Meeting report

Transcriptional networking

Sarah A Teichmann*, Erich Bornberg-Bauer† and Nicholas M Luscombe‡

Addresses: * Structural Studies Division, MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK. †School of Biological Sciences, The Westfalian Wilhelms University of Münster, Botanical Institute, 48149 Münster, Germany. ‡EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK.

Correspondence: Sarah A Teichmann. E-mail: sat@mrc-lmb.cam.ac.uk

Published: 31 August 2005

Genome ***Biology*** 2005, **6:**344 (doi:10.1186/gb-2005-6-9-344)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/9/344

© 2005 BioMed Central Ltd

A report on the European Science Foundation workshop 'Transcription Networks: A Global View', Madrid, Spain, 26-28 May 2005.

The European Science Foundation (ESF) workshop on transcriptional regulatory networks held in Madrid this spring, and sponsored by the ESF Programme for Integrated Approaches for Functional Genomics [http://www.functionalgenomics.org.uk/sections/programme], focused on the theme of transcriptional regulation in the broadest sense. Topics presented ranged from theoretical approaches to experimental work, and from small systems to studies on many thousands of transcriptional regulatory interactions. The small size of the meeting allowed for ample discussion, which provided a lively atmosphere of scientific dialog. To describe a transcriptional regulatory network, one first needs to know which regions of the genome are transcribed. Roderic Guigó (Centre de Regulacio Genomica, Barcelona, Spain) presented the current status of gene annotation in the human genome. Although the genome sequence itself is now very accurate, our knowledge of the number, location and splice variants of genes is still far from complete. He emphasized that whole-genome microarrays harbor great promise for shedding more light on both gene location and splice variation.

The way that genes are transcribed and the timing of expression relies heavily on the higher-order, three-dimensional structure of DNA. Francois Kepes (Genopole, Evry, France) presented a solenoidal model of chromatin structure in the budding yeast *Saccharomyces cerevisiae* that is based on the observed periodicity of binding sites for transcription factors and chromatin-remodeling factors along each chromosome. This particular model potentially allows a small number of factors to efficiently influence the transcription of a relatively large number of genes. Looking at *Escherichia coli*, David Ussery (Center for Biological Sequence Analysis, Lyngby, Denmark) described the role of DNA supercoiling and bacterial 'chromatin' proteins in transcriptional regulation at the most basic level, where many genes are expressed in a relatively 'sloppy' and unregulated manner. He described a weak correlation between gene expression and the predicted DNA curvature, based on GC content, which is abolished in mutants of the chromatin protein HNS.

Below this gross level of gene regulation through chromatin structure, finer control is achieved by the binding of specific transcription factors to *cis*-regulatory motifs. Rekin's Janky (Université Libre de Bruxelles, Belgium) presented a method for detecting potential transcription-factor binding sites in prokaryotes by identifying over-represented dyads (inverted or direct DNA sequence repeats separated by a spacer) in combination with phylogenetic footprinting. In vertebrates, such motifs are often detected computationally through searches using position-specific weight matrices. Mar Albà (Universitat Pompeu Fabra, Barcelona, Spain) presented an assessment of various database compilations of weight matrices in terms of their accuracy in identifying genuine transcription-factor binding sites. She described how detection is improved by including positional constraints between motifs. A closer study of conserved sequence blocks in human and mouse promoters revealed that tissue-specific genes have the most highly conserved promoters, whereas those of evolutionarily older genes that are expressed in a greater range of tissues have fewer regions under selective constraint. The *trans*-regulatory elements that bind to such motifs are generally transcription factors. One of us (S.A.T.) presented a generic method for predicting the repertoire of DNA-binding transcription factors in a complete genome; the method is based on detecting distant sequence homologies to known

DNA-binding domains using hidden Markov models. The transcription factor annotations derived by this method are available for many complete genomes in a transcription factor database called DBD [http://www.transcriptionfactor.org].

In a transcriptional system, the individual components interact with each other; these interactions include both protein-protein interactions among transcription factors and regulatory interactions between transcription factors and their target sites in DNA. These interactions - and thus the system - can be collectively represented as a network. Alvis Brazma (EMBL-European Bioinformatics Institute, Cambridge, UK) discussed how these networks can be studied at different levels of detail, ranging from a whole-genome scale that enables global graph-theoretical analysis down to a single-gene perspective that could allow for control logistic models of systems such as the yeast cell cycle.

Focusing on protein interactions in gene regulation, it is often important to know whether transcription factors act as dimers or physically interact with other non-DNA-binding components in a regulatory pathway. This information may be obtained from high-throughput proteomic experiments, and Benno Schwikowski (Institut Pasteur, Paris, France) described computational approaches that would allow reliable interpretations of such data, for example, by integration of data from different mass spectrometry experiments. A proteomic dataset of different time points in the yeast cell cycle was introduced. John Hancock (MRC Mammalian Genetics Unit, Harwell, UK) explained that many transcription factors - particularly those of the *Drosophila melanogaster* genome - contain simple amino-acid repeats that are likely to promote protein-protein interactions. One of us (E.B.B) showed that several representative transcription factor families in metazoans have evolved dimeric interactions through a series of single-gene and whole-genome duplications.

At a larger scale, Martijn Huynen (Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands) is studying the evolutionary acquisition of new subunits by complexes of the respiratory electron transfer chain and described how the protein-interaction network rarely grows by duplicating entire collections of nodes, but rather in a piecemeal fashion by introducing individual proteins and accompanying interactions. At a more abstract level, Ricard Sole (Universitat Pompeu Fabra, Barcelona, Spain) described how minimal models for network growth (using duplication, deletion and divergence of nodes) can reproduce many general features of biological molecular networks, such as degree and clustering coefficient distributions.

The interactions between transcription factors and their DNA binding sites, and the regulatory effect on the downstream gene, are an essential part of a transcriptional network. Although many such details are unknown for many networks, a potentially powerful approach for predicting regulatory protein-DNA interactions is the use of gene-expression data to reverse-predict such interactions. Joaquin Dopazo (Centro de Investigación Príncipe Felipe, Valencia, Spain) highlighted potential pitfalls in interpreting these types of data, and discussed how robust statistical methods must be used to extract meaningful conclusions. He also suggested how regulatory relationships might be inferred between pairs of genes by searching for complex correlations between gene-expression profiles.

Julio Collado-Vides (Universidad Nacional Autónoma de México, Cuernavaca, México) described the RegulonDB data resource - carefully compiled from reported results in the literature - which currently describes about a quarter of the *E. coli* transcriptional regulatory system [http://www.cifn.unam.mx/Computational_Genomics/regulondb]. Of interest is the internal organization of the resultant network: for example, there are distinct regulatory modules corresponding to different cellular functions, and regulatory events can be classified according to whether they are triggered by internal or external stimuli. The challenges of identifying these partitions in highly interwoven networks, however, are compounded by the fact that we often do not even know the correct paths through which a signal travels. Jacques van Helden (Université Libre de Bruxelles, Belgium) highlighted such pitfalls in relation to graph analysis, and demonstrated the use of a path-finding algorithm applied to metabolic pathways so as to tackle the problem. He described how, by preferentially tracing through nodes that have fewer connections, it was possible to distinguish biologically relevant paths from spurious ones.

Finally, the transcriptional network is not static, but is used dynamically; by combining diverse biological data with the knowledge of regulatory components it is now possible to examine these dynamic properties. Returning to the detailed level, in separate talks, Hidde de Jong (Institut National de Recherche en Informatique et en Automatique (INRIA), Montbonnot, France) and Adrian Garcia-Lomana (Universitat Pompeu Fabra, Barcelona, Spain) described their independent demonstrations that well studied mathematical techniques (such as variants of differential equations in these cases) can be successfully adapted to simulate small bacterial systems such as initiation of sporulation in *Bacillus subtilis* and nutritional stress in *E. coli*. Jan Kim (University of East Anglia, Norwich, UK) presented a formal language for describing regulatory systems, called transsys, which, in combination with Lindenmayer systems (a mathematical theory of plant development), can model plant growth patterns under different conditions, ranging from a single cell to the whole *Arabidopsis* plant. At the genomic scale, two of us (S.A.T. and N.M.L.) described the integration of gene-expression data in order to examine the dynamic usage of transcription factors and their regulatory interactions under multiple cellular conditions such as the cell cycle and sporulation in *S. cerevisiae*.

In one of the few presentations of experimental work, Frank Holstege (Universitair Medisch Centrum Utrecht, Utrecht, The Netherlands) discussed his recent study of gene-expression measurements during the yeast growth cycle. Careful analysis of these data depicted waves of transcription that bring about the transitions between distinct cellular states, in particular, exit and entry into stationary phase. He showed how epistasis studies with microarray analysis revealed the crucial role of the Mediator complex for integrating positive and negative signals and transducing these to the RNA polymerase. He also described how gene-deletion experiments combined with microarray analysis can reveal epistatic genetic interactions.

The many interesting seminars at this workshop covered a wide range of topics, from the structure and evolution of transcriptional systems to their regulatory kinetics both at detailed and whole-organism levels. The quality of the presentations combined with the enthusiasm of the meeting participants clearly reflected the importance of studying transcription regulation. Although we are still far from understanding such systems fully, the continually strengthening ties between bioinformaticists and experimentalists will surely allow us to advance this field at an ever-increasing pace.