

Meeting report

The pleiotropic dividends of genomics

Ann Schlesinger

Address: Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142, USA. E-mail: schlesinger@wi.mit.edu

Published: 12 December 2001

Genome Biology 2001, **3(1)**:reports4003.1–4003.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/3/1/reports/4003>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report on the 19th Whitehead Institute Symposium, Cambridge, USA, 14-16 October 2001.

Genomics is a relatively new field that has arisen from the sequencing of genomes and advances in microarray technology. Until recently, the methods for obtaining and analyzing genome-scale amounts of information were the main research concerns. Talks at the recent Whitehead Institute Symposium on Genomic Information showed that genomics is now emerging from its embryonic stage and beginning to yield significant results in many areas of research.

The sessions covered disparate fields reflecting the diverse disciplines currently benefiting from genomic information. In addition, the talks showed that the increase in breadth of approach, from the single gene to the genomic level, concomitantly expanded the scope of the results. This report summarizes a subset of the Symposium talks, illustrating the powerful contributions of genomics to the topics of human disease, protein evolution, and the origins of species.

Genetics and infectious diseases

Two talks on research into diseases emphasized the benefits of the large database of single-nucleotide polymorphisms (SNPs) being developed by the Human Genome Consortium. Eric Lander (Massachusetts Institute of Technology/Whitehead Institute, Cambridge, USA), in an overview of the progress made at the Whitehead Center for Genomic Research, set the number of known SNPs at over 2 million. Lander's group is using the SNP database to test the 'common disease-common variant' hypothesis which contends that frequent genetic variations, rather than rare mutations, are likely to cause the more prevalent diseases. By comparing the frequency of specific SNPs in a patient cohort to the frequency in a disease-free control group, correlations between a disease and a set of SNPs are determined. The disease-causing gene may be directly identified if

it contains an internal polymorphism. When co-segregating SNPs merely flank the gene, however, linkage-disequilibrium mapping (that is, plotting how often the SNPs and disease are linked) can be used to narrow the region of study and identify the gene of interest. Lander described how this approach was recently used to establish a link between susceptibility to Crohn's disease and a region of chromosome 5q31 that contains a cluster of cytokine genes.

Stephen O'Brien of Cornell University (Ithaca, USA) is also using SNPs, in this case to look for correlations between genomic sequence and the progression of disease in people infected with human immunodeficiency virus (HIV). Focusing on sero-positive individuals who did not develop acquired immune deficiency syndrome (AIDS), his laboratory has previously found a deletion in the *CCR5* gene (encoding a chemokine receptor) that protects individuals from the progression to AIDS after HIV exposure. Of the people homozygous for the *CCR5* deletion, only one out of the 169 exposed to HIV developed AIDS. Heterozygotes for the deletion had an infection rate approximately equal to patients without the mutation, but there was a delay of three years on average in the progression of HIV among these people. Using the same approach, the O'Brien group identified a SNP that correlated with an increased rate of HIV progression. This SNP fell within an intron of the cytokine-encoding gene *RANTES* that confers protection against AIDS by binding to *CCR5*. This SNP down-regulated *RANTES* expression, thereby decreasing signaling through *CCR5* and increasing the rate of AIDS progression.

Evolution and development

David Eisenberg (University of California, Los Angeles, USA) presented his group's work on protein function and evolution that blends database mining with expression data. To study the functional relationship between proteins, Eisenberg's laboratory built phylogenetic profiles, using sequence information from 17 genomes. The presence or absence of genes across genomes was determined, and genes

whose distribution across genomes was similar were deemed likely to be functionally related. These results were combined with microarray expression analysis that suggested functional relatedness through clustering of common transcription patterns. The output of these analyses is a vast network of interactions, all established at this stage through computational analysis of phylogenetic and expression data. The correlation between these inferred protein functional interactions and those in the literature derived from more traditional methods, such as immunoprecipitation, is not yet satisfactory. Eisenberg believes that further refinement of this approach will allow analysis of protein functional interactions on a genomic level. He is also continuing to use the data to study protein evolution, looking at protein domains that in some genomes are encoded together by a single gene but are encoded by separate genes in other genomes. By using this 'Rosetta stone' approach, he hopes to learn about the evolution of functional domains.

Origins of species

Svante Pääbo (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany) has used microarrays and expression analysis in a comparative approach to understanding human origins. To characterize the differences between humans and chimpanzees, Pääbo delved into a new field of inquiry that he labeled 'evolutionary transcriptomics'. Using DNA arrays, transcripts from various tissues were compared between humans and chimpanzees. For most tissues, such as liver, the variation between individuals was only slightly smaller or equal to the differences in transcriptional regulation between species. In some cases, human samples were more similar to chimpanzee samples than to other human ones. The striking exception was the expression patterns in brain tissue. Here, humans differed significantly from chimpanzees, reflecting the large differences between the two species in this organ, and presumably accounting for human cognitive and language skills.

Genomic information, applied to cancer and AIDS, has identified genetic components contributing to the development and progression of these complex diseases. In a similar manner, phylogenetic analysis using genomic sequences should help define the maze-like network of protein interactions and reveal the lineages of protein evolution. On a more subtle level, expression data have revealed distinguishing traits between humans and chimpanzees, two species whose genomes are 99% identical. Taken together, these results presented at the Whitehead Institute Symposium clearly show the significant contributions genomic information is beginning to make to science.