

Meeting report

***Drosophila* by the dozen**

Susan E Celniker and Roger A Hoskins

Address: Berkeley *Drosophila* Genome Project, Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.

Correspondence: Susan Celniker. Email: celniker@fruitfly.org

Published: 13 July 2007

Genome Biology 2007, **8**:309 (doi:10.1186/gb-2007-8-7-309)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/7/309>

© 2007 BioMed Central Ltd

A report of the 48th Annual *Drosophila* Research Conference, Philadelphia, USA, 7-11 March 2007.

This year's conference on *Drosophila* research illustrated well the current focus of *Drosophila* genomics on the comprehensive identification of functional elements in the genome sequence, including mRNA transcripts arising from multiple alternative start sites and splice sites, a multiplicity of noncoding transcripts and small RNAs, identification of binding sites for transcription factors, sequence conservation in related species and sequence variation within species. Resources and technologies for genetics and functional genomics are steadily being improved, including the building of collections of transposon insertion mutants and hairpin constructs for RNA interference (RNAi). The conference also highlighted progress in the use of genomic information by many laboratories to study diverse aspects of biology and models of human disease. Here we will review a few highlights of especial interest to readers of *Genome Biology*.

Comparative genomic analysis

The largest new *Drosophila* dataset comes from the draft genomic sequencing of 11 sibling species of *D. melanogaster* with phylogenetic relationships spanning 40-60 million years. Michael Eisen (Lawrence Berkeley National Laboratory, Berkeley, USA) presented a comparative analysis of these new genomic sequences with a focus on the evolution of gene regulation. Whole-genome shotgun sequences and assemblies for *Drosophila simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis* and *D. grimshawi* have been produced by the biotechnology company Agencourt, and the genome centers at Baylor College of Medicine, the J. Craig Venter Institute and Washington University, St Louis. The

latest assemblies, alignments and annotations of these genomes using the *D. melanogaster* Release 4 genome sequence (see the Berkeley *Drosophila* Genome Project website, <http://www.fruitfly.org>) as a reference are available on the AAA (assembly/alignment/annotation) website (<http://rana.lbl.gov/drosophila>). Eisen discussed how the fruitfly genomic sequence in intergenic regions is some 10-fold more highly constrained than in vertebrates with comparable divergence times. The evolution of gene regulation is being approached by identifying potential binding sites for transcription factors in these genomes from published DNase I footprints (see the *Drosophila* DNase I Footprint Database website, <http://www.flyreg.org>) and confirming them by hybridization of chromatin immunoprecipitation (ChIP) products to whole-genome tiling microarrays (<http://bdtncp.lbl.gov/Fly-Net>). Eisen described how binding sites within a DNase I footprint are frequently not conserved, especially between the more distant species. There appear to be gains in transcription-factor-binding sites in *D. melanogaster* compared with the other species, and a deficit of losses along the *melanogaster* lineage. Because of the difficulty in unambiguously determining functional transcription-factor-binding sites, Eisen suggested that robust identification of control regions by comparative sequence analysis would benefit from genomic sequencing of more divergent fly species. New high-throughput sequencing technologies such as the instruments from 454 Life Sciences (<http://www.454.com>) and Solexa (<http://www.illumina.com>) should make this feasible.

In the meantime, cisDECODER (<http://evoprinter.ninds.nih.gov/cisdecoder/index.htm>), a new tool for the computational analysis of cis-regulatory modules described by Thomas Brody (National Institute of Neurological Disorders and Stroke, NIH, Bethesda, USA), should prove useful for the large-scale discovery and characterization of enhancers. This software identifies short conserved sequence blocks

from comparative genomic sequence alignments and parses them into sets of similar potential enhancers shared by genes that are known to be coordinately expressed.

Comparative studies of the sequence data from the 12 sibling species have also provided new insights into the protein-coding capacity of the *Drosophila* genome. Manolis Kellis (Massachusetts Institute of Technology, Cambridge, USA) described the identification of 1,200 new conserved protein-coding exons in *D. melanogaster*, and one of us (S.E.C.) reported the experimental validation of these predictions, which has led to the discovery of hundreds of new protein-coding transcripts. Bill Gelbart (Harvard University, Cambridge, USA) reported that these new gene models annotated by FlyBase will be publicly available as part of release 5.2 of the FlyBase website (<http://flybase.bio.indiana.edu/>). The genes are often interdigitated with genes on the opposite strand, and one of the new genes is the first described case in *Drosophila* of an exon being translated on both strands.

Antonio Bernardo Carvalho (Universidade Federal do Rio de Janeiro, Brazil) discussed Y-linked genes and reviewed how the *D. pseudoobscura* Y chromosome evolved from an X:3L fusion and shares no genes with the Y chromosomes of the other sequenced species. Brian Oliver (National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, USA) described comparative microarray studies showing that, surprisingly, most of the differences in gene expression between male and female adult flies are conserved among the sibling species. It was previously thought that speciation would be accompanied by changes in male gene expression.

Looking to the future, Trudy Mackay (North Carolina State University, Raleigh, USA) presented a proposal for the systematic identification of *Drosophila* genes contributing to quantitative traits. She described a collection of 345 *D. melanogaster* inbred lines that display high variation in many quantitative traits and proposed draft genomic sequencing of 40 of these inbred lines at four times coverage, using 454 Life Sciences technology at an estimated cost of \$2.3 million. Such data would identify most of the sequence variation and could be used to facilitate molecular identification of genes and alleles at many quantitative-trait loci. A white paper on the proposal is to be reviewed by the NIH in the near future. Andrew Clark (Cornell University, Ithaca, USA) pointed out that the new high-throughput sequencing technologies make it feasible to obtain draft-quality sequences of insect genomes at a low cost - around \$40,000 if you already have access to an appropriate machine. He seconded the proposal for genomic sequencing of some more distantly related species, such as the house fly, for improved annotation of both *D. melanogaster* and the mosquito *Aedes aegypti*. Clark also suggested that finishing the draft sequences of the closely related species in the *simulans* group to higher quality will be important for studies of mechanisms of speciation.

Steven Mount (University of Maryland, College Park, USA) presented a comparison of spliceosomal small nuclear RNA (snRNA) genes in the 12 sequenced fly genomes. Candidates for all nine spliceosomal snRNA genes (including those for the U11 and U12 RNAs of the minor spliceosome) were identified. Many display conserved number and synteny, but gene gain and loss was also observed. There was little support for stable snRNA subtypes, which may argue against specialized roles for these variants. Expansion of intron length in U11 and U12 was observed and may be related to the striking loss of U12-type introns in this group of species compared with vertebrates.

Localizing embryonic gene expression

Drosophila is a leading model organism for developmental biology, and the localization of specific mRNAs at different stages of development is of considerable interest. Ben Berman (University of California, Berkeley, USA) presented an update of the Berkeley *Drosophila* Genome Project embryonic RNA *in situ* hybridization project. Images of expression in embryos at multiple stages of development are now available for 6,000 genes (at the Patterns of gene expression in *Drosophila* embryogenesis website, <http://www.fruitfly.org/cgi-bin/ex/insitu.pl>), and web-based tools enable searches of the expression patterns using gene names and controlled vocabularies describing gene ontology and anatomical features. Globally, 46% of *Drosophila* genes show broad or ubiquitous expression during embryonic development, while the patterns of localized expression defy easy classification, with many gene-specific patterns.

Looking at a more restricted set of developmental stages, Eric Lécuyer (University of Toronto, Canada) described a screen for mRNAs localized during early embryogenesis, in which fluorescent *in situ* hybridization was used to analyze mRNAs from over 4,000 genes. An unexpectedly high proportion of mRNAs (70%) have specific subcellular localizations in early embryos, and many novel distribution patterns were identified. Distinct classes of co-localized transcripts are enriched for mRNAs encoding functionally related proteins, suggesting that mRNA localization may control the assembly of diverse protein complexes.

Posttranscriptional regulation of gene expression

Recursive RNA splicing occurs in genes with very large introns and results in the removal of small subfragments of the introns as they are transcribed. In the process, an internal element functions first as a 3' splice site acceptor but restores a 5' splice donor site when joined to the upstream exon. Javier Lopez (Carnegie Mellon University, Pittsburgh, USA) described genome-wide analyses of recursive mRNA splicing. The distribution and conservation of recursive splice sites between *Drosophila* species indicate

roles for this type of splicing in the expression of genes with large introns. Downstream modules consisting of proximal intronic splicing enhancers, a pseudo 5'-splice site and distal splicing silencers are common within 100 nucleotides of a recursive splice sites. This reflects a continuum between non-exonic sites and recursive cassette exons that depends on the presence and relative strength of module components. Interconversion can occur between non-exonic recursive splice sites and recursive cassette exons as a consequence of mutations in the splice site motif, mutations in components of the downstream module, or relocalization of the recursive splice sites to different introns.

Another posttranscriptional modification is the process of RNA editing, which recodes certain mRNA transcripts in the *Drosophila* nervous system and thus contributes to the diversity of proteins produced. Mark Stapleton (Lawrence Berkeley National Laboratory, Berkeley, USA) presented an expressed sequence tag (EST)-based analysis that identified 27 new genes that undergo RNA editing, bringing the total number of identified and validated genes to 55. The newly identified edited mRNAs encode a range of proteins including signaling molecules and ion channels.

Techniques and tools

Tools and resources are being developed to speed up the study of gene function by approaches such as determining patterns of transcript and protein expression and mutant phenotypes. Transposon-mediated insertional mutagenesis remains a central tool in *Drosophila* genetics. Robert Levis (Carnegie Institute, Baltimore, USA) reported on the Gene Disruption Project that aims to create a collection of fly lines in which every *Drosophila* gene is disrupted by insertion of an engineered transposon. A variety of P-element and piggyBac transposable elements have been used to tag over 50% of the genes (see the Gene disruption project website, <http://flypush.imgen.bcm.tmc.edu/pscreen>). Levis described how the Minos transposable element has significantly improved the yield of newly tagged genes in the project and estimated that 90% of genes may be tagged within the next four years. He then described a new Minos element that has been engineered to contain sequences for recombination-mediated cassette exchange. This feature should enable researchers to replace the sequence within an insertion with any other sequence, dramatically increasing the versatility of new fly lines put into the insertion collection.

In an application of insertional mutagenesis, Oren Schuldiner (Stanford University, Stanford, USA) described a mosaic screen designed to identify mutations affecting axon pruning - the process by which the number of neural connections is reduced during development. A piggyBac transposon was engineered to include a splice acceptor site followed by translation stops (a gene trap), which increased its mutagenicity to 25% lethality. Insertions in 1,400 transcription

units were isolated, and a MARCM screen was carried out on these mutants to identify defects in mushroom body development. MARCM (Mosaic Analysis with a Repressible Cell Marker) is a method in which only the mutant cells in a genetic mosaic animal are labeled. For 19% of the lines, defects were observed in various aspects of neural development. For example, mutations with defects in axon pruning were identified in two subunits of the cohesin complex. This screen illustrates the complexity of the *Drosophila* genetic toolkit and the difficulty of producing a single collection of insertion mutants that satisfies all researchers.

RNA interference libraries

Numerous presentations on RNA interference (RNAi) in *Drosophila* highlighted the emergence of independent libraries that are now available for genome-wide RNAi screens in cell culture. These include a collection commercially available from Ambion (<http://www.ambion.com>), described by Steven Suchtya (Ambion, Austin, USA), the *Drosophila* RNAi Screening Center version 2.0 collection (<http://flyrnai.org>), which eliminates the issue of hybridization of double-stranded RNAs (dsRNAs) to non-target genes through perfect repeats, described by Bernard Mathey-Prevot (Harvard Medical School, Boston, USA), and the Heidelberg RNAi Screening Center dsRNA collection (<http://www.dkfz.de/signaling2/rnai/ernai.html>), designed both to optimize RNAi efficiency and avoid off-target effects, described by Thomas Horn (German Cancer Research Center, Heidelberg, Germany). These new libraries, combined with better ways to address some of the caveats inherent in high-throughput RNAi, bode well for the future of functional genomics in cell-based assays.

Two large collections of fly stocks carrying transgenic UAS-hairpin RNAi insertions are now available, one described by Ryu Ueda (National Institute of Genetics, Shizuoka, Japan) and another by Krystyna Keleman (Research Institute of Molecular Pathology, Vienna, Austria). These insertions are used to make inducible loss-of-function phenotypes. The Japanese collection (<http://www.shigen.nig.ac.jp/fly/nigfly>) currently targets about 8,500 genes (13,500 stocks), and the Vienna collection (<http://www.vdrc.at>) targets the complete set of 15,000 annotated genes (22,247 stocks). Initial findings with both collections have been encouraging, and only a small incidence of false positives was reported for the Vienna collection. In addition, Keleman reported that the strength and penetrance of phenotypes observed with the Vienna stocks could be greatly enhanced by coexpressing UAS-*dicer2*. *Dicer2* is required for short interfering RNA (siRNA)-directed mRNA cleavage and facilitates distinct steps in the assembly of the RNA-induced silencing complex (RISC). Therefore, expressing it at the same time and in the same tissue as the dsRNA promotes silencing of gene expression by specific cleaving the homologous mRNA.

Michele Markstein (Harvard Medical School, Boston, USA) presented an elegant approach for insuring reproducible induction levels of UAS-hairpin RNAs in transgenic flies. Hairpin constructs were precisely targeted through the ϕ C31 integration system to a genomic insertion site preselected for low basal activity and high inducibility in the presence of the transcription factor Gal4. Flanking the integration site with *Su(Hw)* insulator sequences achieved even greater and more uniform inducibility in all tissues tested. In addition, the hairpin expression vector contains two repeats of a cassette containing five UAS sites; one of these cassettes is flanked by *lox* sites, allowing stepwise levels of expression after *Cre*-mediated deletion of one of the cassettes *in vivo*, and thus the possibility of multiple phenotypes.

Despite the long period of divergence of human and fly lineages, *Drosophila* provides information useful for understanding human disease. In the final plenary lecture, Eric Rulifson (University of California, San Francisco, USA) described work to establish a fly model for human diabetes. The human endocrine pancreas, with its insulin-producing cells, develops from the developing gut epithelium and so is derived from endoderm, whereas the insulin-producing cells in the fly are a small collection of neurosecretory cells in the brain that derive from embryonic neurectoderm. Despite their origins from different germ layers, the insulin-producing cells in fly and human are similar in form and function and genes and pathways in the regulation of insulin biology are largely conserved. The expression of orthologous genes in the development of these fly and human endocrine cells suggests there is a shared molecular ancestry of the brain and pancreas insulin-producing cell fate. Rulifson concluded that genetic pathways are the unit of conservation in evolution, and that the tissue or germ layer in which they are deployed is secondary. This radical insight has implications for evolutionary biology and for *Drosophila* and other invertebrates as model systems for the study of human disease.

Acknowledgements

We thank Bernard Mathey-Prevot and Javier Lopez for providing details of some talks we were not able to attend ourselves.