

Meeting report

The lords of the genomes

Mark Stapleton

Address: Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.
E-mail: staple@fruitfly.org

Published: 28 September 2004

Genome Biology 2004, **5**:349

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/10/349>

© 2004 BioMed Central Ltd

A report on The Biology of Genomes meeting, Cold Spring Harbor, USA, 12-16 May 2004.

At this year's annual Cold Spring Harbor Laboratory genome meeting, entitled 'The Biology of Genomes', the focus was progress in the emerging field of high-throughput functional biology. All the major players in the genomics field were in attendance, and the cover of the abstract book depicted them as the lords of the (DNA) ring. Genomics is clearly moving from its initial 'structural' phase (sequencing) into a functional phase. Three broad research directions showed particular promise, and some of the significant progress is summarized here.

Regulatory networks

Large-scale genome projects have contributed greatly to our understanding of the structure of the genome and continue to make important contributions to biology. The resources generated by these projects have spawned innovative attempts towards understanding cellular and molecular processes. In particular, a comprehensive understanding of transcriptional regulatory circuits is now being pursued in many organisms. High-throughput 'promoter bashing' can be accomplished using a combination of comparative sequence analysis and computational algorithms to identify and characterize regulatory sequences. Using 'architectural features' that are evolutionarily conserved between different *Drosophila* species, work presented by Mike Eisen (Lawrence Berkeley National Laboratory, Berkeley, USA) expanded on the previous identification of local clusters of known transcription-factor binding sites in *Drosophila melanogaster*. He examined differences between binding-site clusters that were active or inactive with respect to their ability to recapitulate expression patterns of nearby genes using a simple reporter-gene assay in the embryo. When compared with *Drosophila pseudoobscura*, sequence identity

could not reliably discriminate between active and inactive clusters. But functional clusters of binding sites were always conserved, whereas non-functional clusters were not conserved. Eisen proposed that by identifying more of the architectural features such as clusters of transcription-factor binding sites, more could be learned from comparative analysis of genomes.

In a related talk on the identification of *cis*-acting control regions in the human genome, Rick Myers (Stanford University, USA) presented work on the prediction and experimental testing of promoter regions in the human genome. He used the 5' ends of cDNA clones in the human cDNA collection as an aid for making predictions. Predicted regulatory elements were tested in either an enhancer or a promoter assay using transient transfections in tissue culture lines. Of the 25,790 high-quality predicted promoters, 1,105 have been tested and 92% are positive in one or more of four cultured cell types. Interestingly, 10% of the promoters are bidirectional in nature and the two genes, usually situated 100-600 base-pairs apart, are indeed coregulated. In order to clone all human promoters, Myers' group has used a commercially available system for cloning PCR products, called In-Fusion, along with a bacterial plating scheme that allows for high-throughput processing. This approach was piloted on 30 megabases of sequence from 44 regions designated from the encyclopedia of DNA elements (ENCODE) project. Myers has identified 374 high-quality predicted promoters, 305 alternative promoters, 121 low-quality predicted promoters and negative controls from highly-constrained non-coding, non-promoter regions. Of 1,184 fragments tested, 100% were successfully cloned, most high-quality predicted alternative promoters were positive in the assay as were some of the low-quality predicted promoters, whereas all negative controls were indeed negative in the assay.

Taking a more experimental angle towards understanding regulatory circuits, Mike Snyder (Yale University, New Haven, USA) presented his group's comparison of the

mitogen activated protein (MAP) kinase and cyclic AMP (cAMP) signaling pathways in two yeasts, *Saccharomyces cerevisiae* and *Candida albicans*. Using chromatin immunoprecipitation (ChIP) with three highly conserved transcription factors, Snyder's group investigated whether the downstream signaling targets were similar in the two species. In general, with a criterion set at a BLASTp expect value of $\leq 10^{-10}$ for over 80% of the protein length, the homologous transcription factors interact with homologous targets between 3% and 10% of the time. It is surprising that these homologous morphogenic pathways do not share a greater number of homologous targets, even when the 80% length criterion was eliminated. This may be due to the fact that the two genomes being studied are too evolutionarily divergent (100 million years). Snyder concluded that the two related yeasts use different regulatory circuits consisting of different targets, different combinations of factors, and new factors to regulate an apparently similar morphogenic process. This is probably a result of the different ecological niches of these two yeasts and raises the interesting question of how these factors evolved to form two very different circuits. The answer could be the evolution of either new transcription-factor binding specificities or, more likely, of new binding-site clusters at new target genes. On the basis of this work, Snyder postulated that the differences between human and chimp would more than likely be due to differential gene expression rather than differences in gene content.

Environmental (meta-)genomics

Only 1% of microbes can be grown or cultured in a lab setting. A solution to the problem of sampling this untapped biological diversity is a sequence-based approach - that is, directly to collect otherwise uncultivable microbes from their natural habitats and construct shotgun DNA libraries. Eddy Rubin (Joint Genome Institute, Walnut Creek, USA) presented progress on sampling one such community of microorganisms from a toxic drainage site at an acid mine called Iron Mountain. The interior walls of this mine are coated in a thin biofilm that grows in unbelievably acidic conditions (pH 0.5). The microbes survive in this light-free, otherwise sterile, condition by metabolizing inorganic sulfur into sulfuric acid, leading to the toxicity draining from the mine. In a collaboration with Jill Banfield (University of California, Berkeley, USA), the biofilm was sampled and shotgun sequenced. Most (85%) of the sequence reads were assembled into five distinct microbial genomes. The members of this simple community live symbiotically, and one of the organisms, LeptoIII, is the only one that contains the genes responsible for fixing the community's nitrogen. Efforts are now underway to rid the mine of LeptoIII, with the hope that removing this organism will disrupt the community and remove the source of the toxicity. A soil community, which is far more complex than the acid-mine community, was also shotgun-sequenced. Assembly of the soil sequence data was far too complex to obtain any meaningful information on

individual species. Clusters of orthologous genes, or COGs, were therefore constructed and examined, revealing that the soil had around 3,500 unique clusters, as compared to the acid mine which had roughly 1,800. The COGs were then used to build functional fingerprints from the data in order to better understand the biology of these complex communities.

Craig Venter (The Center for the Advancement of Genomics, Rockville, USA), also sampling from the environment, presented initial results of an ambitious effort to describe global gene space, particularly the diversity of microorganisms in the oceans. He and his coworkers filtered seawater from various locations in the Sargasso Sea, and produced about 2 million shotgun sequence reads. Although the Sargasso Sea is considered to be a 'desert', assembly of the data suggested otherwise, yielding a tremendous diversity with approximately 1.2 million genes identified, most of which have no exact matches to any known sequences. Most notably, sampling sites near each other were very diverse. A most interesting example is the great diversity of the rhodopsin genes, suggesting that this part of the ocean gets most of its energy from the sun. Continuing on in a new global ocean-sampling journey, Venter described some recent data and extrapolated that there might in total be 1-2 billion genes that occupy 'gene space'.

In a related approach, David Relman (Stanford University, USA) sampled the subgingival flora in crevices between human teeth to better understand the possible etiology of periodontitis and gum disease. The abundance of rDNA from the TM7 division of bacteria in the subgingival crevice was significantly increased in patients who had been diagnosed with mild periodontitis. Whether TM7 bacteria are the causative agent in the early stages of the disease, or conversely whether the disease allows the bacteria to thrive, is not known. Using microarrays to compile an inventory of the genes in the subgingival crevice community can, however, perhaps serve to provide useful diagnostic and prognostic signatures of dental health.

Chemical genomics

In a plenary talk introducing a relatively new area with great potential, Stuart Schreiber (Harvard University, Cambridge, USA) described the use of diversity-oriented synthetic chemistry as a starting point for chemical genomics. Through two or three simple chemical reactions, diversity is achieved from the skeletal and stereochemical properties of a compound that after initial screening can then be added to using more classical combinatorial chemistry. Schreiber stressed that this type of chemistry uniquely anticipates the need for follow-up chemistry and the need for 'open-source' biological feedback to keep the chemists engaged. This, of course, necessitates well-designed screening platforms, which are the key to taking full advantage of chemical genomics. An example of an investigator-initiated phenotypic screen was

given that resulted in the identification of a compound that disrupts mitosis, which then led on to a better understanding of the spindle-pole apparatus and how kinesin is involved. In order to store and present such information to the public, there are now plans to make small-molecule and assay-data repositories analogous to PubMed and GenBank.

Overall, the work presented at the meeting demonstrated the diversity of functional studies that are being enabled by the availability of genome sequences. The genomics field is certainly moving away from generating sequence data and is now trying to understand how genomes function. We can look forward to plenty of new developments over the coming months and to an exciting meeting next year.

comment

reviews

reports

deposited research

refereed research

interactions

information