Meeting report

Mice and more

Carol J Bult

Address: The Jackson Laboratory, Bar Harbor, ME 04609, USA. E-mail: cjb@informatics.jax.org

Published: 19 September 2003

Genome ***Biology*** 2003, **4:**340

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/10/340

A report on the Mouse Initiatives V meeting 'Genomics of Complex Systems in Biomedical Research', The Jackson Laboratory, Bar Harbor, USA, 30 July-2 August 2003.

The laboratory mouse shared center stage with many of its distant cousins at this year's Mouse Initiatives V meeting. The meeting brought together close to 100 participants to hear about and discuss the state of the art in three broad areas: the impact of genomics on medicine, comparative genomics, and technology.

Comparative genomics

One of the primary themes was the importance of having genome sequences from many diverse organisms for interpreting the human genome. The completion, or near completion, of genome sequences of many eukaryotic organisms has fueled the development of powerful sequence-based strategies in which many different species are compared to identify those regions of the genome - other than protein-coding genes - that are likely to be under selective pressure and thus to be biologically significant. As illustrated by several of the presentations, genome sequences from organisms other than the usual biomedical models provide an evolutionary perspective on genome history and biology that resonates with Theodosius Dobzhansky's famous statement, "Nothing in biology makes sense except in the light of evolution."

Another emphasis was on the emerging technologies that are enabling efficient genome-wide functional analysis in the mouse and other model organisms. The mouse is the best animal model we have for understanding the molecular and genetic basis of biological processes and disease in humans, but there is still much to be done to assemble an experimental toolkit for efficient functional analysis of the mouse genome. A common refrain was how to make the mouse more like the fly, the worm, or yeast with respect to the genetic and genomic strategies that can be applied to understanding the complex connections between genotype and phenotype. These themes are evident in the highlights from the meeting reported in more detail here.

Elliott Margulies (National Human Genome Research Institute, Bethesda, USA) described a comparison of 1.8 megabases (Mb) of genomic sequence data surrounding the *CFTR* gene (cystic fibrosis transmembrane conductance regulator) from 13 species to identify sequences that are highly conserved across great evolutionary distances. He described the use of a multispecies weighted conservation score to identify conserved regions that have a higher than average likelihood of being biologically significant. The method takes into account the different divergence rate of each species (a conserved region in a more divergent species will score higher than one in more closely related species). The so-called multispecies conserved sequences (MCSs) identified by this method are being followed up experimentally to determine their functional significance.

Margulies and colleagues also investigated which combination of organisms most efficiently identified the MCSs that had been determined from a comparison of the *CFTR* region from all 13 species. Taken individually, the mouse had the third best rate of MCS identification, behind the platypus and the hedgehog. When species were grouped together, the quintet of dog, cow, mouse, rat and chicken had a better detection rate than any one species by itself. The trio of hedgehog, rat, and rabbit matched the quintet's sensitivity and specificity of MCS detection. Using more than five species for MCS detection results in only modest gains in sensitivity and specificity. But whether the results from the *CFTR* region hold up for the entire genome is still an open question. The results of the analysis described by Margulies are currently available as a 'Zoo' track from the University of California at Santa Cruz Genome Browser at [http://genome.ucsc.edu].

The sea squirt *Ciona intestinalis* is a particularly attractive subject for comparative genomics studies designed to

explore changes in genome content and organization during vertebrate evolution. *Ciona* is an ascidian, an early chordate that arose approximately 535 million years ago in the early Cambrian period; its 'tadpole' larval stage is considered a modern day stand-in for the ancestral chordate and reveals a close kinship with vertebrates. Dan Rokshar (University of California, Berkeley, USA) reported on progress in sequencing the *Ciona* genome. It is streamlined compared with those of its distant vertebrate cousins: genes that tend to be members of multigene families in vertebrates often have just a single representative in *Ciona*. This makes *Ciona* attractive for dissecting gene-regulation networks related to cell signaling and development. The status of the sequencing effort stands at 8.5-fold genome coverage in shotgun sequencing, 480,000 expressed sequence tags (ESTs), and 5,600 full-length cDNAs. The estimated genome size for *Ciona* is 155 Mb and the net assembled length is currently 125 Mb; to date, 15,800 genes have been annotated in the *Ciona* genome. As reported by Rokshar, 63% (9,900) of the annotated genes have a homolog in protostomes (*Ciona* and the vertebrates are deuterostomes), and thus represent genes that were present before the protostome-deuterostome divergence; 17% (2,600) appear to be chordate-specific because they have no protostome homolog but do have a vertebrate counterpart; 20% (3,400) have no recognizable homolog. These genes may be specific to the ascidian lineage, or they may be too divergent to allow detection by sequence similarity, or they may simply be incorrect gene-prediction models. There is rampant genetic polymorphism in *Ciona*; the average allelic polymorphism rate in an individual organism is 1.5%, a rate 10-15-fold higher than that reported for humans.

New technologies

As an example of the new technologies that are being deployed to study gene function, George Yancopolous (Regeneron Pharmaceuticals Inc., Tarrytown, USA) presented a functional genomics technology called VelociGene, which can produce targeted mutations and transgenesis at a much higher rate than can be achieved with traditional methods. The technology relies on using large-insert clones (bacterial artificial chromosomes, BACs) as targeting vectors to make directed mutations in one or more genes at a time. BAC-based vectors are created with a reporter-selection cassette flanked by oligonucleotides designed to receive a specific gene. This vector is electroporated into a bacterial cell that harbors both another BAC containing the gene of interest and a plasmid that has recombination activities under the control of a transiently inducible promoter. The result of recombination is that the original BAC vector now harbors the gene or genes of interest. This vector can be used to target genes in embryonic stem (ES) cells and to replace the native gene with the reporter via homologous recombination. The method can generate a wide range of genetic alterations in the target genome, including conditional alleles, point mutations and gene swaps (knock-ins), without the need for isogenic targeting vectors or positive-negative selection schemes. The presence of the reporter also allows one to identify, at high resolution, where the gene or genes of interest are expressed in various cells and tissues.

The power of ES-cell technologies in mouse functional genomics was reviewed by Andras Nagy (Mount Sinai Hospital, Samuel Lunenfeld Research Institute, Toronto, Canada), who described both loss-of-function and gain-of-function approaches. He described a strategy of combining ES cells with tetraploid embryos to produce mouse embryos that are derived completely from ES cells. One of the advantages of this method compared with combining diploid embryos to produce chimeric mice is that the embryos can be stored in liquid nitrogen and reconstituted in a matter of weeks to produce a live mouse. Nagy reported on methods developed in his lab that use an integrase from *Streptomyces* phage phiC31 to achieve site-specific genome insertions for transgenesis in ES cells. Germline transmission of the transgene was demonstrated for two ES cell lines expressing the integrase and it appears that integrase expression does not affect normal mouse development. He also discussed ongoing work to develop ES-cell-based methods for gain-of-function screens using Cre-mediated inducible expression of cDNAs. Finally, Nagy described the development of loss-of-function screens for mouse chromosome 5, using mutagenized ES-cell libraries and markers for either positive or negative selection to identify regions of the genome where there has been loss of heterozygosity (the loss of the single functional allele) due to recombination.

Insertional mutagenesis using transposons is another way of introducing mutations into the germline. In *Drosophila*, insertional mutagenesis using endogenous transposons called P elements has been a powerful functional genomics tool for many years. But although transposons such as Tc and mariner are widely distributed in animals, only inactive forms have been identified in vertebrates. Dave Largaespada (University of Minnesota Cancer Center, Minneapolis, USA) reported on the development of insertional mutagenesis for the mouse. One of the transposons being developed is Sleeping Beauty, a Tc1-like transposon reconstituted from inactive elements identified in salmonid fishes. Sleeping Beauty uses a 'cut and paste' mechanism for transposition and therefore requires no host factors for functionality; it has been shown to be active in a wide variety of vertebrate cell types. Sleeping Beauty is a two-component system: it has a transposon and a transposase. The application of the system depends on how the two parts of the system are combined; one can achieve germline transgenesis, somatic cell transgenesis, or both. Largaespada reported that transgenes within the transposon can be expressed, and that chromosomally resident Sleeping Beauty vectors transpose in mouse soma and germline. Most of the mouse genome appears to be accessible to transposition in this way. Sleeping Beauty insertions

in the mouse genome occur within genes approximately 35% of the time. The insertions can occur within introns or exons and in either orientation, and in both 3′ and 5′ regions. Results to date show that one can get an average of two insertions per gamete (up to a maximum of about 10 or 12) and that a large percentage (40-50%) of the local transposition events are closely linked to the donor site. Largaespada outlined several possible research directions and applications of Sleeping Beauty in the mouse, including determination of the mutagenicity of Sleeping Beauty in a phenotype-driven screen, saturation of a genomic region using transposon mutagenesis, and transposon-assisted chromosome engineering.

These highlights represent only some of the exciting developments reported at the conference. Participants went away with new data and new technologies to apply to their favorite system of study and with the sense that there are many new experimental approaches in genetics and molecular biology on the horizon that will further solidify the position of the laboratory mouse as the animal model of choice for biomedical research.