

Meeting report

Shaken not stirred: a global research cocktail served in Hinxton

Samuel Marguerat*, Brian T Wilhelm*[†] and Jürg Bähler*

Address: *Cancer Research UK Fission Yeast Functional Genomics Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. [†]Present address: Institut de Recherche en Immunologie et en Cancérologie (IRIC), Montreal, H3C 3J7 Canada.

Correspondence: Jürg Bähler. Email: jurg@sanger.ac.uk

Published: 22 November 2007

Genome Biology 2007, **8**:320 (doi:10.1186/gb-2007-8-11-320)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/11/320>

© 2007 BioMed Central Ltd

A report of the 2007 Cold Spring Harbor Laboratory/Wellcome Trust Conference on Functional Genomics and Systems Biology, Hinxton, UK, 10-13 October 2007.

The organizers of the 2007 Cold Spring Harbor Laboratory/Wellcome Trust Conference on Functional Genomics and Systems Biology built on the tradition of past workshops by keeping the number of participants low and choosing presentations covering a wide range of topics on multiple aspects of global genomic and systems-biological approaches. Like in a good cocktail, the varied talks blended into an interesting mix. Here we present a selection of talks with emphasis on unpublished work.

Genome-wide cellular screens

Several talks introduced global cellular screens, covering data-intensive experiments and efforts to improve the design and readout of current approaches. Brenda Andrews (University of Toronto, Canada) described an elegant system to screen for factors involved in the regulation of periodic gene expression during the budding yeast cell cycle. She and her colleagues developed a two-color reporter assay with cell-cycle-regulated promoters driving expression of green fluorescent protein (GFP) and a control promoter driving red fluorescent protein (RFP) expression. Using the synthetic genetic array (SGA) platform, expression levels can be assayed in combination with 5,000 deletion mutants or with overexpressing strains. Readouts are the GFP/RFP ratios in the different yeast colonies, measured with a scanner. This approach rapidly identified both known and new cell-cycle regulators. For example, expression of the histone gene *HTA1* was positively regulated by several factors, including the acetyltransferase Rtt109.

Insuk Lee (University of Texas, Austin, USA) presented new developments in probabilistic functional gene networks, which are built from heterogeneous genomic data and provide evidence for 'functional coupling' between genes, that is, probabilities that genes participate in the same process. He and his colleagues used the networks to predict genes most likely to participate in a given molecular process, thus reducing the search-space for cellular screens - an approach called network-guided focused reverse genetics. Lee and colleagues successfully applied this technique in budding yeast, using the YeastNet resource developed by the group [www.yeastnet.org], to discover new members of the ribosome biogenesis pathway; it also proved effective in predicting knockout phenotypes. In a related talk, Andrew Fraser (Wellcome Trust Sanger Institute, Hinxton, UK) reported the construction of probabilistic functional gene networks in *Caenorhabditis* and the development of WormNet [www.functionalnet.org/wormnet]. While searching for new candidates for the dystrophin pathway, WormNet predicted an unexpected connection between the dystrophin and epidermal growth factor (EGF) pathways. This connection was validated by showing that knockdown of members of the dystrophin pathway caused EGF phenotypes. Julie Ahringer (Gurdon Institute, Cambridge, UK) described double RNA interference (RNAi) screens in *Caenorhabditis* to systematically search for functionally redundant duplicated genes. Surprisingly, only around 4% of the genes tested were functionally redundant compared with 15% of unique genes showing an RNAi phenotype, indicating that redundancy among duplicated genes does not account for the low frequency of RNAi phenotypes observed in the worm. Duplicated pairs with one gene on an autosome and one on the X chromosome were enriched among functionally redundant genes, possibly to ensure expression in the germline when the X chromosome is inactivated.

Another set of talks dealt with the analysis of large screens in tissue culture cells. Chris Bakal (Harvard Medical School, Boston, USA) described how quantitative morphological signatures, a method for automatically characterizing changes in cell morphology in tissue cultures, can be used together with double RNAi transfections to search for synthetic phenotypes in *Drosophila* cell lines. He and his colleagues also devised an elegant screen for components of the Jun N-terminal kinase (JNK) pathway by targeting all kinases and phosphatases by RNAi in cells producing fluorescence by intramolecular FRET in response to JNK activity. This screen identified several new components of the pathway.

Global mapping of transcription factors

Several talks focused on identifying DNA-binding sites for transcription factors by chromatin immunoprecipitation followed by microarray analysis (ChIP-chip) to gain insight into regulatory networks. Stewart MacArthur (Lawrence Berkeley National Laboratory, Berkeley, USA) presented a large dataset for 18 fly transcription factors using tiling microarrays. He and his colleagues identified binding sites for multiple factors near individual genes, suggesting a high level of cooperative regulation. Intriguingly, many binding sites for transcription factors were present within exons. These results, together with those of Eileen Furlong (EMBL, Heidelberg, Germany), suggest that the conservation of *cis*-regulatory elements is of limited use for predicting binding sites. ChIP-chip data can also provide insight into the dynamics of enhancer occupancy. Furlong reported ChIP-chip studies that followed the binding of Twist, Tinman, Mef2, and other developmental transcription factors during the development and differentiation of the *Drosophila* mesoderm. These time-course data enabled the temporal changes in target sites bound by various factors to be distinguished, showing that the same transcription factor binds to enhancers of different subsets of genes in coordination with changing target gene expression and cellular states within the embryo.

Duncan Odom (Cancer Research UK, Cambridge, UK) presented ChIP-chip data from a study of binding sites for orthologous transcription factors for genes expressed in the liver in human and mouse. Surprisingly, only around 20-25% of the binding sites were conserved, suggesting that binding sites can rapidly diverge even if transcription factor targets remain conserved. Indeed, only a third of all binding events occurred in aligned regions of synteny between the orthologous target genes. Preliminary data from mice containing a copy of human chromosome 21 suggest that the binding sites on the human chromosome correspond to those found in human cells, providing intriguing insights into the influence of *cis* and *trans* regulatory effects.

Claes Wadelius (Uppsala University, Sweden) discussed both ChIP-chip and ChIP followed by DNA sequencing

(ChIP-seq) as methods for mapping liver transcription factors. The high-throughput, unbiased nature of ChIP-seq makes it a powerful method for mapping protein-binding sites. Among 35 million sequence reads of potential binding sites for HNF3 β , around 15,000 hits were mapped back to the genome. The majority of binding sites for HNF3 β were not in promoter regions of genes, but correlated with upstream stimulatory factor 2 (USF2) homodimer-binding sites predicted from ChIP-chip data. These results show that we are at or close to the theoretical resolution in assigning histone modification status and transcription factor binding sites to chromatin in genome-wide studies.

Synthetic biology and transcriptional networks

Synthetic biology approaches are being applied to learn more about transcriptional mechanisms and networks. Barak Cohen (Washington University, St Louis, USA) is developing quantitative models to predict transcript levels based on *cis*-regulatory promoter elements. He and his colleagues built libraries of yeast reporter genes containing random combinations of activating and repressing promoter elements and measured the transcript levels of the different synthetic constructs using a fluorescent reporter. Even this relatively simple 'toy system' shows plenty of nonlinear behavior such as cooperativity, orientation effects and epistatic interactions between regulatory elements. Weak regulatory elements play virtually no role in gene expression on their own, but the presence of a strong element can convert a weak into a strong element. Cohen and colleagues are also measuring occupancy of transcription factors combined with physical modeling to capture actual cellular chemistry during transcription.

Anat Bren (Weizmann Institute of Science, Rehovot, Israel) has developed another inventive bottom-up approach. She and her colleagues are interested in the gene input function: the relation between levels of multiple environmental signals and the transcription rates of response genes. Using the *Escherichia coli* sugar-metabolism genes as a model system for a two-dimensional input function, expression levels of each gene were measured under 96 combinations of cyclic AMP and sugar concentration. This broad, quantitative survey of input functions revealed diverse and sophisticated responses, highlighting the need for high-resolution measurements to fully understand the computations done by the cell.

Luis Serrano (Centre for Genomic Regulation, Barcelona, Spain) described a systematic study to explore the effects of rewiring gene networks in *E. coli*. By shuffling promoters and transcription factor genes, he and his colleagues created 600 recombined constructs that added new links to the regulatory network without deleting regular links. Surprisingly, around 95% of these constructs are fully viable, and global gene-expression changes are limited. Under

certain conditions, however, specific constructs consistently survive better than wild-type cells. Thus, bacteria can both tolerate and exploit radical changes in regulatory circuitry. It will be interesting to see whether eukaryotic networks are similarly robust to rewiring.

Computational approaches to evolution

Several talks described 'dry' projects to tease out novel biological insight from published data. Sarah Teichmann (Laboratory for Molecular Biology, Cambridge, UK) analyzed how variation in protein sequences contributes to diversity between animal species and among humans. Using a normalized conservation score, they find that enzymes are generally more conserved than regulatory proteins. Other slowly evolving proteins function in metabolism, cell structure or chromatin, whereas proteins related to environmental responses or immunity evolve more rapidly. Interestingly, proteins functioning in transcriptional control or development are conserved within mammals but have diverged in invertebrates, reflecting an evolutionary transition. Some transcription factors show human-specific selection in positions that are conserved in other mammals, indicating distinct evolutionary constraints in humans.

Global organization of metabolism in *E. coli* is surprisingly poorly understood, according to Nick Luscombe (European Bioinformatics Institute, Hinxton, UK). He and his colleagues integrated the *E. coli* metabolic network with both direct and indirect regulatory networks corresponding to rapid control of enzyme activity or much slower control of enzyme concentration, respectively. This research gives comprehensive insight into how direct and indirect control mechanisms selectively regulate catabolism and anabolism by coordinating reaction time scales, specificities and concentrations. As an example, direct regulation is mainly used for anabolic pathways, while indirect regulation is used for both catabolic and anabolic pathways.

Metabolic networks not only teach us about regulatory principles, but they also reflect the environments in which organisms evolved. Eytan Ruppin (Tel-Aviv University, Israel) reported the application of metabolic network analyses to 478 species to infer their growth environments and evolutionary dynamics. Using a graph-theory-based algorithm, he and his colleagues determined the 'seed' compounds, defined as the minimum subset of metabolites that cannot be synthesized from other compounds and need to be extracted from the environment. A phylogenetic tree based on seed compounds reflects taxonomic groups remarkably well. This imaginative approach allows the reconstruction of current and ancient environments from metabolic networks, providing a glimpse into evolutionary history.

Tools and resources

Several useful tools and resources were also described. Alvis Brazma (European Bioinformatics Institute) talked about ongoing efforts to build a gene-expression atlas to mine combined microarray datasets available in public repositories. One approach takes advantage of around 6,100 high-quality hybridizations from a standardized human DNA microarray platform. After normalization and annotation of different conditions (samples), a meta-analysis produces biologically coherent clusters of samples. This merged experiment is available under the ArrayExpress accession number E-TABM-185. Combining experiments from different array platforms is more challenging, and relies on a qualitative assessment of gene expression. Initial tools available in ArrayExpress allow one to find the most informative experiments relating to a gene of interest. Further developments will be crucial to get the most from the increasing amounts of publicly available data. Along similar lines, Tom Freeman (University of Edinburgh, UK) introduced BioLayout [www.biolayout.org], another promising resource to mine large microarray datasets. Based on a simple calculation of correlations between all pairwise combinations of genes combined with powerful visualization, this tool provides a fast, reproducible and intuitive way to construct and analyze large network graphs. Built-in data-mining modules and a highly interactive interface let you explore relationships between large numbers of genes.

Stefan Wiemann (German Cancer Research Center, Heidelberg, Germany) promoted the initiative to capture the 'minimum information about a cellular assay' (MIACA) [<http://miaca.sf.net>]. Researchers and reviewers are increasingly overwhelmed with too much data that are poorly documented. MIACA aims at a standardized description of high-throughput cell-biological analyses, which will help to compare and integrate different datasets and enhance their long-term usability. A manuscript describing MIACA is currently under public review with *Nature Biotechnology*, and everybody can give feedback on its usefulness. Researchers were also encouraged to join and directly contribute to this initiative.

Functional genomics and systems biology are rapidly evolving and diverging in unpredictable and exciting directions. We can look forward to the next meeting in this series in two years time in the tranquil village of Hinxton, which we expect to change much less than the research field motivating the conference.