Systems Biology and Systems Medicine: The Technological Tools of the System Approaches to Complexity

Alessandra Del Bufalo[1], Patrizia Russo[1]*, Mirta Milic[1,2], Christian Pristipino[3], Massimo Fini[4] and Alfredo Cesario[5]

[1]Laboratory of Systems Approaches and Non Communicable Diseases, IRCCS "San Raffaele Pisana", Rome, Italy
[2]Mutagenesis Unit, Institute for Medical Research and Occupational Health, Zagreb, Croatia
[3]Personalized and Systems Medicine Unit, San Filippo Neri Hospital, Rome, Italy
[4]Scientific Director, IRCCS "San Raffaele Pisana", Rome, Italy
[5]Deputy Scientific Director and Coordinator, Area of Systems Medicine, IRCCS "San Raffaele Pisana", Rome, Italy

Abstract

Progresses made in "omics" research have fueled the collection and storage of large quantities of data. Now these data can be used for knowledge gaining. Systems biology aims at studying biological organisms as interconnected systems. Systems Medicine, by applying systems biology approaches, tries to understand illness in the context of altered regulatory circuitry. Studies in a Systems Medicine framework will enable a better understanding of diseases, and the delivery of a personalized healthcare, tailored on the specific phenotype of each single patient. In this relatively new discipline, the review describes those tools useful in this approach.

The paradigm:
"disease → therapy → recovery"
is not still suitable and superseded by the new one:
"person → definition of the problems → quality of life".
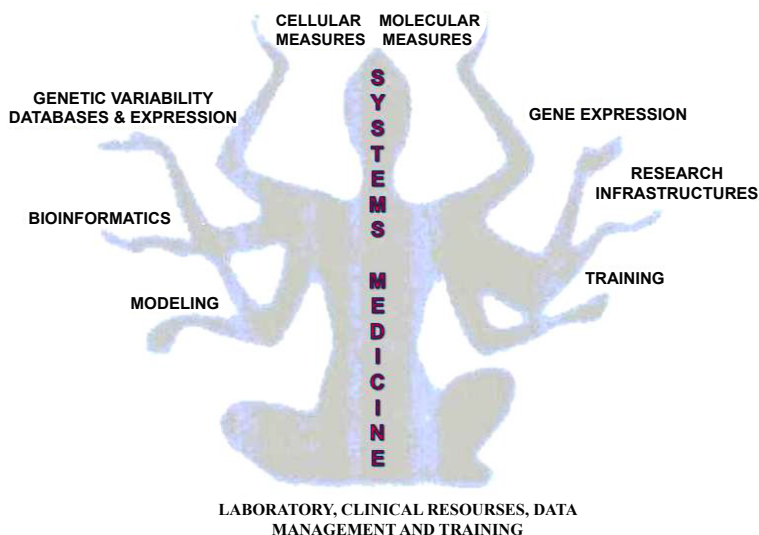
Keywords: Participatory medicine; Personalized medicine; Predictive medicine; Preventive medicine; Systems medicine

Introduction and General Definitions

Since the 19th century, the concept of health was changing, by moving from the concept of "*absence of illness*" to a "*state of complete physical, mental and social well-being*" [1]. The concept of "health" in medicine is now represented by a holistic person-oriented and self-quantification / wellness targeted approach instead of a disease-oriented vision. Thus, the paradigm: "disease → therapy → recovery" is not still suitable and is superseded by the new one: "person → definition of the problem → quality of life". This approach is so widely accepted that the Policy environment has adopted its potentialities (On. Beatrice Lorenzin, Italian Minister of Health, [2]). This new approach is particularly helpful in the case of the chronic conditions. Chronic conditions represent at present a global problem and its resolution will become more and more urgent in the near future. Indeed, the World Economic Forum has estimated that around 47 trillion of dollars will be spent on chronic illnesses in 2030 [3]. A "Systems" approach to Biology and then to Medicine, is the only one able to allow the understanding of the so called "Complex Phenotype" [4-6]. The Complex Phenotype is not a merely additive sum of typical phenotypes that are linked to every single possible competitor in the same individual pathology, but it is the result of the integration and interaction of all components that characterize the "Person". It is important to consider the obvious bias introduced by the traditional categorical ontologies (diagnoses) that are, currently, based on "obsolete" criteria, some of which more than

*Corresponding author: Patrizia Russo, PhD, Chief, Laboratory of Systems Approaches and Non Communicable Diseases, IRCCS "San Raffaele Pisana" Via di Valcannuta, 247, I-00166 Roma, Italia, Tel: 393483339704; E-mail: patrizia_russo@hotmail.it, patrizia.russo@sanraffaele.it

one-century old. This bias is particularly evident in the management of Chronic Non-Communicable Diseases (NCDs) as such as cancer, heart failure, chronic obstructive pulmonary disease (COPD), diabetes, HIV and neurodegenerative diseases [7].

Systems Medicine (SM), actually, includes information stemming from both the cellular and molecular domains, and information from the physiological and clinical domains themselves. This "holistic approach" takes into account the "human person", instead of the individual pathology/pathologies (co-morbidities). By using, among others, computational biology, medical genomics and medical informatics, SM permits the medicine to be not only "Person-centric" but "predictive" and finally "Preventive" and "Participatory" because of "Person-centered" P4 medicine [8,9].

At international level, the research funding programs (i.e. the framework programs of the European Commission) invest significant resources in this field, based on the assumption of a great potential of this holistic approach for the management of optimization processes in Health Systems. This review will describe tools that are used in this relatively new discipline.

From Laboratory, Clinical Resources and Data Management to Training and Education

In the last two decades there has been an unprecedented increase of possibilities offered by "high throughput" technologies in terms of quality and quantity (as well as optimization) and the big boom started with invention of DNA-chips back in the 1990s. To understand the level and importance of these technological developments, it is sufficient to remind that the whole human genome sequencing required about one billion of dollars and it was extremely time consuming even if the challenge was undertaken by an entire community of scientists. Nowadays, we are able to achieve the same result much faster (within days) and with a significantly lower cost by the use of the Next Generation Sequencing methods-NGS-(see detail in the following paragraphs).With this huge amount of progress, soon we will be able to have results of sequencing in a few minutes, with a cost of just around 1000 $, with a reduction of time and costs by the magnitude of more than six times. The evolution of systems measurement (and complex data integration), applied and adapted to health-sciences research, allowed to discern the huge genetic and epigenetic heterogeneity of, for example, neoplastic cells [10,11]. Systems measurements can be now functionally implemented in all areas of NCDs [7]. The final goal is to integrate the experimental evidence (genome, proteome, metabolome, transcriptome, etc) with clinical-epidemiological data to generate predictive models that can – up to the level of the individual patient – serve as efficient stratification tools for prevention, diagnosis and treatment optimization [12]. Systems Medicine (SM) is indivisibly linked to technological developments, which made possible to create models that support the complexity of the clinical picture due to complexity of the human genome and to encompass and partially understand the genetic and epigenetic variability (or, at least, its intrinsic level of complexity). The knowledge of the foundations of the main multidimensional assays developed in these years constitutes the essential vocabulary for those who wish to deal with SM approaches [12].

Molecular and cellular measurements

DNA genomic techniques: The technological process underpinning the functioning of DNA microarrays is now in the public domain [13]. Basically the whole process of detection consists of sample DNA hybridization to the short nucleotide "probes" that are already attached to the DNA chip. When there is a "match" between the sample of the DNA (or single-stranded cDNA) with the probe, after washing process, the matched couple is identified with appropriate targets associated with detectable reading systems (fluorophores, chemiluminescence, etc.).

The large variety of DNA microarray systems allows easily measuring changes in the expression of genes, genotyping of single polymorphisms (single nucleotide polymorphisms-SNPs) or counting and characterization of copies of concerned DNA segments (i.e. copy-number variations) [13]. In addition to functional analyses of proteins (through methods of over-expression or silencing) the DNA microarray is used for the characterization of gene signatures through bioinformatic analysis of the expression of thousands genes in tissue samples. This is particularly important, for example, in the field of oncology where, in addition to the direct causality analysis, the characterization of these genes allows the understanding of tumor-specific features such as chemo-resistance, patterns of progression and recurrence, etc. [11,14]. In some cases the correlation of expression patterns of gene signatures with prognostic stratification of items is so solid (for example in some types of breast cancer) that is now (mandatorily) used in classification purposes in normal clinical practice [14]. A methodology that relies on DNA Microarrays is the so-called Comparative Genome Hybridization (array-CGH) where the DNA sample for testing is placed in a competition with a reference DNA sample to compete for the binding on the probe on the DNA chip. This method is widely used to identify genomic microdeletions and duplications in carriers of congenital abnormalities or to detect copy number variations in cancer samples [15,16]. Since the results of the DNA sequencing and gene expression analyses can be achieved nowadays really rapidly, applying of these new methodologies has inspired the launch of "global" project such as the Worldwide Personal Genome Project (PGP) and large scale projects geared to fully understand the phenomenon of genomic variability.

The technology behind the NGS relies on two solutions: 1) the simultaneous use of millions of "spatially fixed DNA fragments for testing" and then 2) parallel analysis on a large scale. The events of "coupling", identified through polymerase activity or fluorescence, are turned into "digital" events and then with reference sequences hybridized and analyzed with bioinformatics tools. Amplification and coupling are the processes that determine cost and durability of experiments although, in anticipation, these will be reduced to a few thousand dollars per genome (one thousand dollar genome) and a few hours. This is likely to be possible thanks to the elimination of amplification and coupling processes and then by direct sequencing of single DNA molecules [known as "nucleotide-triphosphate methods-polymerase binding" [17], "nanopore sequencing" [18,19], and "scanning probe sequencing" [18] whose description goes beyond the scope of this paragraph. The NGS provides efficiently genome-wide analysis and then virtually identifies all possible polymorphisms. However it is a long and expensive process that cannot be routinely performed. Methods such as the microarray-based genomic selection (MGS) or multiplex exon capture or bead-based enrichment allow us to focus on areas of specific interest to actual pathology (or group of diseases). Technical specifications of DNA enrichment are near to be finally completed. The ability to enrich the areas of specific genome sequencing will make this approach much more accessible for large-scale diagnostics [20,21]. Genome-scanning technologies revealed a large extent of so-called 'structural variations' or "genomic structural variations" [22] that include microscopic and submicroscopic variants (i.e. deletions, duplications copy-number variants or copy-number polymorphisms, insertions, inversions and translocations). Redon et al. [23] defined copy-number variants (CNV) "*as a DNA segment of one kilobase (kb) or larger that is present at a variable copy number in*

comparison with a reference genome". These alterations are important, especially in those processes related to cancer. The understanding of the importance of the CNV (not only in the neoplastic disease) has led to the development of methods based on the NGS better suited for specific measures of this kind of alteration (in the history they were first performed through microscopic analysis of chromosome banding, after with PCR techniques, then FISH and classical or CGH microarray). The Database of Genomic Variants (DGV) [24], updated to 2014 by Mac Donald et al. [25] allows a public accessible, comprehensive catalogue of structural variation (SV) found in the genomes of control individuals from worldwide populations. The introduction of specific advanced algorithms has also allowed the effective analysis of so-called chimeric sequences -sequences that are at the native state in different parts of the genome- and before a specific coupling need to be splitted [26].

Epigenetic changes play a pivotal role in the explanation of how the genetic heritage, represented by the nucleotide sequence, is transmitted from generation to generation. Cytosine methylation is an epigenetic change marker (necessary for the regulation of gene expression or their silencing). Alterations of the physiological methylation process (which does not alter the ability of cytosine pair) have demonstrated important role in development of many diseases and not just in the neoplastic ones, where methylation has been shown as one of the earliest events in cancerogenesis [27]. The prevailing analysis methods for detecting DNA methylation are the reaction of bisulfite conversion or analysis of methylation-specific restriction. The application of techniques of NGS sequentially to the above methodologies now makes possible the analysis of methylation patterns with temporal advantage and substantial economic relation to the classical methods (similar to the analysis of gene expression itself). However, genome wide analysis of this kind is still too expensive for large-scale use, so the technique is still used only for characterizations of areas of interest, suitably enriched [28]. Classical methods of DNA sequencing did not allow the identification of "new" transcribed genes. NGS methods allow, through reverse transcription of mRNA into cDNA, a sequencing that can allow not only the analysis of gene expression but also the identification of new transcribed genes and then facilitate the discovery of new variants "splice" [29].

Trascriptomics sequencing, already carried out on the whole transcriptome of some models (*Drosophila, Arabidopsis*, etc.) and on some cell lines, it is somehow burdened by the loss of the temporal component (loss of the ability to reconstruct patterns of cell kinetics). The use of so-called "nascent" RNA (just created) would allow the overcoming of this limit *via* serial analysis of multiple samples collected in the following moment, allowing the possibility to generate real time function curves. The NGS has also enabled large-scale analyses, through a mechanism of reverse transcription, and then sequencing, of so-called micro-RNAs (miRNAs) that have recently proven to be key regulators in determining cellular evolution. Classical transcriptomics analysis (mRNA) associated with miRNA sequencing allows identification of functional characteristics of RNA and miRNA-target coupling through a quantitative and qualitative analysis and characterization of miRNA recently discovered [30].

Mitochondrial genome sequence: Mitochondrial disorders are a group of metabolic disorders characterized by impaired mitochondrial function that are associated with age and age-related complex disorders. The mitochondrial genome can modulate the clinical severity *via* different mechanisms including: *(i)* inherited mitochondrial genomic haplogroup background, *(ii)* nuclear-encoded gene variants that act together with specific mitochondrial haplogroup backgrounds and *(iii)* variation in the number of mtDNA copies present in a cell [31].

There is a dual genomic expression of mitochondrial proteins from both nuclear (nDNA~1500 genes encoded in the nuclear genome) and mitochondrial DNA (mtDNA of 16.6 kb). The database of mtDNA mutations reports the mutations that have been reported in the literature [32]. The next-generation sequencing of the exome (targeted exome capture) allows to selectively sequence the coding regions of the genome. Comprehensive one-step analysis of the whole mitochondrial genome long range/massively parallel sequencing (LR-PCR/MPS)-based approach using one pair of non-overlapping, back-to-back primers, allows the amplification of the entire mitochondrial genome as a single amplicon [33]. This method allows the simultaneous detection of point mutations with quantified heteroplasmy as well as large mtDNA deletions with accurately mapped break points. The method is rapid and accurate allowing the sequence of large amounts of DNA at lower costs. Recently, a DREAM PCR coined De´ ja` vu PCR using a so called "DNA diode" ensures complex serial amplification steps contamination free utilizing enzymes that specifically digest 5th and 6th bases respectively providing methylation dependent mitochondrial DNA enrichment [34]. The utility of targeted exome sequencing and whole genome sequencing is currently under evaluation.

Since the human mitochondrial genome is relatively small, the sequencing of its transcriptome provides a valuable method to investigate changes in RNA metabolism. Directional RNA sequencing (strand specific) enables the analyses of RNA abundance from each strand of the mitochondrial DNA. Parallel analysis of RNA ends (PARE) enables the analyses of processing of mitochondrial transcripts, their termini, and annotation of any new transcripts [35].

Proteomics and metabolomics

Proteomics is the large-scale study of proteins, of their structures and functions. The data from the proteomic analysis integrate the molecular and cellular measures that are fit to the tasks of bioinformatics and mathematical modeling. Immune-histochemistry (IHC) is the most classical assay used in the approaches of SB and SM where the tissue-state information of large cohorts of patients is necessary. IHC uses the principle of antibodies binding to antigens in biological tissues to detect the antigens (i.e. proteins) in cells of a tissue section. The antibodies may be labeled with fluorescent dye, peroxidase, colloidal gold, alkaline phosphatase and radioactive elements. The use of a laser-capture microdissection (LCM) method allows the analysis of tissue cells subpopulations under direct microscopic visualization. IHC also permits differential analysis of simple and immediate execution in comparing healthy and pathological tissues. Molecular methods to detect proteins and their state (i.e. phosphorylation) have been also developed. The 2D (two-dimensional) gel electrophoresis allows the separation within complex mixtures of proteins (in the order of 1000 to 10000) using as discriminants molecular weight and electric charge. Isoelectric focusing (IEF) and sodyum-dodecyl sulfate and poliacrilamide gel electrophoresis (SDS-PAGE) are the two sequential methods (upgradeable with the differential in gel electrophoresis-DIGE) that allow this separation. Separate "proteins" through "D electrophoresis" are then identified with mass spectrometry [36]. Mass spectrometry (MS) has enabled major progress to identify relevant biomarkers in different processes, all relevant to diagnostic and therapeutic purposes: *(a)* whole proteome analysis); *(b)* identification of post translational modifications (PTM); *(c)* analysis of the quantitative effect of modulation of signaling; *(d)* details of protein-protein interactions (PPI) for the characterization of relevant pathways; *(e)* identification of pathology-specific metabolites [37].

In terms of extreme simplification, proteomics studies can be divided into "discovery approaches" for the discovery of new markers

and quantification/ validation approaches (quantitative analysis of molecules). For the first approach, reference machines are tools that allow measurement of molecular mass with great precision (MALDI/ TOF/TOF, qTOF and LTQ Orbitrap), while the second approach uses tools that are "mass selective" (triple quadropole MS-QQQ MS) with ability to measure quantitatively the molecules of a given mass, by pulling them from complex mixtures (which may be the biological fluids from laboratory tests used for diagnostic purposes). In particular, this type of measurement (quantitative) is particularly relevant to SB approaches. The method of reverse phase protein array (RPPA) identifies not only the quantitative expression of a protein present, for example, in a tumor cell lysate, but also the degree of phosphorylation of the protein being investigated. This qualitative and quantitative approach is used in studies of SB. The absence of a large number of necessary antibodies to detect proteins and their phosphorylated forms is the only bias at the time for this approach that is very promising, for example, in the assessment of the therapeutic effects in the field of oncology [38]. Especially in this field, in fact, the knowledge of the state of phosphorylation of proteins in cell tissue is of particular importance. Some specific enhancements of this methodology were used to improve the quali-quantitative analysis of the phosphorylated proteome, and were performed with the approach described in the 2D electrophoresis where the antibody probe is specifically directed to the phosphorylated proteins. Before performing the classical analysis in MS, it is possible to use innovative enrichment methods such as the PAC (phosphoramidate chemistry), the IMAC (immobilizer metal affinity chromatography) or phospho-specific protein treatments (titanium dioxide). The systematic identification of phosphorylation sites in proteins is of crucial importance for the understanding of disease mechanisms, especially for the purposes of the identification of potentially therapeutic molecules. Systems approaches have allowed the creation of libraries of phosphorylation sites, many of which are derived from the bioinformatic analysis of mathematical models [39].

Metabolomics allow the global assessment of metabolites in a biological sample and reports the closest information to the phenotype of the biological system of interest. The biochemical (and physiological) analysis of the cellular behavior, read through a quantitative analysis of the metabolic assets, gives us important information relevant in understanding the events of deregulation that insist in cell pathological stages [40]. Classical methods as such as the gas (GC-chromatography-gas) and liquid (LC) chromatography and new technologies as MALDI MS technologies, including MALDI-TOF MS, MALDI-TOF/TOF MS, SALDI-TOF MS, MALDI-QqQ MS, and SELDI-TOF MS, may be utilized in discovering and quantification of metabolites. MALDI, SALDI, SELDI and PFM are deemed complementary to each other and, theoretically, all of these technologies could be used in combination, leading to the next generation of the MALDI MS technologies [41]. Cell metabolomics is an emerging field that allows the observation of the metabolic phenomena in cells. Four sequential steps are used for cell metabolomics analysis; they include: *(1)* sample preparation and extraction, *(2)* metabolic profiles of low-weight metabolites based on MS or nuclear magnetic resonance-(NMR-) spectroscopy techniques, *(3)* pattern recognition approaches and bioinformatics data analysis, *(4)* metabolites identification resulting in putative biomarkers and molecular targets [42].

Summary: The techniques of the genome (nuclear/mitochondrial), proteome and metabolome analysis allow us to obtain data supporting phenotype/genotype conclusions as well as patients stratifications and potentially appropriate treatment. Table 1 reports examples of clinical applications of the listed methods [43-71], with no reported examples regarding cancer, since they are largely discussed in overall literature [10-12,72].

Functional studies

Systems Biology/Systems Medicine need "systems" studies of gene function at the level of the whole cell (so-called genetic screens or

Methods	Pathology	References
DNA genomic or mitocondrial techniques		
DNA Microarrays	Juvenile arthritis and spondyloarthropathy Diabetic nephropathy Lupus Human stent-induced neointima	[43-46]
array-CGH	Endometriosis Autism, epilepsy, and schizophrenia	[47,48]
NGS	Retinitis pigmentosa Familial cardiovascular diseases Dementia Type 2 diabetes	[49-52]
CNV	Progression of Chronic Kidney Disease Interstitial Lung Abnormalities Valvular Calcification and Aortic Stenosis Risk of Alzheimer's Disease Rheumatoid arthritis Elevated or reduced risk of type 2 diabetes	[53-58]
DNA methylation	Osteoarthritic disease Alzheimer's disease Dilated cardiomyopathy Diabetes mellitus type 1	[59-62]
Comprehensive one-step analysis of the whole mitochondrial genome	Mitochondrial dysfunction and complications associated with diabetes Disease progression in patients with single, large-scale mitochondrial DNA deletions	[63,64]
Proteomics and metabolomics		
2D-electrophoresis	Endometriosis Amyotrophic lateral sclerosis (ALS) Anderson-Fabry disease.	[65-67]
MALDI/TOF, SELDI/TOF	Systemic Sclerosis COPD	[68,69]
Nuclear magnetic resonance spectroscopy and mass spectrometry	Metabolic syndrome and fatty liver disease Multiple sclerosis	[70,71]

Table 1: Examples of clinical applications of DNA genomic/mitochondrial techniques.

through RNA interference experiments). These studies can associate individual genes/individual proteins with mathematical modeling approaches. Information on the interaction between proteins (protein-protein interaction-PPI) are fundamental elements for the network analysis underpinning systems approaches. There are many techniques now used to identify PPI mappings in humans. Some of these, high-throughput type (protein array), have made possible to analyze the protein-DNA interactions also as a function of time, a multidimensional approach to screening, necessary for the mathematical modeling in SB.

RNAi is a general method for inhibiting gene expression in organisms. RNAi is a process in which RNA molecules inhibit gene expression. Post-transcriptional gene silencing (PTGS) is the term used to indicate RNA interference (RNAi). Briefly, dsRNA introduced into a cell is cleaved by the ribonuclease III, Dicer, into 22-25 nucleotide fragments. The siRNA strand with the most unstable 5' end is incorporated into the RNA-induced silencing complex (RISC) that drives the cleavage of homologous mRNAs. This technology combined with the ability to perform robotic large-scale analysis (so-called high content analysis) makes this method particularly powerful [73].

Cellular, tissue and animal models, biobanks

The upgrading of treatment to "tailored" or "targeted" is certainly dependent on the availability of well characterized cell lines to be used, essentially, as the initial model (i.e. for use during early experiments) in the development of new pharmacological approaches. "Model" organisms represent an evolution in relation to cell lines. These tools are tissue banks that shall be annotated with homogeneous and high quality clinic informations (necessary and fundamental for biomarkers studies, characterization of pathways and molecular networks). All these components are important as "custom" in SB/SM approach to diseases' understanding and treatment [74]. The integration of information (mostly from experimental evidence and of observational nature) from models and human tissues (biobanks) represents a significant challenge. The fundamental work of Hanahan and Weinberg [75,76], regarding the "hallmarks" of cancer, however, has shown that (at least in cancer) it is possible to group the information in a limited number of "strokes" (traits/hits) that cells must acquire to become "malignant". Recently the same rationalization has been proposed for ageing, by grouping informations in a limited number of "three" categories that include nine hallmarks: primary, antagonistic, and integrative [77]. This categorization may facilitate the construction framework for future studies as well as the interventions design. A "systemic" understanding level requires a change in the way we perceive (in basic methodology and investigative power) biological approaches. While, in fact, the understanding of the role of the individual genes or proteins remains important, it is therefore necessary to shift the focus on the dynamics of pathological ontology intended as "system". We must however keep in mind that this type of integrative approaches, where you need both the biological component, as computational and clinical are now possible only thanks to the evolution of tools such as mobile models, animals and Biobanks, as well as the ability to have "omics" information as quickly as possible and at a sustainable cost.

Databases of genetic variability and expression: The "high-throughput" technologies produce a wealth of data that is not only growing in size but also increasingly complex and lays a challenge not indifferent as to how data is generated and analyzed in an effort to locate clusters of significance. In this section we will describe briefly how this data stream is organized in databases that are managed with large scale methodology to organize data so they are useful approaches to SB/SM. Depending on whether the database of genetic variation and gene expression data are, the data can be organized and stored in different

ways. The challenge here is to make them always available for studies that exceed, for example, the ontological classic approach (typical of genetic variance analysis, centered on a single reference pathology). The basic concept of information in the SB/SM studies lies in the ability to connect, with advanced computational methodology, the genetic variants to a single pathology or, better yet, to any phenotype. The information available (which is roughly as reference for further development) can be synthetically classified in:

(i) Database of SNPs [78];

(ii) Databases of structural variants (these variants are not systematically "collected" as the SNPs). Here are the most significant of them in terms of size and update status:

-Database of genomic Variants [79];

-NCBI database for structural variation [80];

-European Bioinformatics Institute-Database of Genomic Variants EBI Archive 7DGVa [81];

-Sanger center Database [82];

(iii) Database for "disease-causing variants": it is commonly believed that these resources are still not optimally managed primarily for information sharing that clarify the correlation of genetic variants and monogenic ontologies. The most significant ones include:

-The Online Mendelian Inheritance in Man [83]. Data collected from the 1960s, not only on monogenic ontology but also on poligenic-type diseases;

-The Human Genome Variation database of Genotype-to-Phenotype information [84]: implementing clinical information-epidemiological from large-scale association studies;

-The Human gene Mutation Database [85], of a commercial nature;

-The Catalogue of Somatic Mutations in Cancer [86]. In 2011 had collected data about 1.7 million of 500,000 cancer experiments.

It is useful, however, to remember that there is a multitude of locus-specific databases that are increasingly connected (also for sustainability issues) to those listed above.

(iv) Large scale repositories: we mention briefly the European genome-Phenome Archive [87] and GAP (dbGAP) database at NCBI which are the two primary resources for information exchange on large-scale genotyping data associated with phenotypic information.

(v) Reference Genomes: after the first sequence of the entire human genome release in 2001, the knowledge of genetic variability has deepened considerably. "Population studies" were made, increasing levels of genotyping tests up to 1000 genome "type project" [88]. Information in this database are definitely useful for any integrative multiscale approach.

Gene expression: Since the end of the '90s, the alterations of gene expression were investigated in order to connect some variation patterns to specific phenotypic expressions. Data for a large number of binding experiments are now in various shapes, arranged in publicly available databases and others in public or commercial connotations (the latter contain relevant data of gene expression linked to specific experimental activities). The availability of these data has pioneered the personalized annotation result of "private" experiments. This continuous exercise thus gave rise to the repositories in "added value". In general, the "added value" is represented by the fact that the original data are transformed in more user-friendly formats. Two are of particular importance: Oncomine [89] and the Gene Expression Atlas

[90] in addition to some commercial types which will not be described in this paragraph.

Research infrastructures and training

In recent years there has been an increase in capacity to acquire and interpretate experimental data, especially from the omics technique. This can only be addressed through approaches that involve the integration of methodologies that come from historically separated fields such as computer science, biology, chemistry and clinical sciences, to convert together their methodologies into systems biology and systems medicine. It is clear that, apart from the technical obstacles, the proper methodological training of professionals hired in these approaches is itself a challenge. Educational tools, are necessary for understanding the tools that are used for integrated multiscale approach, typical of the systems biology and medicine.

Bioinformatics and modeling

The description of mathematical and bioinformatics tools used in approaches of SB and SM is certainly not the purpose of this review. We'll just describe the mechanism/procedure with which it is built and tested as a model of a biological system. The application of bioinformatics tools such as cell signaling mechanisms must necessarily follow the stop procedure that, in a context of extreme synthesis, can be listed as follows:

- Collected data, in the type of the databases described in the previous paragraphs, are used to identify the molecules (proteins) that are involved in a particular type of interaction and particular pathway. Of course, even the most refined biomedical information collected will never complete the level that would be sufficient to build a perfect model. Because of that reason, this approach will also include assumptions mainly related to protein networks and their dynamics;

- The information is then analyzed and organized to provide the blueprint for a pattern and that, as a rule, consists of a set of mathematical equations that are then processed to generate simulations and analyses for detailing the properties of the template "system" and then suggest experiments to transpose from *in silico* to *in vitro*, potentially actually optimizing them;

- The experiments provide informations that are valuable to refine the model. The merging of mathematical and experimental data generates a model calibration process, interactive and continuous, that allows the model to evolve and therefore be available for further analysis and simulations;

- At this point, the model is used to predict the outcome of *in vitro* experiments based on evidence of elaborate simulations on the model itself. The result of this exercise gives the model a predictive value on predictive capacity. The model is then "validated" to know its predictive power.

- A validated model itself is already a very important result because it gives the possibility to verify intervention hypotheses, eventually to be applied in early phases of the pathological process. This can be helpful in selecting targets for pharmacological screening of new molecules when there is no-sense in the use of time and resources with *in vitro* experiments.

A template created in a context of SB would appear as the appropriate approach to address the complexity, given that upstream input is as complete as possible. The complexity can be intrinsic to the system (number of components and interactions considered), or to be connected to multiple space-time scales that are necessary for a better definition and then derive also from perturbations of the system itself

(a complex biological system affected by a disease, for example). We can conclude this paragraph citing at least four examples where a SB approach is deemed appropriate, if not exactly a real requirement:

- Experimental design and validation of assumptions regarding the dynamics, the spatial organization and the structure of biological pathways;

- Study of biochemical networks highly interconnected and integration of data from high-throughput technology;

- Identification and analysis of non-linear dynamics in cellular signaling systems;

- Identification of biomarkers and therapeutic targets based on templates.

It remains, of course, the challenge represented by the "scale" models that poses serious challenges to the computational type.

Perspectives and Conclusions

Measuring methods used in the field of molecular and cellular biology have gone through a huge technological development. In particular (but not only), the output capacity in the field of sequencing has increased by several orders of magnitude and therefore time and costs are significantly reduced. This process has contributed strongly to the development of the discipline of SB, and, therefore, has put its bases in all those technologies that enable genome-wide, DNA, protein, mRNA and metabolite analysis. The SB approaches (technologies, modeling) will greatly contribute to the collection of informations relevant to the identification of new therapeutic targets and therefore innovative approaches to diagnosis and therapy that, if properly interpreted, in turn, will enable the application of a systems approach to a real Personalized Medicine. We can say with confidence that it is now possible to completely analyse the genome and transcriptome of each individual patient. Data from this analysis, if integrated with clinical data, proteomics and metabolomics may give rise to a process whose result is a potentially useful reliable model, for example to determine the response and toxicity profile of a patient to a certain medication or combination of medications.

References

1. Constitution of the World Health Organization. WHO.

2. Lorenzin B, Quaderni del Ministero della salute. n. 23, settembre-ottobre 2013 ISSN 2038-5293.

3. The World Economic Forum.

4. Hood L, Balling R, Auffray C (2012) Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J 7: 992-1001.

5. Hood L, Flores M (2012) A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. N Biotechnol 29: 613-624.

6. Hood L, Tian Q (2012) Systems approaches to biology and disease enable translational systems medicine. Genomics Proteomics Bioinformatics 10: 181-185.

7. Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, et al. (2011) Systems medicine and integrated care to combat chronic non communicable diseases. Genome Med 3: 43.

8. Hood L, Friend SH (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 8: 184-187.

9. Cesario A, Auffray C, Russo P, Hood L (2014) P4 Medicine Needs P4 Education.

Curr Pharm Des.

10. Mwenifumbo JC, Marra MA (2013) Cancer genome-sequencing study design. Nat Rev Genet 14: 321-332.

11. Offit K, Bradbury A, Storm C, Merz JF, Noonan KE, et al. (2013) Gene patents and personalized cancer care: impact of the Myriad case on clinical oncology. J Clin Oncol 31: 2743-2748.

12. Chen R, Snyder M (2013) Promise of personalized omics to precision medicine. Wiley Interdiscip Rev Syst Biol Med 5: 73-82.

13. Papp K, Szittner Z, Prechl J (2012) Life on a microarray: assessing live cell functions in a microarray format. Cell Mol Life Sci 69: 2717-2725.

14. Pinder SE, Brown JP, Gillett C, Purdie CA, Speirs V, et al. (2013) The manufacture and assessment of tissue microarrays: suggestions and criteria for analysis, with breast cancer as an example. J Clin Pathol 66: 169-177.

15. Boyd SD (2013) Diagnostic applications of high-throughput DNA sequencing. Annu Rev Pathol 8: 381-410.

16. Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61: 437-455.

17. Hollenstein M (2012) Nucleoside triphosphates--building blocks for the modification of nucleic acids. Molecules 17: 13569-13591.

18. Lund J, Parviz BA (2009) Scanning probe and nanopore DNA sequencing: core techniques and possibilities. Methods Mol Biol 578: 113-122.

19. Thompson JF, Oliver JS (2012) Mapping and sequencing DNA using nanopores and nanodetectors. Electrophoresis 33: 3429-3436.

20. Shen P, Wang W, Krishnakumar S, Palm C, Chi AK, et al. (2011) High-quality DNA sequence capture of 524 disease candidate genes. Proc Natl Acad Sci U S A 108: 6549-6554.

21. Dufva M, Petersen J, Poulsen L (2009) Increasing the specificity and function of DNA microarrays by processing arrays at different stringencies. Anal Bioanal Chem 395: 669-677.

22. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7: 85-97.

23. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 444–454.

24. Database of Genomic Variants (DGV).

25. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 42: D986-992.

26. Louhimo R, Lepikhova T, Monni O, Hautaniemi S (2012) Comparative analysis of algorithms for integration of copy number and expression data. Nat Methods 9: 351-355.

27. Kaelin WG Jr, McKnight SL (2013) Influence of metabolism on epigenetics and disease. Cell 153: 56-69.

28. Bergman Y, Cedar H (2013) DNA methylation dynamics in health and disease. Nat Struct Mol Biol 20: 274-281.

29. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY (2011) Understanding the transcriptome through RNA structure. Nat Rev Genet 12: 641-655.

30. Esteller M (2011) Non-coding RNAs in human disease. Nat Rev Genet 12: 861-874.

31. Horan MP, Cooper DN (2014) The emergence of the mitochondrial genome as a partial regulator of nuclear function is providing new insights into the genetic mechanisms underlying age-related complex disease. Hum Genet 133: 435-458.

32. Database of mtDNA mutations.

33. Cui H, Li F, Chen D, Wang G, Truong CK, et al. (2013) Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. Genet Med 15: 388-394.

34. McKernan KJ, Spangler J, Zhang L, Tadigotla V, McLaughlin S, et al. (2014) Expanded genetic codes in next generation sequencing enable decontamination and mitochondrial enrichment. PLoS One 9: e96492.

35. Rackham O, Filipovska A (2014) Analysis of the human mitochondrial transcriptome using directional deep sequencing and parallel analysis of RNA ends. Methods Mol Biol 1125: 263-275.

36. Imai K, Koshiyama A, Nakata K (2011) Towards clinical proteomics analysis. Biomed Chromatogr 25: 59-64.

37. Matthiesen R, Carvalho AS (2013) Methods and algorithms for quantitative proteomics by mass spectrometry. Methods Mol Biol 1007: 183-217.

38. Colwill K, Renewable Protein Binder Working Group, Graslund S (2011) A roadmap to generate renewable protein binders to the human proteome. Nat Methods 8: 551-558.

39. Fila J, Honys D (2012) Enrichment techniques employed in phosphoproteomics. Amino Acids 43: 1025-1047.

40. Adamski J, Suhre K (2013) Metabolomics platforms for genome wide association studies--linking the genome to the metabolome. Curr Opin Biotechnol 24: 39-47.

41. Ng EW, Wong MY, Poon TC (2014) Advances in MALDI mass spectrometry in clinical diagnostic applications. Top Curr Chem 336: 139-175.

42. Zhang A, Sun H, Xu H, Qiu S, Wang X (2013) Cell metabolomics. OMICS 17: 495-501.

43. Barnes MG, Aronow BJ, Luyrink LK, Moroldo MB, Pavlidis P, et al. (2004) Gene expression in juvenile arthritis and spondyloarthropathy: pro-angiogenic ELR+ chemokine genes relate to course of arthritis. Rheumatology (Oxford) 43: 973-979.

44. Baelde HJ, Eikmans M, Doran PP, Lappin DW, de Heer E, et al. (2004) Gene expression profiling in glomeruli from human kidneys with diabetic nephropathy. Am J Kidney Dis 43: 636-650.

45. Crow MK, Wohlgemuth J (2003) Microarray analysis of gene expression in lupus. Arthritis Res Ther 5: 279-287.

46. Zohlnhöfer D, Klein CA, Richter T, Brandl R, Murr A, et al. (2001) Gene expression profiling of human stent-induced neointima by cDNA array analysis of microscopic specimens retrieved by helix cutter atherectomy: Detection of FK506-binding protein 12 upregulation. Circulation 103: 1396-402.

47. Matsuzaki S (2011) DNA microarray analysis in endometriosis for development of more effective targeted therapies. Front Biosci (Elite Ed) 3: 1139-1153.

48. Poot M, van der Smagt JJ, Brilstra EH, Bourgeron T (2011) Disentangling the myriad genomics of complex disorders, specifically focusing on autism, epilepsy, and schizophrenia. Cytogenet Genome Res 135: 228-240.

49. Wang F, Wang H, Tuan HF, Nguyen DH, Sun V, et al. (2014) Next generation sequencing-based molecular diagnosis of retinitis pigmentosa: identification of a novel genotype-phenotype correlation and clinical refinements. Hum Genet 133: 331-345.

50. Farhan SM, Hegele RA (2014) Sequencing: the next generation--what is the role of whole-exome sequencing in the diagnosis of familial cardiovascular diseases? Can J Cardiol 30: 152-154.

51. Beck J, Pittman A, Adamson G, Campbell T, Kenny J, et al. (2014) Validation of next-generation sequencing technologies in genetic diagnosis of dementia. Neurobiol Aging 35: 261-265.

52. Lohmueller KE, SparsÃ̧ T, Li Q, Andersson E, Korneliussen T, et al. (2013) Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 93: 1072-1086.

53. Parsa A, Kao WH, Xie D, Astor BC, Li M, et al. (2013) APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med 369: 2183-2196.

54. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, et al. (2013) MUC5B promoter polymorphism and interstitial lung abnormalities. N Engl J Med 368: 2192-2200.

55. Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith AV, et al. (2013) Genetic associations with valvular calcification and aortic stenosis. N Engl J Med 368: 503-512.

56. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, et al. (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368: 107-116.

57. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506: 376-381.

58. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, et al. (2014) Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 46: 294-298.

59. Fernández-Tajes J, Soto-Hermida A, Vázquez-Mosquera ME, Cortés-Pereira E, Mosquera A, et al. (2014) Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. Ann Rheum Dis 73: 668-667.

60. Sanchez-Mut JV, Aso E, Panayotis N, Lott I, Dierssen M, et al. (2013) DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. Brain 136: 3018-3027.

61. Haas J, Frese KS, Park YJ, Keller A, Vogel B, et al. (2013) Alterations in cardiac DNA methylation in human dilated cardiomyopathy. EMBO Mol Med 5: 413-429.

62. Stankov K, Benc D, Draskovic D (2013) Genetic and epigenetic factors in etiology of diabetes mellitus type 1. Pediatrics 132: 1112-1122.

63. Blake R, Trounce IA (2014) Mitochondrial dysfunction and complications associated with diabetes. Biochim Biophys Acta 1840: 1404-1412.

64. Grady JP, Campbell G, Ratnaike T, Blakely EL, Falkous G, et al. (2014) Disease progression in patients with single, large-scale mitochondrial DNA deletions. Brain 137: 323-334.

65. Lo Turco EG, Cordeiro FB, Lopes PH, Gozzo FC, Pilau EJ, et al. (2013) Proteomic analysis of follicular fluid from women with and without endometriosis: new therapeutic targets and biomarkers. Mol Reprod Dev 80: 441-450.

66. Mendonca DM1, Pizzati L, Mostacada K, de S Martins SC, Higashi R, et al. (2012) Neuroproteomics: an insight into ALS. Neurol Res 34: 937-943.

67. Vojtova L, Zima T, Tesar V, Michalova J, Prikryl P, et al. (2010) Study of urinary proteomes in Anderson-Fabry disease. Ren Fail 32: 1202-1209.

68. van Bon L, Affandi AJ, Broen J, Christmann RB, Marijnissen RJ, et al. (2014) Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. N Engl J Med 370: 433-443.

69. Pastor MD, Nogal A, Molina-Pinelo S, Melendez R, Salinas A, et al. (2013) Identification of proteomic signatures associated with lung cancer and COPD. J Proteomics 89: 227-237.

70. Dumas ME, Kinross J, Nicholson JK (2014) Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. Gastroenterology 146: 46-62.

71. Vingara LK, Yu HJ, Wagshul ME, Serafin D, Christodoulou C, et al. (2013) Metabolomic approach to human brain spectroscopy identifies associations between clinical features and the frontal lobe metabolome in multiple sclerosis. Neuroimage 82: 586-594.

72. Cesario A, Marcus F (2011) Cancer Systems Biology, Bioinformatics and Medicine. Research and clinical applications; Springer Science: Heidelberg 3-479.

73. Martinez T, Wright N, Lopez-Fraga M, Jimenez AI, Paneda C (2013) Silencing human genetic diseases with oligonucleotide-based therapies. Hum Genet 132: 481-493.

74. Gonzalez de Castro D, Clarke PA, Al-Lazikani B, Workman P (2013) Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. Clin Pharmacol Ther 93: 252-259.

75. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646-674.

76. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100: 57-70.

77. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. Cell 153: 1194-1217.

78. Database of SNPs.

79. Database of genomic Variant. DGV.

80. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, et al. (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res 42: D297-303.

81. European Bioinformatics Institute-Database of Genomic Variants EBI Archive. 7DGVa.

82. Sanger Center Database. DECIPHER.

83. The Online Mendelian Inheritance in Man. OMIM.

84. Thorisson GA, Lancaster O, Free RC, Hastings RK, Sarmah P, et al. (2009) HGVbaseG2P: a central genetic association database. Nucleic Acids Res 37: D797-802.

85. The Human gene Mutation Database. HGMD.

86. The Catalogue of Somatic Mutations in Cancer. 8COSMIC.

87. The European genome-Phenome Archive. EGA.

88. 1000 Genomes.

89. Oncomine.

90. Gene Expression Atlas.