Translating systems biology into medical applications:

Report of the 3rd Bertinoro Systems Biology Workshop.

Clermont, G[1], Auffray, C.[2], Moreau, Y.[3], Rocke, D[4], Delavi, D[5], Dubhashi, D.[5], Marshall, D.[6], Raasch, P.[7], Dehne, F.[8], Provero, P.[9],Tegner, J.[10], Langston, MA.[11], Benson, M.[12]

[1]Department of Critical Care Medicine and CRISMA laboratory, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261 USA, cler@pitt.edu

[2]Functional Genomics and Systems Biology for Health - CNRS Institute of Biological Sciences - 7, rue Guy Moquet, BP8 94801 Villejuif cedex – France; auffray@vjf.cnrs.fr

[3] K.U.Leuven, ESAT/SCD, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee Belgium, Yves.Moreau@esat.kuleuven.be

4Department of Public Health Sciences, University of California, Davis, One Shields Ave, Davis, CA 95616 USA, dmrocke@ucdavis.edu

Pablo Provero

[5]Dept. of Computer Science and Engineering. Chalmers and Gothenburg University, SE 41296 Sweden, dalevi@chalmers.se, dubhashi@chalmers.se

[6]Meharry Medical College, 1005 Dr. D.B. Todd Boulevard, Nashville, TN 37208 USA, dmarshall@mmc.edu

[7]Systems Biology and Bioinformatics Group, University of Rostock, Rostock, Germany, peter.raasch@web.de

[8]School of Computer Science, Carleton University, 1125 Colonel By Drive Ottawa, Ontario K1S 5B6, Canada, frank@dehne.net

[9]Computational Biology Unit Molecular Biotechnology Center University of Torino, Via Nizza 52, I - 10126 TORINO, Italy, paolo.provero -AT- unito.it

[10]Institutionen för Medicin, Karolinska Universitetssjukhuset, Solna, 171 76 Stockholm, jesper.tegner@ki.se

[11]Department of Electrical Engineering and Computer Science, College of Engineering, University of Tennessee, 1122 Volunteer Boulevard Knoxville, TN 37996 USA, langston@eecs.utk.edu

[12]The Unit for Clinical Systems Biology, The Queen Silvia Children's Hospital, Gothenburg, Sweden, mikael.benson@vgregion.se

Address for Correspondence:

Gilles Clermont, M.D.
Scaife 602
3550 Terrace St
University of Pittsburgh
Pittsburgh PA, 16261
email: cler@pitt.edu
tel: 1-412-647-7980

The first week of April 2009, clinical scientists and physicians were to be found walking the cobbled streets of Bertinoro, Italy, with fellow mathematicians, computational scientists and statisticians (http://www.cs.utk.edu/~langston/BSB2009/). The beautiful countryside offered the opportunity for scientists from disparate backgrounds to reflect on the questions: “what are key obstacles in the translation of progress in Systems Biology to the clinical world?” and “how would one define the discipline of Systems Medicine? Technical terms such as gene, rational function, microarray, inverse problem, acute sepsis, algorithm, SNP screen, lattice, macrophage, graph theory, atherosclerosis, compute kernel as well as correlation were meshed and echoed in a collaborative spirit inside the picturesque lecture hall of the Alma Mater Studorium University of Bologna conference center at the ancient monastery and castle located on the top of a wind-swept hill in Bertinoro.

Defining Systems Medicine

The meeting, held over four days, comprised plenary lectures followed by extensive thematic discussions, formal and informal, centered around the theme of Systems Medicine as a distinct translational discipline.[1] We proposed that Systems Medicine is the application of Systems Biology to the understanding and modulation of developmental disorders and pathological processes in human health. While there is no clear boundary between Systems Biology and Systems Medicine, it could be stated that Systems Biology aims at a fundamental understanding of biological processes and ultimately at an exhaustive modeling of biological networks, whereas Systems Medicine emphasizes that the essential purpose and relevance of models is translational, aimed at diagnostic, predictive and therapeutic applications. Accordingly, advances in Systems Medicine need to be assessed on a medical and clinical scale as the correspondence between biology and medicine is intricate: some seemingly straightforward biological

models may have an important medical impact, while some impressively complex molecular models may not be immediately medically relevant. While Systems Biology may have so far focused primarily on the molecular scale, Systems Medicine must directly incorporate mesoscale clinical information into its models (in particular classical clinical variables, biomarkers and medical imaging data). As an example, it has become increasingly clear that prognostic and predictive models for malignant tumors using expression data cannot ignore information from classical prognostic indices. Furthermore, because of the necessary multi-scale nature of the models bridging embedded levels of organization from molecules, organelles, cells, tissues, organs, all the way to individuals, environmental factors, populations, and ecosystems, Systems Medicine aims at discovering and selecting the key factors at each level and integrating them into models that reveal the global, emergent behavior of the biological processes under consideration. Such an approach is expected to be most valuable when the execution of all the experiments necessary to establish exhaustive models is limited by time and expense (e.g., in animal models) or basic ethics (e.g., human experimentation). Systems Medicine as a discipline did not emerge from clinical medicine. Rather, advances in Systems Biology created the necessary conditions, and tools for the emergence of Systems Medicine. Accordingly, it is currently appropriate to position Systems Medicine as an extension of Systems Biology.

Scale-specific modeling vs. multi-scale modeling (Charles, Gilles, Peter)

Participants noted that computational models have for the most part attempted to assimilate massive data streams collected using global measurement technologies (techniques that look at the complete set of genes, transcripts, proteins, metabolites or other features in an organism) using high throughput techniques and have been, by-and-large, scale-specific. Such attempts target the development of predictive mathematical

and computational models of functional and regulatory biological networks. Specific biological hypotheses can thus be tested by designing a series of relevant perturbation experiments.2 There is clear merit in such an incremental approach. Yet, its true potential is likely to be realized only when such data-driven, bottom-up approaches are combined with top-down, model-driven approaches to generate new medically relevant knowledge. An open question is whether integrative systems biology approaches can reveal underlying principles related to the aforementioned biological functions. It is probably improper to speak of the existence of biological laws in the sense of physical laws, yet there probably are deeper dynamical principles guiding the evolution of biological systems. Energetics and physical constrains play an important role in all scale-specific models. Additional principles at play across multiple scales in biological systems are far less apparent. It thus appears prudent at this stage that top-down and multi-scale models seek to recapitulate scale-specific observables. As mentioned before, if computational models are to be validated by experiments (such as randomized clinical trials) and predictive of therapeutic interventions, relevant system observables must be included.

Ontologies relevant to Systems Medicine

Considerable attention was brought to the importance of developing ontologies relevant to Systems Medicine. Such ontologies must reflect knowledge based on biological function, rather than biological structure. Indeed, structure is permissive to function, and there is clearly a wide variety of structures that could have evolved, under genetic, molecular or physical constraints to accomplish a given function. Examples include energy generation and storage, and transmission of information. The recent emphasis at mapping structure unto function is vital to the advancement of Systems Medicine. It does appear also that the development of appropriate ontologies could

serve promote a (re)interpretation of empirical evidence in light of such ontologies. Recent efforts at data reduction of longitudinal expression data, using principal component analysis to identify and follow health and disease “trajectories”, represent an attempt at understanding such “eigenprocesses” from a data-driven perspective.[3, 4] Typically and unfortunately, such processes have limited intuitive meaning. Alternatively, existing community (e.g. http://www.geneontology.org/) or commercial efforts aimed at developing a phenotype-driven ontology (e.g. annotating genes to a priori defined functions such as “cell-cycle” or “inflammatory response”) are commendable and clearly of great value, although it is apparent that extensive cross-contamination exists between such functional assignations and the response to even the simplest experimental perturbation of functions Knowledge representations relevant to Systems Medicine will probably lie within this spectrum and computational efforts will likely be crucial to their development.

The battle of certainty and false discovery

Panel participants reaffirmed that experimental design and statistical analysis play important roles in discovery and validation in systems biology and medicine.[5] As RA Fisher said in the 1938 Indian Statistical Congress, “To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of.” Study design is often the weakest point in many complex molecular studies in systems biology and medicine. For example, patients with a disease such as ovarian cancer may be compared to normal controls to discern aberrant regulation of pathways. But if the controls are not carefully selected to be comparable to the patients demographically and in other ways (age, sex, income, social class), then differences observed may be due to other factors than the disease. Researchers are often unduly optimistic about sample sizes required

to show differences, and fail to consider many confounding effects. Inter-individual variability in humans can be large, often the largest effect in a study. This provides an avenue for exploration of individual effects, leading to personalized medicine, but also can make detection of differences across subjects quite difficult.

High throughput technologies have brought severe challenges to experimental design and interpretation of results. Gene expression microarrays can have hundreds of thousands of probes, with tens of thousands of targets. In order to show that an effect is real, p-values need to be adjusted for the number of tests conducted, for example by the Benjamini and Hochberg false discovery rate[6] (1995), which can raise the bar for significance so high that it is unlikely to be met. Short of greatly increased sample sizes, this problem mostly can be met by using a priori biological knowledge either to trim the list of analytes to a relatively small number for which the multiple testing correction is modest, or by testing pathways or groups of genes.[7] This is best done not by testing every group of genes defined by a GO term or a KEGG pathway, but by selectively testing those thought to be of importance. It is also important to account for the effect of using thousands of measurements in developing signatures for diagnosis, prognosis, drug effectiveness, etc. In particular, evaluating the results on the same data set that was used to develop the method will lead to important optimistic biases, and to studies that cannot be replicated.[8]

As previously stated, addressing the above mentioned challenges may have direct clinical implications. One of the most significant problems encountered by clinicians on a daily basis is that patients that appear to have the same disease may not respond to the same treatment. Some patients even experience severe side effects. Variable treatment response is also one of the most important causes for the huge costs involved in drug development. Taken together, this causes both increased suffering and costs. Ideally, physicians should be able to routinely measure a few diagnostic proteins

in for example saliva or blood to personalize medication for each patient. At present there is not enough knowledge about the causes for variable treatment responses in most common diseases. However, very recent studies of genetic markers for response to treatment with anticoagulants indicate that personalized medication may become clinical reality within the next 5 to 10 years.[9] The main problems involved in finding markers for personalized medication are that each complex disease may involve altered interactions between hundreds or even thousands of genes, in combinations that can differ between patients. This heterogeneity may, in turn, depend on both genetic and environmental factors. On top of this complexity, there are significant problems involved in clinical research. Ideally, a study aiming to find markers for personalized medication would involve a known external cause, a key cell type and a read-out, all of which can be studied experimentally in samples from patients. For most complex diseases, all of these factors are not readily available. It is therefore important to find model diseases, in which all those factors can be studied together in samples from patients using high-throughput technologies and systems biological principles.[10] Such model diseases might be used to develop and apply the methods required to find markers for personalized medication. It is also been suggested that the same methods might be applied to find markers to predict risk of developing diseases.[11] If successful, this may lead to a new era of preventative medicine. Finally, the methods may be of great value for drug development; if it were possible to predict which patients respond to medication, this would result in increased efficacy and reduced risk of not being able to market drugs that have been developed at great cost. Conversely, delineation of patients that do not respond to a medication may help to develop new drugs for that specific subgroup.

Network Analysis

A network represents a set of objects and their mutual relationships. Much

biological and medical knowledge can be naturally represented as networks: protein interaction networks, metabolic networks, gene coexpression networks, disease networks and many more. Participants raised a number of concerns regarding current trends in network analysis in Systems Biology, and potential extension to the clinical arena through the construction of “diseasomes”.[12] Do network representations actually convey new knowledge or are they just a convenient and eye-catching way to represent data. Can such networks be used to extract new information that is relevant to biological understanding and clinical practice? From a mathematical and computational perspective, what are the topological properties that let you identify the nodes in a network, those that cannot be removed without deeply affecting the whole structure? These are open research problems that were extensively discussed at the meeting.

While the methods used to analyze networks might still be primitive, they are already providing useful information, especially on the genetics of disease. Several research efforts discussed at the meeting suggested that it was possible to integrate information from various biological networks to identify genes involved in both Mendelian and complex diseases. Thought must be given to how network inference for microarray and other types of data are evaluated. These tools should be developed ideally in a dialogue with clinicians. Clinicians should easily and reliably be able to assess which tool is best suited to their application. There is need for systematic benchmark testing and comparative evaluation of the major tools available. For example, current methods are typically tested only on simulated data or tested for functional enrichment of GO categories which may not be very relevant to the clinician. Another possibility is the identification of specific interactions that have been extensively validated, a so called *gold standard*. Some of these issues have been raised by the DREAM (Dialogue on Reverse Engineering Assessment and Methods) initiative (http://www.nyas.org/ebriefreps/main.asp?intEBriefID=534).

Participants presented their own efforts to address some of these issues. For example, representing gene interactions with hypergraphs may be a useful method to discover parts of a network that is not fully resolved. The biological plausibility of such representations could then be discussed with basic biologists and clinicians. Approaching network representations from a different perspective, participants noted ongoing efforts to extend network analysis as a tool to develop evaluations of disease-specific ontologies, such as those developed by Bruce Aronow and his group at the Cincinnati Children's hospital.[13]

Conclusions

The meeting was highlighted with several cross-disciplinary discussions. The broad consensus was that improvements in academic infrastructure are sorely needed in order to facilitate cross-disciplinary translational studies that can one day connect what can be learned using model organisms with real-time samples from human patients. Yet, there was a growing consensus that a serious and useful dialogue between the clinic and Systems Biology has begun. Future meetings in this historical location will hopefully provide continuing evidence that the systems biology community has taken this development to its heart, building on a millennium of scholarship and medical tradition.

Acknowledgments

We wish to thank the meeting organizers, Mike Langston, Devdatt Dubhashi and Mikael Benson for their, and the Bertinoro University Center, University of Bologna, for their generous support of the Bertinoro Systems Biology workshop.

References

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

2. Kitano H. Computational systems biology. Nature 2002; 420(6912):206-210.

3. Calvano SE, Xiao W, Richards DR et al. A network-based analysis of systemic inflammation in humans. Nature 2005; 437(7061):1032-1037.

4. McDunn JE, Husain KD, Polpitiya AD et al. Plasticity of the systemic inflammatory response to acute infection during critical illness: development of the riboleukogram. PLoS ONE 2008; 3(2):e1564.

5. Rocke DM. Design and analysis of experiments with high throughput biological assay data. Semin Cell Dev Biol 2004; 15(6):703-713.

6. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995; 57(1):289-300.

7. Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102(43):15545-15550.

8. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002; 99(10):6562-6566.

9. Takeuchi F, McGinnis R, Bourgeois S et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. PLoS Genet 2009; 5(3):e1000433.

10. Mobini R, Andersson BA, Erjefalt J et al. A module-based analytical strategy to identify novel disease-associated genes shows an inhibitory role for interleukin 7 Receptor in allergic inflammation. BMC Syst Biol 2009; 3:19.

11. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004; 306(5696):640-643.

12. Barabasi AL. Network medicine--from obesity to the "diseasome". N Engl J Med 2007; 357(4):404-407.

13. Aronow BJ. Dysfunctional genomics: toward an integrative biology of disease and health--application to IBDs. J Pediatr Gastroenterol Nutr 2008; 46 Suppl 1:E3-E4.