Systems biology: opening new avenues in clinical research

Franck Molina[1*], Matthias Dehmer[2], Paul Perco[3], Armin Graber[2]
Mark Girolami[4], Goce Spasovski[5*], Joost P. Schanstra[6*], Antonia Vlahou[7*]

1. SysDiag, CNRS / BIO-RAD UMR3145, Montpellier, France
2. Institute for Bioinformatics and Translational Research, UMIT, Tyrol, Austria
3. Emergentec biodevelopment GmbH, and Medical University of Vienna, Department of Internal Medicine III, Vienna, Austria
4. Department of Computing Science, University of Glasgow, Glasgow, UK
5. Department of Nephrology, Medical Faculty, University of Skopje, Skopje, R. Macedonia
6. Institut National de la Santé et de la Recherche Médicale (INSERM), U858, Toulouse, France and Université Toulouse III Paul-Sabatier, Institut de Médecine Moléculaire de Rangueil, Equipe n°5, IFR31, Toulouse, France
7. Laboratory of Biotechnology, Biomedical Research Foundation, Academy of Athens, Greece

[*] members of the European Kidney and Urine Proteomics Consortium (www.eurokup.org)

Address all correspondence to:
Antonia Vlahou Ph.D
Laboratory of Biotechnology
Biomedical Research Foundation, Academy of Athens
Soranou Efessiou 4
11527 Athens Greece
tel:30 210 65 97 506
fax: 30 210 65 97 545

Keywords:
Systems biology, functional genomics, biomedical informatics, biological networks, personalized medicine

Systems Biology as a paradigm shift in clinical research

The marked increase in research spending and individual efforts over the years on the development of disease biomarkers, therapeutic targets and new drugs has not been translated into the expected clinical outcome. This may point towards the need for a paradigm shift in the way biomarker and drug discovery is conceived: Identification of an association of individual molecules with a specific phenotype (*e. g.* presence, recurrence, progression, response etc of the disease), even following optimization of all technical parameters involved and initial confirmation of findings, is not sufficient for the establishment of this molecule as a clinically useful test or a potential drug target. The multi-factorial molecular phenotype of disease makes increasingly evident that development of novel therapeutic and disease detection approaches should be based upon the study of the entire "System" simultaneously (1). This is in contrast to the reductionist approach that focuses on individual molecules, as extensively used in the past two centuries to address those complex questions (2). The reductionistic approach is based on the principle that complex problems can be solved by reducing them into smaller, simpler units easier to deal with. However, it is obvious that living system behaviour cannot be predicted only by the sum of observations made on its individual parts (3). This is due to several reasons like inter-dependencies, context sensitivity, dynamics and more. Simply stated, molecules in a living cell are involved in networks of interactions that regulate the cell's basic functions such as proliferation, growth, differentiation and death. Thereby cells can be broken down into smaller biologically relevant entities such as DNA, proteins, amino acids, etc and on the other end cells are also part of tissues and organs, all connected and dependent. Disruption of a partner in these interactions does not result in linear and definable effects, but rather in global and often unpredicted perturbations of the whole network (1,3).

Along this line of thinking, identification of efficient drug targets and development of optimal therapeutic modalities has to rely on a better understanding of biology in a systemic manner, on "Systems Biology", predicting the series of events and resulting response of the biological network triggered by the introduction of the new interfering compound (4). Similarly, individual molecules/ biomarkers are destined to fail in providing optimal sensitivity and specificity in disease detection since they cannot reflect with accuracy the disease complex molecular phenotype. Systems Biology-based diagnostic and prognostic models consisting of relevant panels of molecules - key branches of the cellular network, appear to more accurately reflect pathophysiology, consequently may have a much higher chance of success in the clinical setting (1,5)

Systems Biology requirements

This shift of interest from the identification of individual components to the ways that these components interact (3) as addressed by Systems Biology has raised specific research needs and will change significantly the ways biological and clinical studies are conducted. Of the most pronounced effects is the crossing between the borders and need for integration of multiple disciplines (biology, computing sciences, mathematics, medicine, etc.).

To get a holistic view of a system's biology, multiple and different types of observation have to be combined: clinical (e.g. pathological, demographical, epidemiological etc data), imaging (including molecular, tissue and organ imaging data), as well as molecular (including large scale genotypic, gene expression (mRNA, miRNA), proteomics, metabolomics, lipidomics etc.) data (6-8). These can be applied in a combined way to address questions at different levels and scales of a biological system, ranging from the definition of the transcriptional regulation network of a specific gene, to the cellular response to specific stimuli in time and space including cell interactions with the tissue environment. Therefore, the output of a Systems Biology approach can provide understanding on cell's (i) structures (such as gene regulatory and biochemical networks, physical structures), (ii) dynamics (both quantitative and qualitative), (iii) control and, (iv) design (3).

A Systems Biology approach therefore implies the integration of multi-source data along with an extensive use of mathematical and computational concepts and, additionally, experimental techniques to handle its complexity. It is safe to say that the progress in Systems Biology applications reflects advancements in technologies for data acquisition, development and adaptation of mathematical tools for their integrated analysis in parallel to an increase in computing power. A brief description of the input, analysis tools, and expected output associated with a Systems Biology approach (depicted in **Figure 1**) is given below:

-Study design, standardization, quality control, data acquisition and mining:

A key element in Systems Biology approaches is the availability of reliable data that are associated with a high level of confidence (in other words: the underlying assumption that the data are correct must not be violated) (9-11). This requirement demands the employment of appropriate practices and technologies using standard operating procedures, quality controls, and a proper experimental design from both clinical and statistical points of view. The data are subsequently pre-processed and mined for the identification of traits (e.g. changes of molecules or alterations in tissue structure) with a true association with the disease. For the latter, application of proper uni- and multivariate statistical methods that take into consideration the high data dimensionality and variability appears essential (12). Definition of incorrect associations will result in errors in the subsequent modelling approaches, due to false input information.

-Functional Ontologies and databases

Biological ontologies and classification systems, which contain definitions of terms in a specific domain and the relations among them, are important for efficient integration of experimental and public available datasets and information (13). Ontologies can thereby be viewed as a means to interpret and categorize the data (Figure 1). A widely used classification system is the gene-ontology which links major categories including ‘biological process’, ‘molecular function’ and ‘cellular component’ to molecules. In addition, a variety of databases on molecular pathways and cell signalling cascades (e.g. KEGG, Reactome, Wikipathways, PANTHER etc) can be leveraged. The knowledge embedded in these databases and ontologies is continuously growing due to constant input from the research community (7, 14).

-Biological Network analysis

Following data annotation by linking to available ontologies and databases, mathematical models are developed to describe molecular associations and interactions and/or map them to known pathways (reviewed in 14; 15, 16). These models can be usually represented by networks. Here, nodes typically represent molecular entities (genes, proteins, metabolites etc.) and edges (links) between them represent relationships between pairs of nodes (8, 15-16). Networks can represent knowledge at different levels of abstraction such as gene regulation, protein or metabolite interactions, and others. The vast majority of models generated so far reflect either protein-protein interaction (PPI) networks or co-expression networks as predicted based on one type of data -primarily gene expression- collected from specific biological specimens and pathophysiology (reviewed in 8). Recently, efforts have been made to build biological networks not only on PPI and co-expression data but on the incorporation of different heterogeneous data sources (17-20; e.g. transcription factor binding profiles, functional annotation, information on subcellular location, genomic and proteomic profiles). These generated networks of higher complexity can for example be used in order to analyse lists of differentially regulated features from different -omics experiments and/or identify novel clusters of biomarker candidates for disease diagnosis.

- Model generation and validation

Systems Biology studies rely on the iterative cycle of "model building and validation, perturbation of model, hypothesis generation, experimental verification, model refinement, hypothesis refinement", as outlined in more detail in the accompanying editorial by Dominiczak *et al*. During model development a variety of mathematical techniques have to be employed ranging from simple Pearson correlations to Boolean or Bayesian networks (21,22). Various software tools have been developed to facilitate creation, analysis and *in silico* validation of models (23). As expected from the status quo: datasets that may in part be biased or not entirely accurate, when analyzed using high dimensional mathematical algorithms (which by themselves lead to several possible solutions of a given problem), will generate models highly unlikely to be correct. Therefore, a tight cross-talk of the *in silico* with the experimental data in the form of a "validation loop" is of paramount importance: The generated models must be evaluated based on experiments in the biology laboratory, and the generated experimental data are in turn being used to refine and further improve on the *in silico* models.

- Platforms for data deposition and dissemination

Platforms that can accommodate and support the analysis and presentation of data from multiple sources and dissemination of the respective results are an evident requirement for Systems Biology and are continuously evolving. Multiple software packages and web-based tools are now available that can perform network mapping and visualization and in several cases network manipulation and inference (reviewed in 7, 24). Nevertheless there is much room for improvement in all cases in particular targeting more comprehensive data integration from multiple sources, providing methods to address information consistency and ensuring compatibility across datasets and disciplines.

Challenges-Outlook

Systems Biology is a new and exciting discipline which evidently involves several challenges:

- processes that ensure high quality and repeatability/reproducibility of the –omics data have to be implemented and adhered to;
- multi-source data integration in a comprehensive way remains elusive, a problem that is at least in part linked to data incompatibility issues;
- the development of techniques to analyze complex pathway structures and/or dynamics is still challenging; classical graph-theoretic methods can often not be used. In addition, since such biological networks are often inferred probabilistically, erroneous graphs may be generated (15) pointing towards a need to adapt statistical methods for overcoming this problem.
- multiple tools for data analysis are available, nevertheless their application is still prohibited due to the need for increased computational skills by the user. This further underscores the need for either adaptation of these tools to become more friendly for cross-discipline users or for development of educational programs emphasizing on the acquisition of computer skills.

Once those challenges are met, Systems Biology, defined as the holistic investigation of a biological system and determination of the complex network of interactions among its components, clearly offers exciting opportunities in clinical research: Systems Biology approaches hold the promise of substantially improving the current state-of-the-art in medicine by providing the ability to distinguish between multiple disease states and enabling the identification of the disease underlying molecular causes. This is particularly important for multi-factorial diseases such as Chronic Kidney and Cardiovascular diseases (see Dominiczak et al, accompanying editorial; also 25). Such disease definition, relying not only on the clinical symptoms but also on the molecular components involved and the dynamics of their interactions, is a prerequisite for the development of personalized medicine approaches.

Conflict of Interest Statement: There is no conflict of interest or involvement of any of the authors that might raise the question of bias in the work reported or in the conclusions, implications or opinions stated. The manuscript has been submitted solely to this journal and is not published, in press, or submitted elsewhere.

Reference List:

1. Hurst RE. Does the biomarker search paradigm need re-booting? *BMC Urol*. 2009; 9:1

2. Ahn AC, Tewari M, Poon C-S, Phillips RS. The Limits of Reductionism in Medicine:Could Systems Biology Offer an Alternative? *PLoS Med* 2006; 3(6): e208

3. Kitano H. Systems biology: a brief overview. *Science* 2002; 295 (5560): 1662–1664

4. Hopkins AL Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008; 4(11):682-690

5. Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; 1(1):2

6. Kohl P, Noble D. Systems biology and the virtual physiological human. *Molecular Systems Biology* 2009; 5:292

7. Wheelock CE, Wheelock AM, Kawashima S *et al* Systems biology approaches and pathway tools for investigating cardiovascular disease. *Mol. BioSyst.* 2009; 5: 588–602

8. Schadt EE, Zhang B, Zhu J. Advances in systems biology are enhancing our understanding of disease and moving us closer to novel disease treatments. *Genetica*, 2009; 136: 259-269

9. Mischak H. Apweiler R., Banks RE *et al.* Clinical Proteomics: a need to define the field and to begin to set adequate standards. *Proteomics Clin. Appl.* 2007; 1:148-156

10. Brazma,A. Hingamp P, Quackenbush J *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 2001; 29, 365-371.

11. Dudley JT, Tibshirani R, Deshpande T, Butte JA Disease signatures are robust across tissues and experiments. *Molecular Systems Biology* 2009; 5:307

12. Perco P, Rapberger R, Siehs C, *et al* Transforming omics data into context: bioinformatics on genomics and proteomics raw data. *Electrophoresis* 2006; 27(13):2659-75

13. Meier S, Gehring C. A guide to the integrated application of on-line data mining tools for the inference of gene functions at the systems level. *Biotechnol J*. 2008; 3(11):1375-87.

14. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. .*Molecular systems biology* 2009; 5:290

15. Emmert-Streib F**,** Dehmer M. Analysis of Microarray Data: A Network-based Approach. Wiley-VCH, Weinheim, Germany, 2008

16. Viswanathan GA, Seto J, Patil S *et al* Getting Started in Biological Pathway Construction and Analysis. *PLoS Comput Biol* 2008; 4(2): e16

17. Bernthaler A, Mühlberger I, Fechete R *et al* A dependency graph approach for the analysis of differential gene expression profiles. *Mol Biosyst* 2009; 5(12):1720-31

18. Perco P, Wilflingseder J, Bernthaler A *et al* Biomarker candidates for cardiovascular disease and bone metabolism disorders in chronic kidney disease: a systems biology perspective. *J. Cell. Mol. Med.* 2008; 12(4):1177-1187

19. Ishii N, Nakahigashi K, Baba T *et al*. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* 2007; 316: 593-597

20. Zhu J, Zhang B, Smith EN *et al* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genetics* 2008; 40:854-861

21. Markowetz F, Spang R Inferring cellular networks-a review. *BMC Bioinfor*matics2007; 8 (6)S5

22. Schlitt T, Brazma A. Current approaches to gene regulatory network modelling. *BMC Bioinformatics* 2007; 8 (6) S9

23. Endler L, Rodriguez N, Juty N, *et al* Designing and encoding models for synthetic biology. *J R Soc Interface* 2009; 6 (S4):S405-17.

24. Ng A, Bursteinas B, Gao Q *et al*. Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Briefings Bioinf.* 2006; 7: 318-330

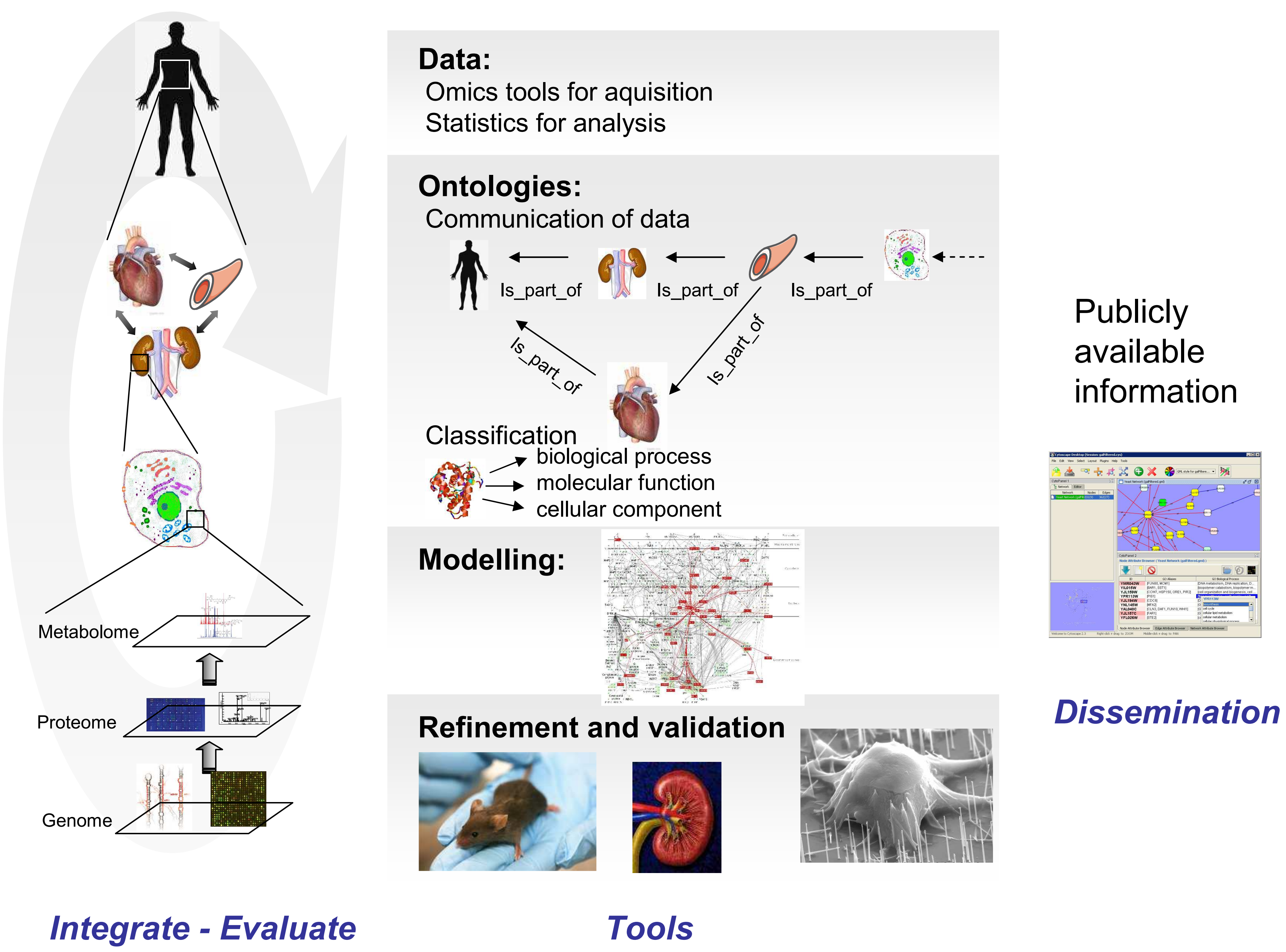
25. Solez K, Colvin RB, Racusen LC, *et al*. Banff '05 Meeting Report: differential diagnosis of chronic allograft injury and elimination of chronic allograft nephropathy ('CAN'). *Am J Transplant*. 2007; 7:518-26

Figure Legends

Figure 1

Representation of the major elements and requirements of Systems Biology.

The definition of an individual based on the correlation of the molecular elements may be seen as the ultimate goal of Systems Biology, as presented on the left. To reach this goal, data on molecules, physiology, and biology have to be collected and linked to pathology. Ontologies are employed to display communication between data, and correctly interpret datasets. This combined know-how is used in mathematical modelling, which generates networks of molecules, but ultimately also smaller (cells) and larger (tissue, organs, organisms) biological structures. This process is constantly refined by the iterative alignment of the *in silico* models with actual biological data.


Metabolome
Proteome
Genome
Data:
Omics tools for aquisition
Statistics for analysis
Ontologies:
Communication of data
Is_part_of
Is_part_of
Is_part_of
Is_part_of
Is_part_of
Classification
biological process
molecular function
cellular component
Modelling:
Refinement and validation
Publicly available information
Dissemination
Integrate - Evaluate
Tools