Meeting report

The changing face of genomics

Manolis Kellis

Address: The Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02138.
MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139.
E-mail: manoli@mit.edu

A report on the 5th annual AGBT / AMS meeting on Advances in Genome Biology & Technology
Marco Island, Florida, USA, February 4-7, 2004.

The annual meeting on Advances in Genome Biology and Technology was very different this year. Only a handful of talks covered the latest large-scale sequencing projects and the next species to be sequenced. The meeting took for granted that we can sequence, assemble, and align complete genomes, achievements that only a few years ago seemed daunting, if not unthinkable. The focus instead shifted towards the new challenges in genomics, particularly in the areas of gene regulation, cell dynamics, and genome evolution.

(1) Regulation: ***Systematic discovery of all regulatory elements***

Given the primary sequence of a species, a major goal of genomics is to understand the regulatory mechanisms and control circuitry of the cell. Towards this goal, Rick Young (MIT / Whitehead Institute) presented the completion of the protein-DNA interaction map in yeast. Using Chromatin Immuno-Precipitation (ChIP) technology in combination with microarray chips containing all intergenic regions, his group has studied the genome-wide targets of all 200-some transcriptional regulators in yeast under multiple environmental conditions. They searched for the sequence-specificity of these regulators using numerous motif discovery tools, together with evolutionary sequence conservation, and protein structure information. The resulting regulatory map revealed general principles of regulation in yeast, including the organizational architectures of promoter regions (single motif, multiple sites, multiple regulators, factor combinations), and the different types of regulatory response to environmental changes (off/on, invariant, expanded, altered). The group additionally studied 20 chromatin regulators that do not directly recognize DNA sequences, but instead rely on their association with transcription factor partners for binding, and can keep a record of transcription. The Young Lab is now moving this technology into studying transcriptional regulator binding in the human genome, with applications to understanding diseases from diabetes to cancer.

(2) Networks: ***From the 'one regulator, one motif' model to enhancer architectures***

Michael Levine (UC Berkeley Biology Department) presented work towards understanding the cis-regulatory circuitry of promoter regions in the fly. He proposed that enhancer complexity might be a better measure of organismal complexity than overall gene count, and noted that the majority of enhancer elements act cooperatively in higher eukaryotes –autonomously-acting elements would in fact be the exceptions. His group studied the promoter architecture of developmental genes responding to different levels of the activation gradient of the developmental protein Dorsal in Drosophila. Five activation levels were detected, created by the combinatorial action of a three-response-level activator and a repressor. They searched for conserved sequence elements in the promoter regions of genes belonging to the same activation level, and they discovered a common grammar in the organization of three basic enhancer elements. Searching for a similar regulatory grammar in the mosquito genome revealed ten genes with similar clusters, two of which contain the same architecture, despite 230 million years of divergence.

(3) Development: ***Systematic identification of vertebrate developmental genes***
Both Rick Young and Michael Levine benefited from classical studies of gene function in yeast and known developmental genes in the fly that have set the foundations for future work. In that spirit, Nancy Hopkins (MIT Biology Department) presented her program to systematically identify all developmental genes in the Zebrafish. Zebrafish is an appealing model for studying early vertebrate development, due to its transparent body and short time, a mere four days, between fertilization and free-swimming larvae. However, since time to adult takes another four months, following multiple generations can be prohibitively slow. To face this challenge, Hopkins and colleagues have used insertional mutagenesis to create mosaic parents whose germ cells contain each a different mutation. Following the fish with developmental defects and classifying each mutation has allowed her to screen 32000 founder fish, identify 550 mutants and 390 loci, 298 of which have human homologues. As many as 20% of these genes have no previously known biochemical function, providing a great starting point for experimentation and new biological discoveries. Additionally, the systematic approach allows one to estimate the total number of developmental genes, which Nancy Hopkins sets at 1600, 25% of which have already been isolated. However, the cost of identifying additional genes increases as the study approaches saturation, and her group is not planning to pursue the systematic discovery phase of the work. They are currently working on understanding the genes identified and have revealed important new insights for genes involved in kidney, jaw, liver, and myeloid cell formation.

(4) Proteomics: ***Dynamic behavior of proteins during development***
Beyond the identification of genes involved in a developmental process lies the major challenge of understanding their dynamic patterns of behavior during development. Josh LaBaer (Harvard Medical School) presented his lab's proteomics work enabling such a pursuit, by developing a Protein Expression Clone Repository that contains full protein sequences for every gene in a number of model organisms, including yeast, bacteria, and human. These are inserted within master clone vectors, and can be easily transferred to specialized clones for expression, GFP tagging for localization, two-hybrid assays for determining protein interactions, or MS/MALDI-specific clones for detecting protein modification states. The system architecture is designed to be flexible, modular, reliable, comprehensive, and catalogued. Mark Vidal (Harvard Medical School) described a similar system for understanding the worm proteome. His lab is now mapping the localization of all 19,000 worm genes across development, building 'chronograms' of gene expression from the head to the tail. His group is then clustering patterns of protein localization across space and time, thus constructing the dynamic aspect of the protein interaction network of the worm. They have currently completed 10% of the interactome matrix, and moving towards completeness, with the goal to understand not only the components, but also emerging protein behaviors during complex organismal tasks.

(5) Evolution: ***Evolution of regulatory mechanisms across related species***
The protein interaction network of a species provides the foundation for understanding the organism's responses to environmental changes and developmental signals. Across evolutionary time, these responses change and the regulatory circuits shift towards new responses. Lisa Stubbs (Lawrence Livermore National Laboratory) presented work towards understanding the evolution of such regulatory mechanisms across species. Her group studied the KRAB-ZNF family of chromatin-interacting zinc-finger transcriptional regulators. These arose 400 million years ago, after divergence from fish, and have recently undergone lineage-specific expansions in the human and mouse lineages via tandem duplications and deletions. Comparative genomics tools typically shy away from analysis of such expanding gene families, since orthologous pairs are hard to determine.

In absence of orthology information, Stubbs constructed multiple alignments of all paralogous gene copies within each species in turn, rather than across species. This clever methodology - although atypical - yielded biologically meaningful intergenic sequence elements that are highly conserved across paralogs, and indeed act as enhancers based on reporter assays. It additionally allowed the identification of species-specific elements since divergence of the two lineages. The group is now working out the protein-level differences between paralogs, with the goal to understand how structural changes affect function, in particular with respect to tissue-specific regulation and parental imprinting.

(6) Diversity: ***Identifying regions of recent selection in the human genome***

Differences between more closely related species can be much more subtle than the protein family expansions observed between human and mouse. For example, divergence between human and chimp shows as few as 12 changes every 1000 nucleotides, which makes signal discovery a real challenge. Mike Zody (MIT / Broad Institute) presented a very interesting way to use such a close relative. Assuming that neutral divergence between species and neutral diversity within species are driven by the same underlying mutational mechanisms, Zody and colleagues used human-chimp divergence information to model the background mutation rate for each region of the genome. Using this information, they could distinguish regions of low within-human diversity that were due to selective sweeps, rather than simply a lower mutation rate. These regions of low diversity were confirmed by genotyping nucleotide polymorphisms in humans of different ethnicity and building allele frequency profiles. FOXP2, a well-known gene involved in language development, showed only a moderate signal, whereas some regions that are devoid of annotated genes showed very strong evidence of selection, raising new questions about what genetic elements may be under selection in the human genome.

(7) Anthropology: ***The evolution of language in human populations***

With the aim to understand human variation and how genetic differences may relate to phenotypic differences, Joanna Mountain (Stanford University) presented her work relating language similarities to population polymorphisms. In particular, she has studied the genetic relationships of two geographically isolated populations in Africa, both of which speak with clicking languages (Khoisan). Using a multitude of metrics across various loci and populations, she found that the genetic data clearly supports two distinct groups for the two languages. To explain the results, Mountain brought forward three possibilities: that the two languages arose independently, that they traveled across populations, or that instead the ancestral language was click-based and all intervening groups lost the click sounds. One advantage to such click languages might be the ability to communicate while hunting without alerting the prey. The use of modern genomics tools in the population genetics of language evolution illustrates how diverse the field has become, and in fact, the maturity of genomics as a science applicable well beyond the boundaries of the modern laboratory.

Overall, the meeting showcased a wide range of innovative talks, combining ground-breaking technological inventions with important scientific applications. The attendance was unusually low this year, especially on the industry side, witnessing the tight economic situation in the US and internationally. At the same time, the changing focus of the meeting is evidence of a maturing field. Genomics has mastered its initial challenges, and is now extending its arms to embrace a growing number of fundamental questions in the study of life.