Genome **Biology**

MEETING REPORT

Dynamic systems

Chris Bakal*

Abstract

A report of the Wellcome Trust Functional Genomics and Systems Biology Conference, Hinxton, UK, 29 November to 1 December 2011.

One of the central aims of systems-biology is to gain a holistic, quantitative, and predictive understanding of how the complex interplay between genotype and the environment determines specific phenotypes in the context of health and disease. Pioneering efforts over the past decade have led to insights that are impressive in scale but are often static portraits of dynamical cellular systems. Moreover, initial systems-biology research has done little to address how genetic variation impacts cell systems architecture and information flow. However, work presented over 3 days at this conference consistently demonstrated that striking advances have occurred that now allow the study of systems over time as well as in specific cell types of the same organism, and that it is now possible to gain understanding of how genome sequence and epigenetic modification affects cellular systems. Not only did speakers at the meeting describe their work on model organisms, but there were also first glimpses of how systems-biology approaches can be used to decode the oncoming flood of personalized genomic data.

Quantifying the number of parts

A long-standing question that is highly relevant to all of biology is how many genes, and at what level, are expressed in any cell. Using deep sequencing methods, Jürg Bähler (University College London) demonstrated that in *Schizosaccharomyces pombe* global gene expression follows a continuous normal distribution, meaning that there is some expression of most genes. However, he noted that some genes, such as stress response genes, are expressed at very low levels. Through integration of data derived by thousands of expression experiments and

deposited in Array Express, Alvis Brazma (EMBL European Bioinformatics Institute) also provided evidence from human cells that most genes are in fact expressed in most cell types. Although there are differences in gene expression between different cell types (neuronal, blood, cancer), they are much less than perhaps might be expected from functional or morphological differences. Stephen Watt (CRUK Cambridge Research Institute) and Lars Dölken (University of Cambridge) demonstrated the existence of non-coding RNAs with rapid turnover - species of RNA that probably evaded detection before the development of recent technologies, but that are widely expressed. Whether this abundant gene expression is essential for survival, or rather that exquisite regulation of transcription would waste valuable resources, is still an open question. By creating haploinsufficient yeast strains, Stephen Oliver (University of Cambridge) showed that the genes required for growth are different from one medium to the next. The Oliver group identified another set of strains, which he termed 'haploproficient', in which reduction in gene dosage accelerates growth. The implication is that some genes have evolved as part of quality control mechanisms to slow growth and proliferation to ensure fidelity in DNA replication. Together this work overturns the long-held notion that there is cell-type specificity in the expression of different RNAs, and forces the revision of many models that propose of how phenotypic specificity is generated.

Although RNA sequencing methods and analysis of massive datasets such as ArrayExpress provide strong evidence that most genes are expressed, there is still some controversy as to whether this is reflected in protein expression. However, proteomic analysis by Bähler demonstrated that most proteins are expressed in fission yeast, and by using the antigenic peptides generated for antibody production by the Human Atlas Project as standards in quantitative mass-spectrometry experiments, Mathias Uhlen (Royal Institute of Technology, Stockholm) has confirmed that most of these mRNAs are also translated in metazoan cells. Interestingly, cell-specific proteins tend to be expressed at the surface of cells.

Despite the fact that gene expression may be more ubiquitous than previously imagined, there is very low expression of some genes in specific cell types. Using a combination of experimental and statistical techniques,

*Correspondence: cbakal@icr.ac.uk
The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Sarah Teichmann (MRC Laboratory of Molecular Biology, Cambridge) demonstrated the existence of two mRNA populations in mouse T helper 2 cells, a largely normally distributed population of widely expressed mRNAs and a subpopulation of mRNAs that were expressed in a noisy manner in only a handful of cells. Some of these noisy genes are critical fate determinants, which suggests that, like the stress response in yeast, differentiation may occur in a probabilistic manner. The idea that particular cells in a population may be better suited for unexpected changes in the environment is consistent with work I presented showing that there is phenotypic heterogeneity in normal cells that may facilitate rapid response of a population to diverse signals. One model that might unite many observations presented at this meeting is that, in most cells, the bulk of genes are typically expressed, and all are probably largely essential for population survival in fluctuating environments, but that there are a select group of genes expressed in a stochastic manner and/or in sub-populations, which allows individual cells in the population to survive acute stress or undergo rapid differentiation as part of a bet-hedging strategy.

Transcription networks

Many researchers provided systems-level mechanistic insights into transcription across time and in different cell types, which have been realized through the development of new experimental and computational methods. To describe transcriptional dynamics on a systems level, Ido Amit's group (Weizmann Institute of Science) has devised iChIP, a cost-effective and high-throughput means by which to monitor the binding of multiple transcription factors (TFs) across time. Eileen Furlong (EMBL) also presented an elegant system by which to monitor TF binding in different cell types isolated from *Drosophila* embryos during development. Simon Tavaré (University of Cambridge) also stressed the importance of monitoring transcriptional regulation over time as he demonstrated a robust method that combines ChIP-seq approaches with Bayesian-based methods to monitor temporal replication schedules. It is striking that, even recently, performing ChIP-seq to monitor the binding of multiple TFs across time in specific cells would have been considered unfeasible, but that numerous laboratories can now undertake such studies routinely.

Not only is the development of new technologies changing how we should think about which genes are expressed in time and space, but it is also providing new ideas concerning non-coding regions of the genome that regulate this expression. Through use of deepCAGE methods, Piero Carninci (RIKEN) provided evidence that the expression of genes from promoters is 'sharp' - in that it consistently begins from the same exact site in the genome - rather the majority of genes being expressed in a 'broad' manner where transcription of the same gene begins at many different sites. Broad expression correlates with the presence of CpG islands, whereas TATA boxes delineate sharp expression boundaries. deepCAGE sequencing also identified non-coding RNAs that stabilize proteins. Duncan Odom (University of Cambridge) showed that transcriptional output is not dictated by any single TF or motif in a *cis*-regulatory module (CRM), but rather by binding of TF complexes. Within CRMs TF binding motifs can play a 'ring-around-the-posey' in terms of their ordering, with little phenotypic consequence - which probably allows sequence within the CRMs to rapidly evolve. Byung-Kwan Cho (Korea Advanced Institute of Science and Technology) provided a nearly complete systems-level view of transcriptional regulation in *Escherichia coli* and showed that different sigma factors regulate different classes of transcriptional start sites.

Of course, to fully understand the regulation of transcription it is essential to describe the architecture and dynamics of signaling networks that respond to environmental flux by altering gene expression. Frank Holstege (University Medical Center Utrecht) described the combined use of expression profiling, high-throughput genetics, and statistical approaches to gain remarkable insights into the architecture of signaling networks regulating transcription. Brenda Andrews (University of Toronto) has used high-content microscopy of *Saccharomyces cerevisiae* strains in which individual proteins have been fluorescently tagged to quantify changes in the localization of the *S. cerevisiae* proteome in response to different signals. Such changes in protein localization enable specific phenotypes to be generated even when most genes are expressed in a cell. Stefan Wiemann (German Cancer Research Center) showed how signaling pathways regulating transcription in breast cancer cells are regulated by microRNAs. Finally, through live cell imaging Michael White (University of Manchester) discussed how different TFs show different oscillatory behaviors in terms of nuclear localization. Not only do different oscillatory dynamics of a single TF influence the genes that a TF will regulate, but TFs can affect each other's oscillatory behavior in complex ways.

Much of systems research is aimed at using data to generate models that will predict phenotype based on specific parameters (such as genotype and environment), and a large body of research has focused on the development of modeling techniques. Nick Luscombe (EMBL European Bioinformatics Institute) presented a simple but effective means to predict pattern formation in *Drosophila* embryos in three dimensions based on large-scale image datasets, and Tom Freeman (University of Edinburgh) described a system to visualize the dynamic modularity of macrophage regulatory networks.

Personalized dynamics

A particularly exciting aspect to this meeting was that discussions of systems approaches in model organisms were integrated with work focused on the development of systems-based tools to be used for personalized diagnosis and therapy. Mike Snyder (Stanford University) discussed the highly personal subject of the comprehensive profiling of his own genome, transcriptome, proteome, and metabolome over a period of years. Through this work he has not only been able to link much of his own genomic variations to his phenotype, but it has also led to changes in his lifestyle as he witnessed in real time the onset of his diabetes, which was foreshowed by his genomic sequence. The theme of monitoring disease phenotypes over time was continued by William Cookson's (Imperial College London) description of epigenetic profiling of asthma patients over time. For example, he showed that methylation status can be affected by the seasons. Mark McCarthy (University of Oxford) described the development a network-based strategy to understand how the loci associated with type 2 diabetes drive the development of the disease. Finally, in thought-provoking dissection of scientific fraud, Keith Baggerly (MD Anderson Cancer Center) highlighted the fact that as basic systems-biology work has ever growing impact in the clinic, scientists must ensure that their methods and findings are reproducible before patients are treated based on research outcomes. Baggerly stressed that high-throughput data, the key to decipher the data, data analysis code, and descriptions of the non-scriptable steps must be made freely available.

Although all biologists appreciate the fact that no phenotype is static, but rather they are constantly changing over time, studying cellular, tissue, or organism dynamics has long been considered challenging. At Hinxton, however, it became clear that technology now exists not only to quantify biological dynamics, but to do so in specific cell types and on a systems level.