

Meeting report

Integrating systems biology with clinical research

Rainer Pepperkok* and Stefan Wiemann†

Addresses: *Cell Biology and Cell Biophysics Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

†Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.

Correspondence: Rainer Pepperkok. Email: Pepperko@embl.de

Published: 1 July 2008

Genome Biology 2008, **9**:314 (doi:10.1186/gb-2008-9-7-314)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/7/314>

© 2008 BioMed Central Ltd

A report on the conference 'Systems Genomics 2008', Heidelberg, Germany, 2-3 May 2008.

High-throughput techniques in genomics, proteomics, and cell biology hold the promise of systems-level analyses to elucidate fundamental biological principles and to understand and predict the behavior of cellular systems in health and disease. With this challenge in mind, the recent Systems Genomics 2008 conference [<http://www.dkfz.de/mga/SG2008>] in Heidelberg brought together researchers in the fields of genomics, functional genomics, and systems biology to discuss the latest technological developments and their possible implications for clinical research.

Systems and biological networks

Two keynote presentations dealt with the understanding of interaction networks, and with the implications of such knowledge for developing new therapeutic strategies in cancer. Peter Sorger (Harvard Medical School, Boston, USA) highlighted the central role of stochastic processes in the fate of cells and of the importance of determining them quantitatively. He and his colleagues have shown that, even in a visually homogeneous cell population, not all cells react the same way in response to a specific perturbation. While the causes of such an effect are complicated enough for binary decisions like apoptosis, differentiation or cell division, they will be even more so for continuous events such as signaling and cell migration. Such processes need to be quantitatively analyzed at single-cell resolution - by live-cell imaging or flow cytometry, for example. Neither genomics nor proteomics approaches can be carried out at this level of resolution yet.

Yossi Yarden (Weizmann Institute of Science, Rehovot, Israel) discussed the robustness of biological systems as one prerequisite for cell survival in an unfriendly environment.

Taking the ErbB family of growth factor receptors and breast cancer as an example, he showed that such systems have evolved to withstand perturbations such as those induced by common therapies. He suggested that compensatory pathways provide the plasticity needed to confer drug resistance, and that this would be responsible for the long-term failure of many therapies. However, despite the fitness of cells to deal with such common perturbations, he claimed that unusual types of perturbations would render the system fragile and should re-establish drug potency. For example, targeting of ErbB2 with two different antibodies should more efficiently attract natural killer cells to the tumor cells, and combination of chemotherapy with monoclonal antibodies would better remove this receptor from the cell surface and result in reduced signaling. Cancer cells should then not be able to compensate for such uncommon perturbations, resulting in a much enhanced therapeutic response. A detailed knowledge of the signaling pathways and networks involved would define the optimal treatment regime, depending on the status of the cancer cell in terms of key signaling factors.

High-throughput technologies

Much of our current knowledge about genes and their expression is based on the approximately 7 million expressed sequence tag (EST) sequences in the UniGene database. Sumio Sugano (University of Tokyo, Japan) introduced the next-generation DNA sequencers, such as those marketed by Illumina, ABI, and Roche, which promise to determine more than 10 million sequences from just one experiment. With this remarkable technology, genes, promoters and transcription start sites will in future be able to be mapped in single cell types with unprecedented precision. Sugano showed the preliminary results of such an unbiased approach, where an Illumina sequencer had been used to map transcription start sites and transcripts. He concluded that some start sites are indeed cell-type specific and that the

huge number of tags generated permits fine-grained analysis of gene expression. But in every million cDNAs captured and sequenced by these techniques, gene expression turns out to range from a few transcripts to thousands of transcripts from the same gene. Sugano pointed out that with the depth of data achievable with next-generation sequencing, sparse transcription cannot be distinguished from what could be termed 'transcriptional noise'. There are no clear cutoffs, which complicates the detection of rarely expressed genes, and especially of intergenic and antisense transcription.

Caroline Shamu (Harvard Medical School, Boston, USA) discussed the many challenges associated with currently fashionable genome-scale screening by RNA interference (RNAi) using small interfering RNAs (siRNAs). She reported on projects where high-throughput transfection methods such as reverse transfection are combined with a conventional plate-and-assay design and high-content read-out to conduct more than 20 large-scale primary screens in different human and mouse cell lines. In her talk she concentrated on technical issues of RNAi screening in her central facility, stressing the importance of spending enough effort to make the assay robust, and to work on plate designs in order to circumvent edge and plate effects as these hamper data analysis. Once these issues are addressed, RNAi seems to be rather robust, as she screened for phenotypes in cancer, infectious diseases, neurobiology, and stem-cell biology, utilizing a number of different cell lines in combination with diverse transfection reagents and siRNA concentrations. While initial RNAi screens had mostly been done with plate readers, data acquisition is increasingly shifting towards high-content screening microscopy. Dorit Arlt (German Cancer Research Center, Heidelberg, Germany) reported that RNAi is also ideal for identifying functional interaction networks of genes. She presented data where knock-down of a single network component did not have a phenotype itself, yet the parallel perturbation of two or more genes did, thus revealing their functional interactions with the network. First she established a literature network of cell-cycle regulation consisting of the ErbB receptor family, AKT1 and MEK1 signaling intermediates, estrogen receptor alpha and Myc transcription factors, and cyclins D1 and E1 as well as cyclin-dependent kinases Cdk1, 4 and 6 as effector molecules. The input was epidermal growth factor (EGF), and the phosphorylation state of the retinoblastoma (Rb) protein was measured in response to siRNA treatments. She systematically perturbed the network components alone and in combinations to identify critical components in the regulation of that network. Indeed, she found novel edges in that network, most of which indicated feedback regulations, for example from cyclin D1 to AKT1 and MEK1. There was a common feeling that such screens will unravel the molecular mechanisms of cellular processes and potentially define major targets for interventions to cure human diseases. But it also became clear that such experiments take months rather than days, which needs to be improved.

The complementation of functional gene-interaction experiments with information on physical protein-protein interactions is a logical next step in the generation of protein networks. Using tandem affinity purification (TAP), Anne-Claude Gavin (EMBL, Heidelberg, Germany) and her collaborators have found that at least 80% of the proteins in yeast exert their function in complexes with other proteins. She stressed the point that protein complexes are, in general, highly dynamic structures, and often the same proteins are components of several protein complexes. To fully understand the modularity of the proteome in all its dynamics and stoichiometry will thus be a true challenge for the coming years.

Two array-based platforms were discussed as tools for qualitative and quantitative proteomics. The nucleic acid programmable protein array (NAPPA) presented by Joshua LaBaer (Harvard Institute of Proteomics, Boston, USA) enables the *in situ* production of large numbers of different protein probes with a success rate of greater than 90%. For use in this system, comprehensive collections of expression plasmids harboring the protein-coding regions of genes are being established at Harvard (in the Flexgene project) and by an international project (the ORFeome Collaboration). LaBaer described how NAPPA arrays have been used to generate protein-protein interaction maps, to test for serum-responsive proteins in the *Pseudomonas* proteome, and to detect tumor-associated antigens as a way of monitoring responses to cancer therapy. On the quantitative side, protein microarrays consisting of spotted protein lysates or antibodies tagged with Odyssey IRDye 680 or IRDye 800 were introduced by Ulrike Korf (DKFZ, Heidelberg, Germany). Detection of signals in the near infrared led to low background, low variability between samples and a high dynamic range. The highly parallel setup of these arrays enabled the dynamics of activation of the kinase ERK after stimulation with erythropoietin to be quantified in cell lines, for example. A problem in applying this method on the genome scale is the availability of high-quality antibodies that must be highly specific for their respective targets.

Challenges for computational biology and data integration

Ways of automating the analysis of complex phenotypes such as cellular morphology will be needed to speed up the type of screens described above. Advanced methods of image and data analysis for evaluating cellular morphology were described by Wolfgang Huber (European Bioinformatics Institute (EBI), Cambridge, UK), who showed that supervised learning approaches are able to quantify the occurrence of a number of cellular morphological phenotypes in an unbiased manner. Known complex cellular phenotypes are first user defined in a limited number of 'teaching images'. Automated analysis algorithms then recognize these phenotypes and quantify their occurrence in automatically

acquired microscope images. Using this approach, Huber and his collaborators have established a cell-morphological phenoprint of the human genome by siRNA-based screens assayed by high-content microscopy.

In regard to data integration, Henning Hermjakob (EBI, Cambridge, UK), who has been involved in developing Human Protein Organization standards for reporting and data collection, stressed the necessity of common standards as prerequisites for efficient data exchange. Given the rapidly increasing number of huge and diverse datasets being generated in the 'multi-omics' sciences, the proper analysis and, even more, the integration of data depends on annotation with enough information to enable researchers to evaluate and understand how the data have been collected and for what purposes they can be sensibly exploited. Reporting guidelines and data-exchange formats from many research communities are in existence, for example, MIAME, MIAPE, and MIACA for microarray, proteomics and cellular assay data, respectively. Hermjakob noted that, unfortunately, many of these guidelines are not yet in general use by the scientific community. Harmonization of the different guidelines to enable the integration of multi-omics data, a prerequisite for systems biology, also remains a challenge.

With human genome sequencing now entering the era of '1,000 genomes' and our personal genomes coming within reach, Rolf Apweiler (EBI, Cambridge, UK) reported on the ongoing implementation of the Human Proteome Project as a natural next step. This project aims to catalog the parts list of all proteins, splice variants, and modifications. Apweiler stressed the need for appropriate technologies, cooperation, data sharing and integration in order to tackle the individual proteomes of cells, tissues, and organisms during growth and development. Compared to the 'mere' 3 billion bases of the human genome, a definitive catalog of the expression pattern of each and every protein in a human being appears to be a Herculean task.

Taking it into the clinic

The clinical session focused on the impact of genomics on two major human disease areas, cardiomyopathies and cancer. Norbert Frey (University of Heidelberg, Germany) described examples of candidate gene approaches for dilated cardiomyopathy, starting from an *in silico* identification of potential effectors, and leading via ORFeome resources to *in vitro* and *in vivo* validation in zebrafish and mouse models. Frey and his colleagues have identified one gene that is specifically expressed in the heart and localizes to the Z-discs of sarcomeres. When this gene was knocked-down in zebrafish, severe cardiomyopathic phenotypes were seen. This and other genes identified in this study are currently being screened for mutations in patients with cardiomyopathy with the aim of improving diagnosis.

Alexander Marmé (University of Tübingen, Germany) raised the question of what impact genomics had already had on the prognosis and treatment of cancer patients. In breast cancer, the age of the patient, tumor grading, and the receptor status (estrogen receptor, progesterone receptor or ErbB2) are currently utilized to decide on a therapeutic regime. But a decision based on so few biological markers is often of little benefit. A number of gene signatures for breast cancer have already been approved (for example, Oncotype DX (Genomic Health) and MammaPrint (Agendia)) or are in clinical testing (for example, H3E-MC-So80). They are based on sets of molecular and non-molecular predictive factors and should permit tailored therapies, according to Marmé. They should be better suited to a fine-grained stratification of patients, allowing personalized therapies and decreasing the likelihood of overtreatment or wrongly treating patients.

A complex interplay between tumor and stromal cells was highlighted by Daniel Mertens (University Hospital, Ulm, Germany). He and his collaborators found that chronic lymphatic leukemia cells quickly die in culture unless they are co-cultured with nurse-like stroma cells. Testing which factors convey the survival message to the tumor cells, they found IL-4 and CD40 to be most effective. The finding was then validated in samples from patients. This stresses the importance of paracrine signals for the growth and survival of tumor cells, and emphasizes the need to study cancer cells within their complex environment.

As the meeting showed, one major challenge is the need for cooperation between different disciplines to push forward and exploit the 'omics' sciences. "We are all looking at the same elephant, just from different angles", says Yarden. "It could turn out in the end though that it had been an octopus all the time", adds Sorger. Acquiring knowledge at the systems level raises the hope that a more comprehensive understanding of cells and tissues in health and disease will open up new avenues for the treatment of patients.

Acknowledgments

This meeting report is dedicated to Annemarie Poustka, a pioneer in genomics and genome biology and one of the organizers of Systems Genomics 2008, who died on 3 May, 2008.