

Meeting report

How is the *Drosophila* research community making use of the genome sequence?

Stephen Richards

Address: Human Genome Sequencing Center, Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX 77006, USA. E-mail: stephenr@bcm.tmc.edu

Published: 30 April 2003

Genome **Biology** 2003, 4:319

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/5/319>

© 2003 BioMed Central Ltd

A report on the 44th Annual *Drosophila* Research Conference, Chicago, USA, 5-9 March, 2003.

Drosophila researchers have capitalized on the genome sequence in the three years since it was released. From identifying paralogs and orthologs to microarray analysis of specific cell types, it is clear that the genome sequence is changing approaches to *Drosophila* research. In this review of the 2003 *Drosophila* research conference, I have focused on talks illustrating the various ways in which the genome sequence is being used.

The Larry Sandler Award is presented annually to the graduate student with the best thesis that uses *Drosophila* as a model system. This year's recipient, Sinisa Urban (University of Cambridge, UK), gave perhaps the best presentation of the meeting. His thesis focused on how the cell controls the release of the Spitz epidermal growth factor (EGF) signal. In *Drosophila*, a single EGF receptor is used repeatedly in over 60 different contexts; control of how the signal is released is therefore critical. Spitz is a transmembrane protein, which must be cleaved by a protease to release the signaling portion for secretion. Firstly, Urban presented biochemical data showing that cleavage of the Spitz protein by Rhomboid occurs in the Golgi apparatus and is followed by glycosylation and secretion of the Spitz signaling portion. Secondly, using biochemical and mutagenic approaches, he demonstrated that Rhomboid is a serine protease: it contains the residues necessary for serine protease catalysis, and is inhibited by known serine protease inhibitors. He also identified a family of seven Rhomboid-like proteins in *Drosophila* and additional Rhomboid-like proteins in species as diverse as the Gram-negative bacteria *Pseudomonas aeruginosa* and *Providencia stuartii*. Finally, he discussed the cleavage site in the Spitz protein: a seven-amino-acid

sequence, ASIASGA in the single-letter amino-acid code, in a transmembrane region of the Spitz protein. This motif is also present in the TGF α and Delta signaling proteins, suggesting they may also be substrates for Rhomboid cleavage. The ASIASGA motif seems to have two functions that allow it to be cleaved by Rhomboid: it produces a kink in the transmembrane α helix and it forms a hydrophilic pocket at the top of the helix allowing water, which is necessary for protease activity, to enter the cleavage site.

Michelle Markstein (University of California, Berkeley, USA) used a computational method (Fly Enhancer [<http://flyenhancer.org>]) to search the genome sequence for clusters of enhancers that are targets of the Dorsal transcription factor. Looking for clusters of enhancer sequences appears to improve the sensitivity of such methods and has allowed the identification of approximately one third of the genes estimated to be directly affected by Dorsal. Besides known targets such as *zen*, *sog* and *brinker*, she found novel targets, including *Phm*, *Ady* and CG12443; these were confirmed by embryonic *in situ* hybridization and expression of *lacZ* under the control of the putative enhancer. Interestingly, it seems that clusters of different enhancer binding sequences may be more diagnostic for the identification of *cis*-control regions than clusters of a single binding site.

A number of groups described research using microarray analysis. Ulrike Gaul (Rockefeller University, New York, USA) presented an analysis of glial cell transcription. Glial cells labeled with green fluorescent protein under the control of the *repo* promoter were chemically dissociated from embryos and sorted by fluorescence-activated cell sorting (FACS). Gene expression in glial and non-glial cell fractions was assessed using an Affymetrix gene array, and 255 strongly expressed genes were identified. CG11910 is expressed only in midline and longitudinal glia; reduction of the transcript level by RNA interference (RNAi) prevents midline glial cells from separating axon tracks in embryonic

commissures. Other examples of new genes found in this screen include molecules affecting axon guidance, cell migration and shape, and axon wrapping. The combination of microarray analysis and RNAi provides a new paradigm for rapid screening.

Amir Orian (Fred Hutchinson Cancer Center, Seattle, USA) has investigated the binding sites of the Myc-Max-Mad (MMM) transcription factor complex. Fusions of *Dam* methylase to these proteins were introduced into transgenic flies, then genomic DNA was digested with a methylation-sensitive restriction enzyme and the fragments were analyzed on a microarray. Interestingly, methylation of genes encoding synaptic-vesicle and mitochondrial proteins was observed, suggesting that the MMM complex may exert previously unknown influences on these processes.

Greg Gibson (North Carolina State University, Raleigh, USA) used long-oligonucleotide arrays to study the inheritance of gene expression. Gene expression was measured in seven strains of *D. melanogaster* and all F1 progeny of crosses between those strains. His data show that approximately 10% of genes are differentially expressed between any two of the strains studied, and that 20% of genes are expressed differently in the F1 compared to the parental strains. It is possible to divide these differences into several classes: some are expected, such as additive, dominant and recessive patterns of inheritance of the expression level; in other cases, the level of gene expression in the F1 is significantly greater or less than can be explained by additive expression of both parental strains.

Many researchers are making use of the expanding *Drosophila* gene collection. Mark Stapleton (Lawrence Berkeley National Laboratory, Berkeley, USA) identified RNA-editing substrates by comparing the high-quality cDNA and genomic sequences. He found 27 adenosine deaminase substrates, the majority of which are ion-channel transcripts. Pavel Tomancak (University of California, Berkeley, USA) presented a comparison of *D. melanogaster* and *D. pseudoobscura* embryonic expression patterns for a number of genes. The vast majority of the 176 genes investigated showed identical expression patterns in the two species. But two genes with different expression patterns were identified. The expression of the midline fasciclin transcript is moved from the neuroectoderm in *D. melanogaster* to the mesoderm in *D. pseudoobscura*. *Ecdysone-inducible gene E2* (described in a poster presented by Amy Beaton, University of California, Berkeley, USA) is expressed in the anterior of early embryos and in the developing foregut by stage 11 in *D. melanogaster*, but in *D. pseudoobscura* it is expressed in the posterior of early embryos and in the developing hindgut by stage 11.

Laura Lee (Massachusetts Institute of Technology, Cambridge, USA) identified seven novel substrates for Pan gu, a protein

kinase required early in the cell cycle during embryogenesis. Her biochemical screen made use of coupled transcription-translation of cDNA clones from the *Drosophila* gene collection to produce [³⁵S]-labeled proteins in a 384-well format. Pools of 24 proteins were then screened in a variety of binding, degradation and enzymatic assays. Examples include screens for Disheveled-binding proteins, microtubule-binding proteins and the Pan gu kinase assay based on band shifts on electrophoretic gels.

One talk highlighted the imprecise art of gene prediction. Marc Hild (University of Heidelberg, Germany) presented a microarray constructed using a less stringent gene-prediction program and a possible 21,396 putative ORFs. Expression data acquired using this array suggests that there are 3,000 more *Drosophila* genes than were predicted in the Release 3 version of the genome. Some of these sequences produce phenotypes in S2 tissue culture cells when inhibited by RNAi. Once the data are made public and analyzed in detail, many of these 'novel' genes will no doubt be found to have exons that overlap those of previous predictions. Other differences may be 'philosophical': for example, should a gene prediction be considered if it has an open reading frame of less than 100 amino acids? It is clear that biological evidence is required to positively identify a gene.

The *D. pseudoobscura* sequence, available from the *Drosophila* Genome Project [<http://hgsc.bcm.tmc.edu/projects/drosophila>], may be the surest way to identify the meaningful sequences of the *D. melanogaster* genome. Richard Gibbs (Baylor College of Medicine, Houston, USA) presented the initial release of the *D. pseudoobscura* sequence. A tBLASTn comparison of the two *Drosophila* genomes identified putative orthologs in *D. pseudoobscura* for 95% of *D. melanogaster* genes. Alignment of the two genomic sequences identified both large features, such as chromosomal inversions, and small ones, such as conserved non-coding regions. A comparative genomic approach using both sequences will improve gene prediction and allow the identification of *cis*-regulatory sequences for the majority of *Drosophila* genes. We can hope that the sequence of *D. pseudoobscura* will be as informative to *Drosophila* research as that of *D. melanogaster*, and many presentations on 'the other *Drosophila*' can be expected at future annual conferences.