

Meeting report

## Comparative genomics comes of age Leonard Lipovich

Address: Department of Genome Sciences, University of Washington, Seattle, WA 98195-7730, USA. E-mail: LL@u.washington.edu

Published: 15 July 2002

Genome **Biology** 2002, **3**(8):reports4024.1-4024.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/8/reports/4024>

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

---

A report on the 2002 annual Cold Spring Harbor Laboratory meeting on Genome Sequencing and Biology, Cold Spring Harbor, NY, USA, 7-11 May 2002.

---

A publicly available draft sequence of the mouse genome at 6.3X coverage (each base sequenced an average of 6.3 times) - announced by Robert Waterston (Washington University, St Louis, USA) in the opening session of this meeting - can now be compared with the available human draft sequence. But what can, and can't, the mouse tell us about being human? Has mammalian comparative genomics advanced enough to enable us to understand why humans and chimpanzees look and behave so differently despite an estimated 98.8% genomic DNA sequence identity? And are mammalian genes more complex than they were thought to be in the heady early days of counting gene numbers, when only crude automated annotations and meager cDNA collections were available? Most of the material at the 2002 annual Cold Spring Harbor meeting that was not presented in some form in 2000 and 2001 was relevant to these three fundamental questions.

### Human versus mouse: what is conserved?

Mike Kamal (Whitehead Institute and Massachusetts Institute of Technology, Cambridge, USA) was the first of many speakers to emphasize the surprisingly high extent of non-coding sequence conservation between human and mouse. Kamal revealed that only 50% of conserved elements in the total genomic sequence (exons and introns) of orthologous genes correspond to exons. So, what are the putative non-exonic conserved sequences? One possible answer was suggested in a poster presented by Emmanouil Dermitzakis (University of Geneva, Switzerland). As detailed by Dermitzakis, 62% of sequence blocks on human chromosome 21 that are conserved in the mouse are predicted to be non-exonic by existing annotations. But many of them correspond to expressed sequence tags and long open reading frames, and they therefore probably do in fact represent

novel exons of known, and novel, genes. The utility of human-mouse comparisons is limited, however; in fact, of the 1,822 exons on human chromosome 21, only 68% have equivalents in the mouse (poster presented by Katsuhiko Murakami, RIKEN Genomic Sciences Center, Yokohama City, Japan).

Eric Green (National Human Genome Research Institute, Bethesda, USA) helped expand the horizons of comparative genomics at this meeting beyond human-mouse comparisons. He has analyzed sequences syntenic to portions of human chromosome 7, which were obtained from multiple vertebrates in a targeted sequencing project. Intronic sequence conservation was absent from mammal-bird and mammal-fish pairs, and among mammals the degree of intronic sequence conservation varied from gene to gene.

### Towards a sequence-level basis for species-specific phenotypes

A comparative analysis of a gene family rapidly evolving in great ape lineages was presented by Evan Eichler (Case Western Reserve University, Cleveland, USA). Eichler discussed the LCR16A duplications on human chromosome 16. The duplicated regions contain multiple copies of a novel gene, *MORPHEUS*, which is absent from the mouse and has undergone amplification and apparent positive selection in apes. Some of the lineage-specific LCR16A insertions in human and chimpanzee chromosomes disrupt gene-rich regions, and ongoing gain and loss of the duplicated copies is taking place in human populations. According to Eichler, LCR16A may exemplify the remodeling of an entire chromosome in a manner unique to the human lineage.

Recent lineage-specific genome structure modification in primates is, of course, not limited to a single gene family on a single chromosome. Kelly Frazer (Perlegen, Mountain View, USA) designed a tiled set of long-PCR amplicons covering the entire available human chromosome 21 sequence. She then amplified chimpanzee genomic DNA with human

primers and concentrated on those amplicons where a product size difference, suggesting an insertion or deletion (an 'indel'), distinguished the human and chimpanzee PCR products. Of the 57 indels, 20 were within or near genes. Some of the 20 resulted in gene structure differences between human and chimpanzee, such as the species-specific deletion of an entire exon of a gene.

The International Chimpanzee Genome Sequencing Consortium (poster presented by Hidemi Watanabe, RIKEN Genomic Sciences Center, Yokohama City, Japan) has compared the sequences of bacterial artificial chromosome (BAC) ends from chimpanzee and human, and found that a sizeable proportion of genomic sequences, both from autosomes and from the Y chromosome, differ by as much as 5% between the two species. These regions may be candidates for having experienced accelerated sequence evolution after the human-chimpanzee divergence took place.

As different as humans may be from chimpanzees in some parts of the genome, humans may be even more different from other humans. The difference between a reference human genome and a somatic-cell cancer genome, as defined by the proportion of BACs from a cancer cell line that do not hybridize to BACs from a reference library, approaches 10% (poster presented by Shaying Zhao, The Institute for Genomic Research, Rockville, USA). In the closing session, Vivian Cheung (University of Pennsylvania, Philadelphia, USA) discussed intraspecific transcriptome differences. Cheung probed human cDNA microarrays with cDNA from different individuals, verifying expression-level differences between individuals by RT-PCR. Several genes, including major histocompatibility complex HLA genes and those encoding cytochromes, consistently had very high variation of expression level between individuals, suggesting that heightened intraspecific expression level variability is an intrinsic property of some genes.

### **Antisense and imprinting**

With more finished genomic sequences and more human and murine cDNA clones available than ever before, it is time to re-examine earlier presumptions regarding the mammalian gene count and gene-structure complexity. Over 60,000 nonredundant cDNAs have been reported in the mouse. More than 5,000 of them participate in endogenous *cis*-antisense pairs, according to results from the RIKEN Genomic Sciences Center (Yokohama City, Japan; poster presented by Yasushi Okazaki). Transcription from opposite strands in complex genomic regions is now taken into account during probe selection for cDNA microarray design by Affymetrix (Santa Clara, USA; poster presented by Simon Cawley). In the meantime, the 'human chromosome 7 workgroup' at the Hospital for Sick Children, Toronto, Canada (poster presented by Kazuhiko Nakabayashi) is exploring the complexity of known imprinted regions. They reported the

identification of two novel imprinted genes - one of which is a noncoding antisense transcript - and further intricacies of imprinted loci, such as isoform specificity and epigenetic heterogeneity of imprinting.

### **Glimpsing the not-so-postgenomic future**

In a keynote address, Svante Pääbo (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany) emphasized the relevance of comparing the human and chimpanzee genomes to understanding the basis for medically relevant human-specific phenotypes, ranging from speech and its disruptions to malaria susceptibility and the high incidence of cancer. Pääbo described work that has shown that the human brain is a hotspot for human-chimpanzee gene expression differences and reported on a comparative atlas of great ape gene expression in six areas of the brain.

A second keynote speaker, Richard Gibbs (Baylor College of Medicine, Houston, USA) suggested that the production capacity provided by the major genome centers has better uses than sequencing obscure model organisms, given how pitifully little is known about the genomic basis of human disease. Gibbs outlined how the identification of all human Mendelian disease traits could be completed in one year by large-scale resequencing of candidate genes in the small families with large linked regions that account for the majority of Mendelian diseases for which the causal gene is unknown. He further recommended that the genome centers make inroads into somatic-cell genomics: "sequencing a brain" would be useful, given the popularity of organ-specific transcriptomics, and the mutational theory of aging could finally be tested.

If Gibbs' inspiring call to action is answered, it could mean only good news for the numerous small laboratories that, because of low experimental throughput, are still unable to derive practical benefits from genomic sequence in their thorny positional-cloning projects and oncogenomic endeavors. Fortunately, the keynote speakers' suggestions at the Cold Spring Harbor Laboratory annual genome meetings seem to be quite effective in mobilizing researchers to bridge the gap between plan and reality.