup> Department of Physics and Astronomy, University of Catania, and INFN Section, Catania, Italy

<sup>4</sup> Department of Pharmacology, Gynecology, Obstetrics and Pediatrics, University of Sassari and Institute of Population Genetic, CNR/SHARDNA, Alghero/Cagliari, Italy

---

**Abstract:** Now that complete genome sequences are available for a variety of organism, the elucidation of single gene function involved in the metabolism, includes necessarily a better understanding of the cellular responses on all levels of gene products, mRNA, proteins and metabolites. Such knowledge's growth is essential since the observable properties of living organisms, so called phenotypes, are the results of the genotype in juxtaposition with the environmental factors. If much has been done and is in continuous evolution to make clear the mRNA and transcript profiling, considerably less effort has been required to characterize the products of gene expression, the metabolites. Now it is possible to represent cellular processes as metabolic networks and to describe them at different levels of resolution, focusing the attention on the connection map of the metabolic network. Here we describing just the structure of a metabolic network, instead of considering the details of the metabolic flux. For each approach, a number of examples have been given and the potential application have been discussed. Mathematical studies of simple metabolic networks show clearly that, already from an oversimplified approach based only on the analysis of the *structure of the network*, it is possible to reach important conclusions on the network's design principles. Of course one of the most important constraint on metabolic networks comes from the influences of the external environments. For this reason we focus some examples of the effects of the environments on metabolic networks, making of transcriptomic a new paradigm of the life.

**Key words:** postgenomic era, metabolic network, environments, transcriptomic

---

### INTRODUCTION

All the functions and cellular processes of a living organism are regulated by the complex dynamics of the interactions among the cellular constituents: proteins, DNA, RNA and other small molecules. Thanks to the recent progress in experimental techniques of molecular biology, we now have access to a wealth of genome data and to a huge amount of information on cellular processes. In particular we have now the possibility to represent cellular processes as metabolic networks and to describe and study them at different levels of resolution<sup>[1]</sup>. An extremely simplified description, which nevertheless allows a characterization on the large-scale of organization properties of metabolic networks, has recently given a series of very important results. It consists in focusing the attention on the connection map of the metabolic network, i.e. in describing just the structure of the metabolic network, instead of considering the details of the metabolic flux. Nevertheless, the results obtained by such oversimplified (purely structural) analysis, seem to evidenciate the existence of common design principles, results of the networks' evolutionary history and their

interactions with the environmental constraints. The aim of this paper is to open a "brain storming session" on cell path from DNA to protein. Today we can no more consider this as a fairly straightforward process (DNA → RNA → different splicing → protein), but as an economical system enabling the gene to produce a plethora of different (both in structure and in function) proteins, whose expression does not fully reflect the original DNA instruction<sup>[2]</sup>. Therefore the strong need to define the new concept of transcriptomic, intended as the study of a new functional unit, the transcriptosoma.

The paper is organized as follows:

In *chapter I*, we discuss how from a logic organization of Life, where the priority place was occupied by the expression-activity regulation of the enzymatic proteins and by the main regulatory mechanisms of enzymes (the posttranscriptional level), we are moving now towards a new logic organization of Life mediated at the transcriptional level and closer to the influences of the environment. In this way the transcriptomic emerges naturally as the intermediate level between proteomic and genomic.

In *chapter II*, we introduce the graph representation of a metabolic network and we review the main results

obtained by studying the structure (the connectivity properties) of the network. We briefly mention how the transcriptosoma can find its coherent localization in the study of the dynamics of the metabolic network in response to the influences of the external environment. In *chapter III*, we report some examples of the influence of environmental factors on metabolic network, which open the way to a new vision of metabolic pathways and suggest that the metabolic network, dynamically represented, may be influenced and modified in a general economy of entire organism. In *chapter IV*, we conclude that this panoramic review of experimental information may be synthesized in a micro array analysis.

### I. The logic organization of Life: basic concepts

Virtually every cellular chemical reaction occurs at a proper rate only because of the presence of biological catalysts that greatly enhance the rate of specific chemical reactions. Although in a few cases RNA molecules may have catalytic roles, metabolic catalysts are usually proteins i.e. *enzymes* and some times *abzymes*. With a few exceptions, each enzyme protein catalyzes a specific reaction and therefore each cell requires thousands of different enzymes. The multiplicity of enzymes, their specificity and their susceptibility to functional modulation are at the base of cellular processes and of their modulation. The enzyme-catalyzed chemical reactions in cells are functionally organized into different sequences of consecutive reactions, the so-called “pathways”, in which the product of one reaction becomes the reactant in the next. Hence, in a living organism, an impressive number of different molecules are simultaneously synthesized in the precise proportions required by the specific cell in that particular time of the cell cycle. To achieve this goal, key enzymes in each metabolic pathway, as well as the expression of different proteins are finely tuned. Thus the enzymes are at the same time, regulated in their production and regulate various metabolic pathways. The regulated expression of the numerous enzymes as well as of the other proteins may be performed at DNA or RNA or protein levels in different ways also at the same time and elicited by various biotic or abiotic factors. In a system as complex as a living cell, the different mechanisms of enzyme regulation all together constitute what we may call the “logic” of that specific living organism. Since the human genome has been sequenced, it is possible to study the processes generating gene product diversity at different levels: transcriptional, posttranscriptional, translational, metabolic, metabonomic and interatomic.

#### I.A The posttranscriptional level

At the posttranscriptional level a priority place is occupied by the regulated expression-activity of the

enzymes and proteins. Hereby below we list some of the main mechanisms by which the enzymes (proteins) may be regulated:

1) Some enzymes require the presence of free sulfidryl groups and are inactivated if these groups are oxidized. In this respect, glutathione (GSH) may be regarded as a natural intracellular modulator for the proteins having sensitive and essential sulfidryl groups. Thus, the functionality of several proteins can be modulated by controlling the intracellular concentration of reduced GSH, since the level of GSH may interfere with the level of –SH protein groups. It should be noted that this is not a specific effect; rather it depends on the fact that GSH is a natural antioxidant that can be easily synthesized in the cell from readily available precursors (see Fig. 1).

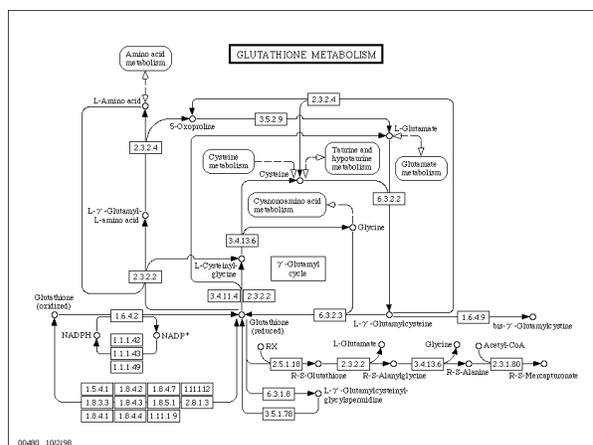


Fig. 1. Pathway of glutathione metabolism (from: [www.genome.ad.jp](http://www.genome.ad.jp))

2) An enzyme involved in a metabolic pathway can also be modulated by an allosteric transition induced by binding of different effectors. In this case the effector is a molecule that drives the enzyme from an inactive to an active conformation. Examples of this type are also provided by many of the ATP regulated enzymes: they are modulated by the rise in ADP concentration that occurs when ATP levels drop or viceversa. For these enzymes the ADP/ATP concentration has a purely regulatory role, in contrast to the substrate role played by ATP in the function of hexokinase.

3) In addition, it has become increasingly clear that an important means of enzyme regulation involves covalent modification. These reactions are brought about by enzymes known as protein kinases and protein phosphatases, which act on other enzymatic proteins as their substrates. The interconversion of an enzyme from the phosphorylated to the dephosphorylated form usually is associated with a marked change in its activity and conformation. An elegant example of this mechanism is clearly shown by glucose metabolism in which, for example, the phosphorylation/dephosphorylation mechanism regulates the glycolytic

rate controlling the mitochondrial oxidative decarboxylation of the pyruvate or glycogen synthesis/breakdown rate.

4) Moreover, in looking at the modifications of enzymes it should be mentioned that some proteins are synthesized as enzymatically inactive form, called proenzymes, which are later modified to the active forms, usually by hydrolytic degradation. Sometimes there is more than one active species depending upon the extent of the degradation. Finally, active enzymes are destroyed by deliberate mechanisms that involve hydrolytic digestion. Thus, as with other cell constituents, the synthesis of enzymes is balanced by degradation. The overall process is described as protein turnover.

5) In addition, there are complex systems named enzymatic cascades whose function is essentially to exert a much more dynamic biological control working also as biological amplifiers of a specific signal, given, for example, by adrenalin or glucagon, collected before by the G proteins and after by adenylate cyclase, a membrane-bound enzyme normally found on the inner surface of the cell membranes. A classic example of this phenomenon is referred to the glycogen metabolism, in which glycogen breakdown and glycogen synthesis are reciprocally activate.

The cascade initiated in the cytosol by activated adenylate cyclase may thereby stimulate, in several steps, a wide variety of enzymes, usually by a phosphorylating step. A key factor in the scheme outlined above is represented by cyclic nucleotides usually indicated as “second messengers”, a term that outlines their importance in trans membrane signaling. The amplification factor of an enzyme cascade is thus a powerful regulatory device.

6) The mechanisms of feedback inhibition regulate most biosynthetic pathways; that is, the final product of the pathway inhibits the enzyme that catalyzes the first step, thus preventing both production of the intermediate products and unnecessary metabolic activity. Feedback inhibition of enzyme function is reversible. If the concentration of free feedback inhibitor (e.g., CTP) falls, the bound inhibitor dissociates from the regulated enzyme, which then reverts to its active conformation. The binding of a feedback inhibitor to an enzyme and its subsequent release can be described by the equilibrium binding constant  $K_i$ , which is similar to the Michaelis constant  $K_m$  used to describe substrate binding to enzyme.

7) A further degree of control is provided by the spatial localization of the given molecular species. This is achieved through the various compartments existing in a eukaryotic cell. Among these we may mention nuclei, mitochondria, lysosomes, micro bodies and vacuoles, where the cytosol tubules and vesicles of the endoplasmic reticulum separate off another membrane-bound compartment. Hence, while many enzymes appear to be dissolved in the cytosol and without any

long-term association with other proteins, in other instances enzymes that catalyze a series of consecutive reactions may be attached to a membrane and held close together as for oxidative enzymes of mitochondria. In other instances several enzymes associate to form a high molecular weight complex as in the case of  $\alpha$ -ketoglutaric dehydrogenase complex or ACP(acyl carrier protein)-lipidic acid synthase.

### I.B The transcriptional level.

For most genes the regulation of transcription initiation is the principal mechanism for controlling gene expression.

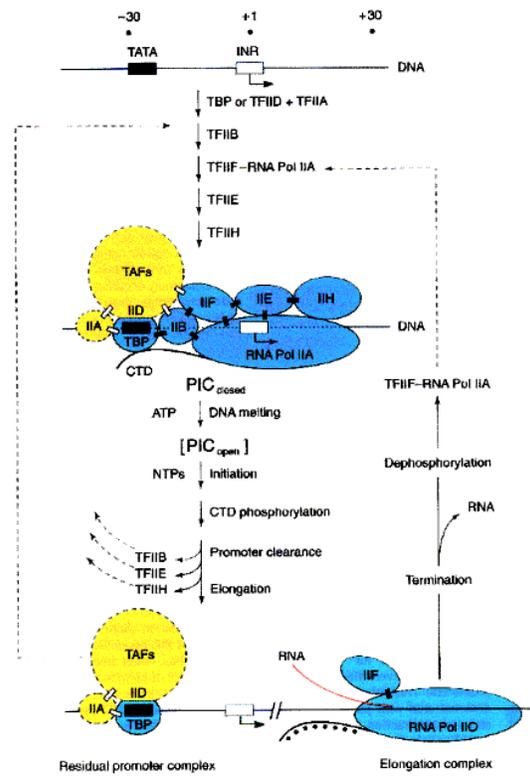


Fig. 2 Eukaryotic Transcriptional Regulators (from: [www.biochem.ucl.ac.uk](http://www.biochem.ucl.ac.uk))

As we have seen in the preceding subsection, however, expression of many highly regulated genes is also controlled at one or more additional steps: elongation and termination, alternative RNA splicing, 3' cleavage and polyadenylation, and even nuclear export. In eukaryotes RNA polymerase II guides the transcription of genes encoding proteins and its activity is regulated other than by a core promoter upstream of the gene containing a set of General Transcription Factors, also by enhancer regions that may be both very distant and upstream/downstream of the gene (Fig. 2). Other forms of post-transcriptional control have been discovered to help regulate expression of some genes. These include

alterations in the protein-coding sequence of a pre-mRNA after its synthesis, regulation of the stability and translation of mRNAs, and regulation of the sub cellular location of specific mRNAs (Fig. 2). Recent works have widely demonstrated that the remarkable variability of transcripts are generated first of all by three mechanisms: 1) alternative splicing; 2) RNA editing process; 3) transcriptional factors. Some basal transcription factors involved in RNA polymerase II-dependent transcription. General transcription factors are multi subunit protein complexes. First of all TFIID binds to TATA box of the promoter; TFIIB binds to TATA-bound TFIID; TFIIB recruits pol II and TFIIF; TFIIE and TFIIH are recruited (largely by pol II); ATP hydrolysis is required for transcription to start (Fig. 2).

### 1) Alternative splicing

Many genes are first transcribed as long mRNA precursors that are then shortened by a series of processing steps to produce the mature mRNA molecule. One of these steps is *RNA splicing*, in which the intron sequences are removed from the mRNA precursor. Often a cell can splice the primary transcript in different ways and thereby make different polypeptide chains from the same genes: such a process is called alternative RNA splicing (Fig. 3).

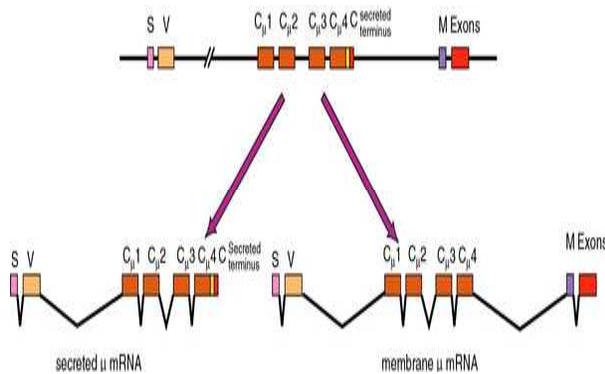


Fig. 3. Alternative splicing (TiBS, 25 August 2000).

Two possibilities are shown for one such gene, the gene for the  $\beta$  heavy chain of the mouse IgM immunoglobulin: on the top is shown the DNA structure of this gene region. The exons are shown as colored boxes, the introns as lines. A pre-mRNA is transcribed from this DNA. It can be spliced in two different ways. On the left, the RNA is splice to include the exons S, V, C $\beta$ 1, C $\beta$ 2, C $\beta$ 3, C $\beta$ 4, and C (the terminus of the secreted form of the protein). This form is translated and sent out of the cell as part of a secreted antibody. On the right is shown a splicing pattern that includes S, V, C $\beta$ 1, C $\beta$ 2, C $\beta$ 3, C $\beta$ 4 and then the M exons. This form of the mRNA is translated into a protein with a trans membrane anchor region (M) and therefore winds up in the plasma membrane of the cell

that produces it. In this way the immune system can produce two different forms of the protein: one that is sent out of the cell as a soluble antibody, and the other remains on the surface of the cell to help identify it to other cells of the immune system (see <http://www.blc.arizona.edu/marty/411/Modules/altsplice.html>). A substantial proportion of higher eucaryotic genes produce multiple proteins in this way. When different splicing possibilities exist at several positions in the transcript, a single gene can produce dozens of different proteins. In some cases, however, alternative RNA splicing is *regulated* rather than constitutive. In the simplest examples regulated splicing is used to switch from the production of a nonfunctional protein to the production of a functional one. The transposase that catalyzes the transposition of the *Drosophila* P element is produced in a functional form in germ cells and a nonfunctional form in somatic cells of the fly, allowing the P element to spread throughout the genome of the fly without causing damage in somatic cells. The difference in transposon activity has been traced to the presence of an intron sequence in the transposase RNA that is removed only in germ cells. In addition to switching from the production of a functional protein to the production of a nonfunctional one, the regulation of RNA splicing can generate different versions of a protein in different cell types, according to the needs of the cell.

### 2) RNA editing

This is a widespread phenomenon, in eukaryotes from protozoa, plants and flies to mammals, concerning in the post-transcriptional modification of RNA sequences (Fig. 4). The RNA editing is an additional "economical" mechanism by which one gene may be translated into multiple proteins with different structures or functions. RNA editing, discovered in mammals in the 1990s is defined as an alteration in the coding capacity of mRNA other than splicing or 3' processing. It may occur after the RNA has been transcribed from DNA (pre-mRNA editing) or before the RNA is translated into protein (mRNA editing) and by two distinct mechanisms: substitution editing and insertion/deletion editing. These alterations are guided by RNA molecules that base pair with the RNA to be edited or serve as template for addition (or removal) of nucleotides and in addition they present at the same time tissue specificity. RNA editing perhaps enable the cell to try out slightly different "recipes" for assembling different proteins without irrevocably abandoning the original version or it may be a way to create proteins with slightly different functions useful in specialized circumstances. An example of this mechanism is the case of antibodies which were produced against any different antigens without any pre constituted information.

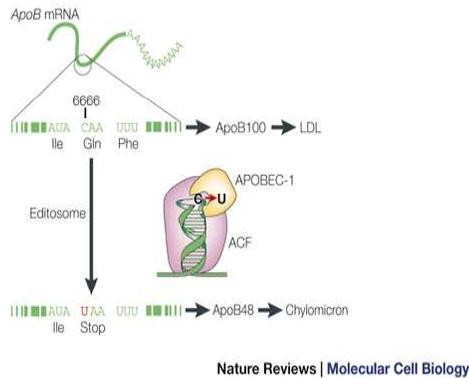


Fig. 4. RNA editing in mammals (*Nature Reviews Molecular Cell Biology* 2; 493-503 (2001))

This is one of nature's ways of creating economically functional diversity!

### 3) Transcription Factors act as great orchestra's conductors

In order to survive, all organisms must be able to respond rapidly and effectively to changes in their environment: These changes might include externally applied stresses such as heat, ultraviolet light, radiation, free radicals or the presence of infectious agents such as viruses, bacteria, parasites and their cellular products or increased concentrations of signaling molecules such as hormones, cytokines and adhesion molecules). Most types of signaling molecules induce cellular responses by binding to specific cellular-surface receptors that respond to occupancy by undergoing structural or biochemical changes that can be transmitted to the interior of the cell. One of the most common responses to receptor ligation is the synthesis of new proteins through alteration of the pattern of gene expression. Consequently, the relatively few transcription factors that regulate inducible gene expression can be the targets for many distinct signal transduction pathway triggered by a wide variety of stimuli. So, differential activation of transcription factors may lead to different cellular responses. The recent knowledge of the increasing number of genes that are being reported whose induction requires activation of so few number of transcription factors, together with the vast array of identified inducers of transcription factors activity, demonstrate the critical role that the transcription factor play in many aspect of cellular responses to environmental changes. Thus it easy to imagine the transcription factors as orchestra's conductors to play the symphony of the cellular life.

1) They are very important genes produced in small quantities, which are able to regulate other genes since their binding sites are present in the promoters of other genes. Some important examples are the AP-1, AP-2, Nf-kB, CREB, PPAR family. The transcription of eukaryotic genes is regulated by combinations of

activators and repressors that bind to specific DNA regulatory sequences. Whether or not a specific gene in a multi cellular organism is expressed in a particular cell at a particular time is largely a consequence of the binding and activity of the transcription factors that interact with the regulatory sequences of that gene. Clearly, since different proteins are expressed in different cells at different times in development, the activity of transcription factors must be controlled. At cellular responses (Fig. 5), obviously, a critical control point for cells is transcription of the genes encoding transcription factors themselves. The transcription factors expressed in a particular cell type, and the amounts produced, are a consequence of multiple regulatory interactions between transcription-factor genes that occur during the development and differentiation of a particular cell type. Concluding the expression of a particular gene is further controlled by

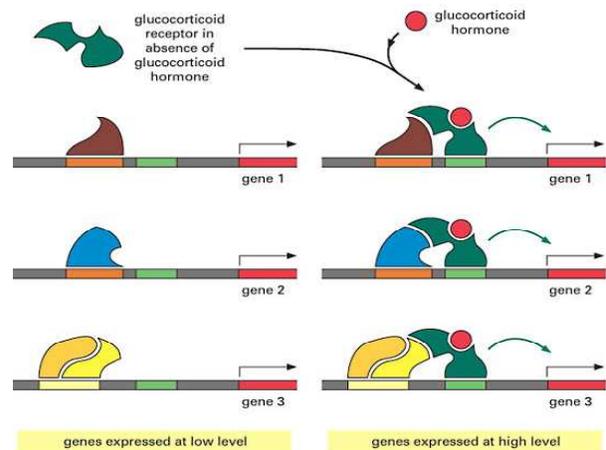


Fig. 5. Regulation of the level of gene expression by some factors required for its transcription ([www.accessexcellence.org](http://www.accessexcellence.org))

regulating the activities of the factors required for its transcription. A limited proportion (approx.1%) of individual with the  $\beta$ -thalassemia carrier phenotype show completely normal  $\beta$ -globin gene sequences as well as normal cluster regulatory regions (LCR). Moreover has been recently demonstrated that mutation in the general transcription factor TFI-IH, coding for a specific helicase, results, in addition to trichothiodystrophy, also in a  $\beta$ -thalassemia carriers phenotype<sup>[3]</sup>. Furthermore mutation in the GATA-1 transcription factor in a sequence motif of the N finger, interacting with the transcription factor (FOG) determines a complex clinical picture characterized by dyserythropoietic anemia and thrombocytopenia<sup>[4,5]</sup>. Finally a mutation in the same GATA-1 factor located of a different site of the N finger results in thrombocytopenia as well the  $\beta$ -thalassemia carrier phenotypes<sup>[6]</sup>. These situations demonstrate that a mutation in trasncription factor may be responsible of defect in  $\beta$ -chain synthesis of the same entity of

mutation in structural genes, such to interact with classical  $\beta$ -thalassemia genotypes

## 2) Nonsense surveillance machiner

Another mechanism by which the cells may control the gene expression is to release mRNAs of high quality only. Nonsense surveillance machinery or nonsense-mediated mRNA decay is an RNA surveillance

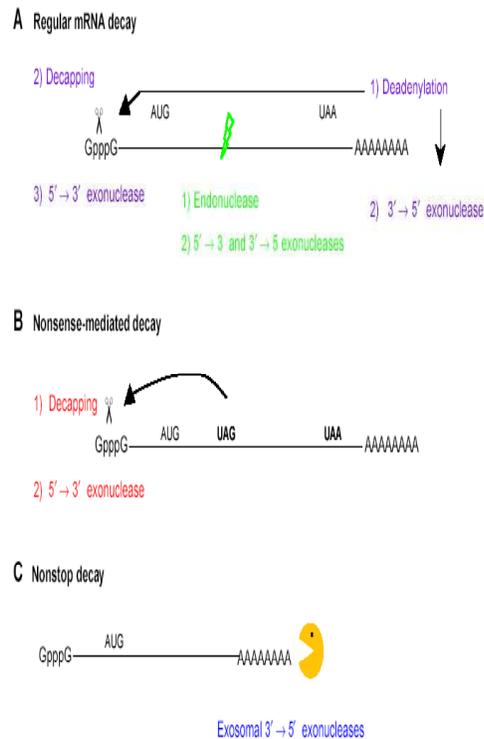


Fig. 6.: RNA turnover

mechanism some times coupled with alternative splicing to regulate gene expression (Fig. 6).

The nonsense RNAs are able to truncate prematurely the synthesis of one protein introducing a stop signal in the active producing cellular protein machinery. The surveillance is exerted through two functionally distinctive cellular mechanisms: a) nonsense-mediated RNA decay (NMD), b) nonsense-mediated altered splicing (NAS)<sup>[7]</sup> NMD seems to exert its function by degrading nonsense RNA before they release altered proteins; NAS seems to splice the new mRNA, working into the nucleus to excise the aberrant stop signal. Two proteins, called regulator of nonsense transcript 1 and 2 (*rent 1- rent 2*), are integrant part of the nonsense surveillance machinery. However it is not clear how the nonsense surveillance machinery processes the newly transcribed RNA (premRNA) in the nucleus before it is spliced in mRNA. This is in any case an economical natural control exerted into preventing the production of abnormal, not-functional proteins as well as enzymes. This mechanism occurs in numerous genetic

diseases as resulting from nonsense mutations and it has been widely studied in human iron metabolism. Today DNA microarray technologies enable investigators to monitor the expression of thousands of genes in biological sample as an useful tool to produce accurate molecular diagnosis of individual disorders in absence of clear clinical or biochemical diagnostic markers.

Normal RNA are degraded by different pathways: from 3' end by exosome, or decapped and degraded from 5' end. RNA can be cleaved by endonucleases and the resulting fragment can be degraded from both ends. Exosome degrades the termination-codon-less transcripts.

## 3) RNA interfering

Finally an additional mechanism to control gene expression is a different mRNA surveillance pathway linked to the mRNA degradation, so called RNA interfering – RNAi - (Fig. 7). This activity is elicited by a double stranded RNA (dsRNA) which, contacting an endogenous, homologous gene, is able to degrade it. Biochemical studies have suggested the involvement of amplification or propagation of RNAi signal in one stage of RNAi. In the first step cellular factors bound dsRNA processing it in 21-23 nucleotide pieces. One of these factors is a DICER, an RNase III-related protein. In the second step the cleaved RNA (siRNA) guides the recruitment of a degradation complex called RISC (RNA induced silencing complex) to a target mRNA, which is then cleaved by endonucleases. Since some of SMG (surveillance modulated genes) proteins affect RNAi, a connection between mRNA surveillance and RNAi has been suggested<sup>[8]</sup>. However the world of small RNAs just got bigger, since at this time at least two classes of small RNA produced by the ribonuclease Dicer, small interfering RNA (siRNA) and micro RNA (miRNA), have been characterized in the RNAi pathway depending on the extent of sequence

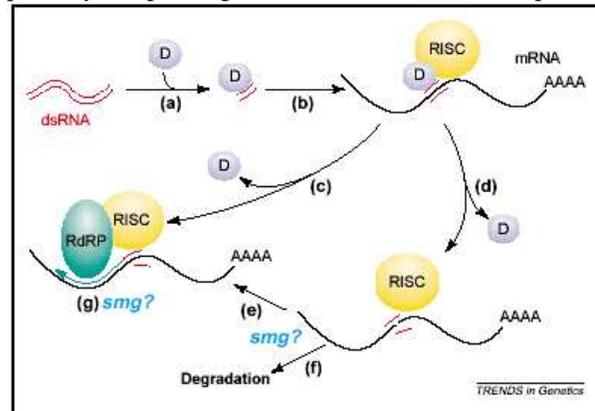


Fig. 7. A two-step mechanism for dsRNA mediated degradation.

complementary between the small RNA and its target<sup>[9]</sup>. These molecules have emerged as powerful post-transcriptional regulators of gene expression in many

different organisms, thus making the analysis of small RNA molecules increasingly important, both in vivo and in vitro, other than for their pervasive and essential role in animal and plant development, also for taking into account the power of RNAi to shut down expression of pathological genes<sup>[10]</sup>. The RNAi pathway consists of initiation and effector steps and usually functioning as a defense mechanism against invading nucleic acids, such as viruses and transgenes. First, the ribonuclease Dicer generates siRNA or miRNA from long dsRNA or short hairpin RNA (pre-miRNA). Second, the siRNA is incorporated into the RISC complex for sequence-specific mRNA degradation, whereas the miRNA is assembled into the miRNP complex to suppress translation of target mRNA. However siRISC and miRISC, as reported in figure 9 seem to be distinct complexes that regulate mRNA stability and translation respectively<sup>[11,12]</sup>. Probably many "riboregulators" miRNAs as well as siRNA, with a range of function must be yet identified and the challenge will be to define the potential targets and the exact functions of this fascinating small nucleic acid. In the Molecular Economy of the living organisms these five mechanisms, and probably other unknown by us, play an interesting role in the so-called enzymes induction. Thus, living cells also regulate the synthesis of their own catalysts, not only by producing quantitatively regulated amounts of constitutively expressed enzymes but also switching on/off the synthesis of an enzyme required in a date physiopathological moment to make a given product. Thus, for example, some of the enzymes involved in glucose metabolism may be induced by increasing the load of glucose and, in a similar way, some of the enzymes of the amino acid metabolism are inducible either by loading doses of the specific amino acid or by certain hormones. A clear example is provided by *E. coli* cultivated in the presence of lactose. Under these conditions, a series of enzymes required for the catabolism of that disaccharide, and normally repressed, are formed. The genes coding for these proteins are kept turned off through the action of a specific repressor protein with allosteric properties. The repressor binds to a specific site on the DNA and blocks transcription of the genes under its control. However an inducer such as an isomeric form of lactose, allolactose, may bind to an allosteric site of the repressor causing a decrease in the affinity of the repressor for DNA leading to derepression. Synthesis of most enzymes is apparently repressed most of the time. The appearance of specific enzymes at a particular time in the life of an organism or in certain differentiated tissues results from derepression brought about by the accumulation of specific metabolites or by other still unknown mechanisms. In this respect, it is suggestive that in eukaryotic cells control may be exerted both at the transcriptional and translational level. The phenomenon of induction is a very clear example of how much the

functional activity of enzymes may depend on the physiologic status of the tissue in which they are found. Moreover, these self-regulating properties allow cells to maintain a dynamic steady state adjusting themselves to the fluctuations of the environment in order to respect the principle of maximum economy. So the expression of enzymes is regulated by environmental factors representing the dynamic change of substrates.

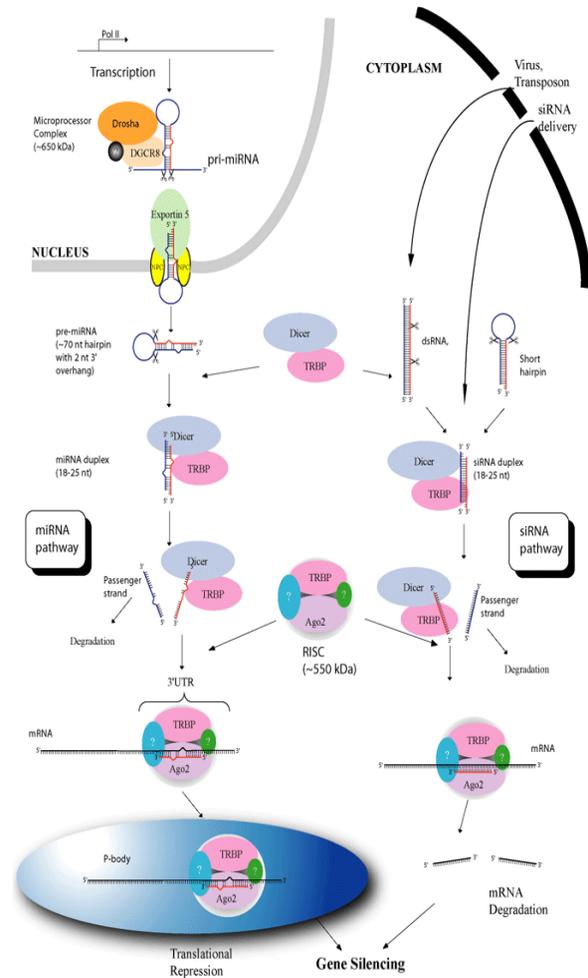


Fig. 9. The proposed siRNA and miRNA pathways (from: www.cell-research.com)

The graphical representation of all this is today possible with the utilization of the biochip technology, which constructed for the first time in 1996, have completely changed the methods of DNA and protein analysis. In fact by means of these new silicium frames has been possible to sequence more than 30 living organisms, *Saccharomices cerevisiae*, *Caenorhabditis Elegans* and *Drosophila Melanogaster* among the others. By photolithography it will be possible to collect in a slide of 1.5 x 1.5 cm specific sequences of DNA of 20-25 nucleotides or specific cDNA. In a single microchip one can store up to 12,000 sequences of gene. With such a method it will be soon possible to acquire and

compare data on the modification of expression levels of different metabolic networks, to establish the possible mathematical correlation among various genes differently modulated and to check in this way the pharmacological as well as nutritional target molecules of gene expression.

## II. The graph representation of a metabolic network

A cellular process can be represented as a metabolic network. The detailed modeling of each gene products' activity as function of the concentration of substrates and products is extremely difficult even for a small metabolic network because the enzymes and the other proteins usually exhibit a rich and extremely complicate dynamical behavior. A more manageable level of description consists in describing the network of metabolic reactions as a set of coupled stoichiometric equations. The solutions to the equations represent the admissible steady states for metabolic fluxes. An even more simplified approach which has given recently a series of interesting results, consists in neglecting all the details of the metabolic flux, and considering only the structure (the connectivity properties) of the metabolic network<sup>[1]</sup>. In such an approach the metabolic network is represented by a graph, a mathematical entity described by  $N$  nodes (also called vertices) and  $K$  connections (links, edges) between couple of nodes. In a metabolic network the vertices are the substrates (the chemical compound such as ATP, ADP, H<sub>2</sub>O) that can be either educts or products of a metabolic reaction, and the links between substrates are the actual metabolic reactions: two substrates are linked if they occur in the same chemical reaction, i.e. the only information retained is about the existence or the absence of a link, and nothing is known about the physics of the link. The physical entity of the link is the temporary educt-educt complex itself, in which enzymes provide the catalytic scaffolds for the reactions yielding products, which in turn can become educts for subsequent reactions. In this oversimplified representation the only role played by the enzymes is to connect or to not connect two substrate<sup>[1,13]</sup>. There are many advantages in such a description. First it is possible to use the same analysis tools already developed by mathematicians to investigate and characterize the structure of other complex networks, as social systems (a social system or a population can be imagined as a graph where the individuals are the vertices and the links are the social interactions among them), neural networks, and technological networks (computer networks, Internet). Moreover, the study of the connectivity properties of a network is the first step towards the comprehension of the dynamics and the history of the network (as the structure of a social network is important to study the spreading of information, fashions, rumors but also of epidemics, and the structure of a computer network affects the efficiency of the communication). In

particular, it has been found that the structure of a metabolic network is a fingerprint of the evolutionary process leading to an optimal organization of the cellular functions.

We now summarize the main results obtained by the analysis of the structure of metabolic networks.

1) *Small-world networks*. Fell and Wagner, 2000<sup>[13]</sup> have studied the metabolism of the *Escherichia coli* (involving about 300 substrates), and they have found that the properties of such metabolic network are neither those of a regular graph nor those of a random graph. In fact the metabolic network of the *E. coli*, as other (social, neural and technological) networks is better described by a new class of networks recently introduced by Watts and Strogatz, 1998<sup>[14]</sup>, the small-world networks. Small worlds are systems having on one hand a short distance between any couple of nodes (as random graphs), and on the other hand strong clustering properties (as regular graphs). The network of *E. Coli* is a small-world because it is indeed highly clustered locally and has an average distance between nodes (called the characteristic path length) slightly higher than 3<sup>[13]</sup>. What is the meaning of the small-worldness of the metabolic network of the *E. coli*? A characteristic path length slightly higher than 3, means that we can go from a node to any other node in an average of three steps, a very short path indeed, that implies, in the formation of the metabolic networks a need to minimize the transition time between metabolic states. In fact, as other examples of small-world networks, (like the collaboration graph of actors in feature films, the nervous system of *C. elegans*, the network of connections between cerebral cortical areas, and also the transportation systems of the Boston area), metabolic networks are systems efficient both on a local and on a global scale, i.e. networks following a perfect balance between local necessities and wide scope interactions<sup>[15]</sup>.

2) *Scale-free networks*. A different study of metabolic networks has been proposed by Jeong et al 2000 and is based on the calculation of the number of reactions in which every substrate is involved<sup>[16]</sup>. The authors have analyzed the metabolism of 43 organisms representing all three domains of life, and for all the 43 metabolic networks analyzed they have found that the probability  $P(k)$  that a substrate participate in  $k$  reactions is a power law distribution, i.e. a slowly decreasing function of  $k$ <sup>[17]</sup>. This means that metabolic networks as other social networks<sup>[18,19]</sup> are very heterogeneous networks because there are many substrates involved in only a few chemical reactions, but also a few substrates that play a key role since they participate to many reactions. Networks with a power law connectivity distribution  $P(k)$  are named scale-free networks<sup>[20]</sup> and are very important because, differently from random networks, display a high degree of error tolerance: the ability of

either nodes to communicate is unaffected by the failure of some randomly chosen nodes<sup>[21]</sup>. It is also interesting to note that the most connected substrates are practically the same for all the 43 organism. This indicate the generic utilization of the same very important substrates, the most highly connected metabolites (that should also be the philogenetically oldest), while the difference between organisms and species comes from the less connected substrates. The mathematical studies of metabolic networks we have described above, show clearly that, already from an oversimplified approach based only on the analysis of the *structure of the network*, we can draw important conclusions on the network's design principles. In particular the presence of the *small-world* and of the *scale-free* topology as the optimal structural organizations, is the fingerprint of an evolutionary selection of efficient and robust (error-tolerant) architecture that may characterize all cellular networks. It would be extremely important now to move one step further and to extend the mathematical description from the static study of the structure to the dynamics of the network, its response to the internal and external constraints. Of course one of the most important constraint on metabolic networks comes from the influences of the external environment. For this reason the next section we focus on a list of important examples of the effects of the environment on metabolic networks.

### III. Environmental influence on metabolic network

Accumulating evidence supports the hypothesis that the metabolic network may be influenced by external factors, which conferring a new image at the genomic expression. Nutrients can regulate the expression of genes at both transcriptional and posttranscriptional level. The family of enzymes involved in lipogenesis may represent a model system for understanding how a cell adapts to dietary energy intake in form of carbohydrate versus energy in the form of triacylglycerol. Most lipogenic enzymes undergo significant large changes in the rate of gene transcription in response to dietary changes. However glucose-6-phosphate dehydrogenase (G6PD), which is regulated at a step subsequent to transcription, is peculiar in this group of enzymes in that it participates in multiple metabolic pathways: reductive biosynthesis, including lipogenesis; protection from oxidative stress; and cellular growth. G6PD activity is enhanced by dietary carbohydrates and is inhibited by dietary polyunsaturated fats<sup>[22]</sup>. These changes in G6PD activity are a consequence of changes in the expression of the G6PD gene. This step is involved in the rate of synthesis of the mature mRNA in the nucleus, specifically regulation of the efficiency of splicing of the nascent G6PD transcript. Understanding the mechanisms by which nutrients alter nuclear

posttranscriptional events will help to design a new network of carbohydrate metabolism and will give additional information on the breadth of mechanisms involved in gene regulation. Following this observation the same group of authors<sup>[23]</sup> measured the amount of G6PD mRNA at different stages of processing in RNA isolated from the nuclear matrix fraction of mouse liver (this nuclear fraction contains nascent transcripts and RNA undergoing processing) and found, using a ribonuclease protection assay, with probes that cross an exon-intron boundary in the G6PD transcript, an excess of mRNAs that contains the intron (unspliced) and without the intron (spliced). Refeeding resulted in 6- and 8-fold increase in amount of G6PD unspliced and spliced RNA, respectively, in the nuclear matrix fraction. However, the amount of the unspliced fraction was at most 15% of the amount of the spliced one. During refeeding, G6PD spliced RNA accumulated at a rate significantly greater than unspliced RNA. Further, the amount of partially spliced RNA exceeded the amount of unspliced RNA, indicating that the enhanced accumulation occurs early in processing. Starvation and refeeding did not regulate either the rate of polyadenylation or the length of the poly(A) tail. Thus, the G6PD gene is regulated during refeeding by enhanced efficiency of splicing of its RNA, and this processing protects the mRNA from decay, a novel mechanism for nutritional regulation of gene expression. The authors conclude that the expression of G6PD gene during starvation and refeeding is regulated by a posttranscriptional mechanism occurring in the nucleus. But G6PD does not represent the only model for the influence that the alimentary habits may have on the metabolic network. The gene for cystathionine  $\beta$ -synthase (CBS) is located on chromosome 21 and is over expressed in children with Down syndrome (DS), or trisomy 21<sup>[24]</sup>. In 42 children with karyotypically confirmed full trisomy 21, Pogribna et al 2001<sup>[25]</sup> evaluated the impact of over expression of the CBS gene on homocysteine metabolism to determine whether the supplementation of trisomy 21 lymphoblasts in vitro with selected nutrients would shift the genetically induced metabolic imbalance. All metabolites involved in homocysteine metabolism were measured and compared to those of 36 normal siblings used as controls. Lymphocyte DNA methylation status was determined as a functional endpoint. The results indicated that plasma levels of homocysteine, methionine, S-adenosylhomocysteine, and S-adenosylmethionine were all significantly decreased in children with DS and that their lymphocyte DNA was hypermethylated with respect to that in normal siblings. Consequently plasma levels of cystathionine and cysteine were significantly increased, plasma glutathione levels were significantly reduced, reflecting an increase in oxidative stress due to the overexpression of the superoxide dismutase gene, also located on chromosome 21. When methionine, folinic acid,

methyl-B12, thymidine, or dimethylglycine were added to the cultured trisomy 21 lymphoblastoid cells the metabolic profile *in vitro* improved. The increased activity of CBS in children with DS significantly decreases the availability of homocysteine for the methionine re-synthesis and promotes well established "folate trap," creating a functional folate deficiency that may contribute to the neurological manifestations of this characteristic genetic condition. In a recent work it has been stressed how the homocysteine level in the plasma in subjects with DS does not always follows the number of gene implicated in its metabolism. Bosco et al 2003<sup>[26]</sup> found higher values in subjects with serious neural damages, where others have found lower levels according to the over expression of cystathionine  $\beta$ -synthase, also located in chromosome 21. These results suggest that the organization of a metabolic network can be influenced by a pre (CBS), post-transcriptional (G6PD) control, as already mentioned, but other mechanisms previously described can be able to influence the organization of a metabolic complex, where some specific nodes are the main hubs of the system. Historically, the theory by Lyon, demonstrating the activation of one of the two X at random, has been an important milestone for a control at molecular levels of something was, many years ago, thought to be only limited to chemical or receptor level.

#### IV. Conclusions

In light of the considerations reported here, a given protein molecule should be considered as one piece of a multidimensional structural and functional mosaic in continuous interaction with a number of other biological systems. The complexity of the picture is further enhanced by the time dependence of these interactions and the inter-relationships with the life span of the molecule and its degradation products. On the whole, the new emerging picture that we name transcriptomic (meaning the study of transcriptosoma), indicates a complex organization of the information contained in the DNA molecules, which is highly coordinated with others proteins, first in the interior of the cell and later in the transmission of information to other organism districts. With the help of biochips it will be soon possible to acquire and compare data on the modification of expression levels of different metabolic networks, to establish the possible mathematical correlation among various genes differently modulated and to check in this way the pharmacological as well as nutritional and environmental target molecules of gene expression. Thus being now clear that the human genome is only a static image of our genetic patrimony, the function of the cells, tissues and/or organisms are defined by proteins. Every cell of each one organism contain an equivalent set of genes, but not every cell expresses the same (quality and quantity) ones. Furthermore by each

gene it can arise multiple proteins, through alternative splicing, nonsense RNA regulation or trough post-transcriptional modification. In addition it is important to underline that the RNA abundance and steady-state protein levels are not correlate. This is the reason for which one of the most prominent area in life sciences is today the proteomics, that range from research in gene products to protein function (functional proteomics). In conjunction with micro array analysis and two-dimensional gel electrophoresis the proteomics revolution is becoming and researchers, using mass spectrometers are identifying proteins, sequence of peptides, posttranscriptional modifications and multi protein complexes. However by considering the information previous reported by us, it is reasonable to hypothesize the birth of pharmaco-transcriptomic and nutritranscriptomic or environment-transcriptomic era to better afford our knowledge in the complexity of life, first of all, of one cell, and to open the scientific view on the new dynamical map of different metabolic networks. In our rapidly moving from the genomic to the post genomic era, between genomic and proteomic, we think it is necessary to afford our hands and our minds in the study of transcriptomic to better clarify the mechanisms under posed to life sciences. But, at the same times, we are sure that only an interdisciplinary study of this three molecular mechanisms will enables us to reveal the main properties of the molecular networks. Only the integrated study of the transcriptomic and proteomic can guide the researchers versus a more complete comprehension of the internal (genetic) and external (environmental) mechanisms, involved into phenotype definition. Culture of keratinocytes have demonstrated<sup>[27]</sup> through the products of gene transcription the presence of unknown genes, the under expression of known genes and the over expression of housekeeping genes, which were important for epidermic differentiation. This should be considered the landmark of the birth of the new era of transcriptomic.

#### REFERENCES

1. Wagner A. and D.A. Fell, 2001. The small world inside large metabolic networks. *Proc Biol Sci.* 268: 1803-1810.
2. Kinoshita K., and T. Honjo, 2001. Linking class-switch recombination with somatic hypermutation. *Nature Reviews Molecular Cell Biology* 2: 493-503.
3. Viprakasit V., R.J. Gibbons, B.C. Broughton, J.L. Tolmie, D. Brown, P. Lunt, R.M. Winter RM, S. Marinoni, M. Stefanini, L. Brueton, A.R. Lehmann, D.R. Higgs 2001 Mutations in the general transcription factor TFIIF result in  $\beta$ -thalassemia in individuals with trichothiodystrophy. *Hum. Mol. Genet.* 10: 2797-2802.

4. Freson K., K. Devriendt, G. Matthijs, A. Van Hoof, R. De Vos, C. Thys, K. Minner, M.F. Hoylaerts, J. Vermeylen, C. Van Geet. 2001. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* 98: 85-92.
5. Nichols K.E., J.D. Crispino, M. Poncz, J.G. White, S.H. Orkin, J.M. Maris, M.J. Weiss. 2000. Familial dyserythrocytic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat. Genet.* 24: 266-270.
6. Yu C., K.K. Niakan, M. Matsushita, G. Stamatoyannopoulos, S.H. Orkin, W.H. Raskind. 2002. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood*, 100: 2040-2045.
7. Mendell J.T., C.M. Rhys, H.C. Dietz. 2002. Separable roles for *rent1/hUpf1* in altered splicing and decay of nonsense transcripts *Science* 298 (5592): 419-422.
8. Mango S. E. Stop making non sense: the *Caenorhabditis elegans* smg genes. 2001 *Trends in genetics*, 17: 646-652.
9. Tang G. 2005. siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* 30: 106-114.
10. Dykxhoorn D.M., J. Lieberman. 2006 Running Interference: Prospects and Obstacles to Using Small Interfering RNAs as Small Molecule Drugs. *Annu. Rev. Biomed. Eng.* Apr 17.
11. Kawasaki H., K. Taira, R. Wadhwa. 2004. World of small RNAs: from ribozymes to siRNA and miRNA. *Differentiation.* 72: 58-64.
12. Lee R.C., C.M. Hammell, V. Ambros. 2006. Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* 12: 589-597.
13. Fell D.A. and A. Wagner. 2000. The small world of metabolism. *Nature Biotechnology* 18: 1121-1122.
14. Watts D.J. and S.H. Strogatz 1998. Collective dynamics of small-world networks. *Nature* 393: 440.
15. Latora V. and M. Marchiori. 2001 Efficient behavior of small-world networks. *Phys. Rev. Lett.* 87: 198701.
16. Jeong H.B., R. Tombor, Z.N. Oltvai and A.L. Barabasi. 2000. The large-scale organization of metabolic networks. *Nature* 407: 651-654.
17. Podani J., Z.N. Oltvai, H. Jeong, B. Tombor, A.L. Barabasi and E. Szathmary. 2001 Comparable system-level organization of Archea and Eukaryotes. *Nat. Genet.* 29: 54-56.
18. Liljeros F, Edling CR, Amaral LA, Stanley HE, Aberg Y. The web of human sexual contacts. *Nature* 411: 907.
19. Pastor-Satorras R. and A. Vespignani. 2001. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.* 86: 3200-3203.
20. Barabasi A.L. and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286: 509-512.
21. Albert R., H. Jeong, A.L. Barabasi. 2000. Error and attack tolerance of complex networks. *Nature* 406: 378-382.
22. Salati L.M., and B. Amir-Ahmady 2001. Dietary regulation of expression of glucose-6-phosphate dehydrogenase. *Annu. Rev. Nutr.* 21: 121-140.
23. Amir-Ahmady B. and L.M. Salati. 2001 Regulation of the processing of glucose-6-phosphate dehydrogenase mRNA by nutritional status. *J. Biol. Chem.* 276: 10514-10523.
24. Skovby F., N. Krassikoff, U. Francke. 1984. Assignment of the gene for cystathionine beta-synthase to human chromosome 21 in somatic cell hybrids. *Hum. Genet.* 65: 291-294.
25. Pogribna M., S. Melnyk, I. Pogribny, A. Chango, P. Yi, S.J. James. 2001. Homocysteine metabolism in children with Down syndrome: in vitro modulation. *Am. J. Hum. Genet.* 69: 88-95.
26. Bosco P., R.M. Gueant-Rodriguez, G. Anello, C. Barone, F. Namour, F. Caraci, A. Romano, C. Romano, J.L. Gueant. 2003. Methionine synthase (MTR) 2756 (A --> G) polymorphism, double heterozygosity methionine synthase 2756 AG/methionine synthase reductase (MTRR) 66 AG, and elevated homocysteinemia are three risk factors for having a child with Down syndrome. *Am. J. Med. Genet. A.* 121(3):219-224.
27. Van Ruissen F., B.J. Jansen, G.J. de Jongh, P.L. Zeeuwen, J. Schalkwijk. 2002. A partial transcriptome of human epidermis. *Genomics* 79: 671-676.