Meeting report
Interactome networks: the state of the science
Guy J Warner*, Yeyejide A Adeleye* and Trey Ideker[†]

Addresses: *Safety and Environmental Assurance Centre, Unilever Colworth Laboratory, Sharnbrook, Bedfordshire MK44 1LQ, UK.
[†]Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, Mail Code 0412, La Jolla, CA 92093, USA.

Correspondence: Guy J Warner. Email: Guy.Warner@Unilever.com

A report on the joint Cold Spring Harbor/Wellcome Trust Meeting 'Interactome Networks', Hinxton, UK, 31 August-4 September 2005.

High-throughput analyses are identifying the DNA, RNA, proteins and metabolites within a biological system with increasing accuracy and speed. As a result, we now have a relatively detailed understanding of the components that make up the dynamic and temporal characteristics of the cell. In most cases, however, we know very little about how the individual components work together to carry out specific biological functions. To get over this hurdle, it will be necessary to map how individual biomolecules interact with one another within a larger network of molecular interactions (the so-called 'interactome') in the cell as a whole.

Mapping this network is the shared goal of an increasing number of researchers from the UK, Europe, US, and Japan, who gathered at the first annual Cold Spring Harbor/Wellcome Trust meeting on interactome networks in Hinxton. This meeting provided an opportunity to review the recent experimental and computational advances that have been applied to uncover biomolecular interactions. Here we report a few of the key advances in the areas of new interaction-mapping techniques, new experimental reagents and resources, and new computational tools for understanding interaction networks that were presented at the meeting.

Assays for interactome network mapping
Although tens of thousands of protein interactions have been uncovered so far, these are only a small part of the complete set of pairwise interactions between all proteins. Therefore, one of the main themes at the meeting dealt with current approaches and resources available for mapping interactions efficiently across a variety of organisms, including *Saccharomyces cerevisiae, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster* and humans. Chris Sanderson (University of Liverpool, UK) described how the focused application of the yeast two-hybrid system can provide insights into disease pathways. Protein-protein networks were generated for specific biological processes (including DNA degradation, multivesicular formation and ubiquitin conjugation) using stringent yeast two-hybrid screens, and the resulting protein-protein interaction networks were used to gain insights into the role of previously uncharacterized proteins in hereditary spastic paraplegia, which is most commonly caused by mutations in the gene (*SG4*) for the protein spastin. Several proteins were identified as interacting partners with spastin, including one called CHMP1B, which is associated with the endosomal sorting complex required for transport (ESCRT) III. This suggests that spastin has a role in intracellular membrane trafficking, supporting the hypothesis that defects in intracellular membrane traffic are a significant cause of motor neuron pathology.

An alternative interaction mapping strategy was described by Anne-Claude Gavin of Cellzome (Heidelberg, Germany). Cellzome has a long track record of applying tandem affinity purification (TAP) in conjunction with liquid-chromatography tandem mass spectrometry (LC-MS/MS) to map protein interactions, and Gavin reported that this integrated approach has now been used to elucidate protein interactions in the pathway mediated by tumor necrosis factor α (TNFα) and the transcription factor NFκB. Cellular complexes containing one or more of 32 TNFα/NFκB pathway components were isolated using TAP and then screened *in vivo* for both constitutive and signal-induced protein interactions. A total of 680 interacting proteins were identified using LC-MS/MS, of which 33 were dependent on stimulation with TNFα or NFκB-inducing kinase. Upon filtering the

network for specificity and consistency, 131 high-confidence interactors remained. Network analysis combined with sequence information for the proteins was then used to select 28 candidates for functional validation using systematic directed perturbation by RNA interference (RNAi), in which ten proteins showed reproducible responses consistent with a modulatory role in TNFα/NFκB signal transduction. This combined and rigorous elucidation of a medium-scale map of human protein interactions shows that specific components of human pathways can be elucidated and subsequently validated with RNAi.

Another approach to the determination of protein complexes was covered by Oda Stoevesandt (University of Tübingen, Germany), who used a combination of peptide microarrays and fluorescence spectroscopy to investigate the early stages of T-cell signaling. Microarrays of peptides corresponding to motifs that interact with binding domains of signaling proteins were used to detect activation-dependent changes in protein-protein interaction patterns; proteins binding to the array were detected by indirect immunofluorescence techniques. Stoevesandt suggested that such a method can detect protein interactions more rapidly than established techniques.

Resources for interaction mapping projects

A prerequisite for determining a protein interaction network *ab initio* is to clone the open reading frames (ORFs) that encode each protein in the network. Gary Temple (National Institutes of Health, Bethesda, USA) discussed the work being done by the Mammalian Genome Collection (MGC) [http://mgc.nci.nih.gov], which aims to provide a community resource of publicly accessible full-length ORF clones of human and mouse protein-coding genes. At the time of writing, MGC has more than 20,000 and 16,000 full ORF clones for human and mouse, respectively.

Susan Celniker (Lawrence Berkeley National Laboratory, Berkeley, USA) presented the work being done by the Berkeley *Drosophila* Genome Project to produce the *Drosophila* Gene Collection (DGC) [http://www.fruitfly.org/DGC], a resource of sequenced full-length *D. melanogaster* cDNAs. She also stressed the importance of accurately annotating the genomic sequences in such collections, which should include data that allow the temporal and spatial patterns of the expression of the genes to be understood. Studies that require the use of such ORFeome resources can be adversely affected by inaccuracies in gene prediction and annotation. An attempt to improve the quality and coverage of the *C. elegans* ORFeome resource Worfdb [http://worfdb.dfci.harvard.edu] was detailed by Philippe Lamesch of Marc Vidal's group (Dana-Farber Cancer Institute, Boston, USA). The first version of the database was based on an early release of the genome sequence (Wormbase WS9), and since then there has been a continuous effort to further annotate and better predict

coding regions. Lamesch presented recent work to update Worfdb with newly predicted ORFs; approximately 12,500 ORFs are now available that can be characterized using high-throughput techniques, including yeast two-hybrid approaches. This type of iterative ORFeome construction was strongly supported at the meeting and argues that the complete cloning of an organism's coding regions must be an ongoing process that takes into account improved gene predictions and incorporates strong experimental validation through PCR.

Jean-François Rual (also from Vidal's group) described how their resource of approximately 8,000 MGC clones transferred into Gateway entry vectors from Invitrogen, has been used to perform large-scale screens for human protein-protein interactions. A stringent yeast two-hybrid system was used to test all possible pairwise interactions between the products of the ORF clones, from which around 2,800 high-confidence interactions were detected, with 80% of them being were verified using co-affinity purification.

Rual's talk, along with several others, highlighted a major shift in interaction research from model species to human: two other human interaction networks were described at the meeting by Ulrich Stelzl (Max Delbrück Center for Molecular Medicine, Berlin, Germany) and Anne-Claude Gavin, respectively, based on primary mass spectrometry or two-hybrid assays. Interestingly, as is typical for high-throughput interaction screens, these new human interaction datasets have little apparent overlap with each other or with the previous literature. Although this situation could arise because of false-positive interactions (and it is difficult to rule these out), the principal cause is probably related to the low overall coverage of interactions: that is, if a single study sampled one fifth of the 'true' network, the expected overlap between two such studies would also equal one fifth - a seemingly low number. It was suggested that the community could collectively work to maximize coverage by producing 'networks of networks', built up from separate interaction-mapping initiatives, each of which would target a specific functionally related subset of proteins and interactions.

Computational analyses and tools for studying interactome networks

Apart from the recent high-throughput interaction datasets, interactions reported in the past scientific literature represent another valuable resource. Given the overwhelming number of relevant articles, the challenge is to effectively extract these interaction data from the literature. Along these lines, Edward Marcotte (University of Texas, Austin, USA) described a bioinformatic approach to predict human protein interactions based on automated literature extraction and homology to interaction networks of other species. He reported a final network of over 31,000 interactions between 7,748 human proteins. Allan Kuchinsky (Agilent Technologies, Palo Alto, USA) presented a freely available tool

[http://www.agilent.com/labs/research/mtl/projects/sysbio/sysinformatics/downloadv2.html] for constructing literature-based networks *de novo* that can be combined with experimental data. This tool has been implemented as a plug-in to the network modeling software Cytoscape [http://www.cytoscape.org].

Several initiatives aimed at generating interaction databases were described, including BIND [http://www.bind.ca/Action] (Chris Hogue, Samuel Lunenfeld Research Institute, Toronto, Canada), HPRD [http://www.hprd.org] (Akhilesh Pandey, Johns Hopkins University, Baltimore, USA) and Reactome [http://www.reactome.org] (Ewan Birney, European Bioinformatics Institute, Hinxton, UK). In addition, integration of interaction databases with visualization and analysis tools requires standards for the representation of molecular interaction data. Hennig Hermjakob (European Bioinformatics Institute) described the latest developments in the Protein Standards Initiative Molecular Interaction (PSI-MI) XML language, which is increasingly accepted as the standard for the exchange of interaction data. PSI-MI has recently been adopted by the partners in the International Exchange Consortium (IMEx) who will exchange their data in the form of XML files, following the PSI-MI standard [http://imex.sourceforge.net].

Beyond literature curation and databases, interaction networks present a number of major challenges to bioinformatics researchers, such as how to enrich for the true interactions in noisy measurements, how to best associate high-level information about protein interactions with functional roles and, most importantly, how to organize individual interaction measurements into higher-order models of cellular signaling and regulatory machinery. Several speakers discussed strategies for constructing pathway models through the integration of interaction datasets with each other and with other genomic sources. One of us (T.I.) presented methods for combining interaction and expression data to model regulatory pathways, and for comparing networks across species to identify their conserved regions, which then serve as markers for evolutionary change. In his closing remarks, Marc Vidal underscored the importance of assembling interactions into biological models that are freely accessible to all.

The proposed data sharing between groups is reminiscent of the early stages of the Human Genome Project. And indeed, the similarities go further: as with genome sequencing ten years ago, interaction-mapping projects are at different stages of completion for human and for each model organism; experimental techniques are still being optimized and further advances are needed; large databases are being constructed; and bioinformatic analyses are just beginning. We will see at future meetings how far this analogy can be extended, and whether, in fact, mapping the interaction network will have the same revolutionary impact on biology as mapping the human genome has had already.