

Meeting report

With the finished human genome in hand, what next?

Julio Collado-Vides*, Arturo Medrano-Soto* and María Teresa Tusié-Luna†

Addresses: *Program of Computational Genomics, Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México, Ave. Universidad s/n, Cuernavaca, Morelos 62100, Mexico. †Unidad de Biología Molecular y Medicina Genómica Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Instituto de Investigaciones Biomédicas de la Universidad Nacional Autónoma de México, Vasco de Quiroga #15, Tlalpan 14000, Mexico City, Mexico.

Correspondence: Julio Collado-Vides. E-mail: collado@cifn.unam.mx

Published: 27 June 2003

Genome Biology 2003, **4**:328

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/7/328>

© 2003 BioMed Central Ltd

A report on the eighth annual meeting of the Human Genome Organization (HGM2003), Cancun, Mexico, 27-30 April 2003.

The eighth annual Human Genome Organization (HUGO) meeting was a very special one, starting as it did two days after the 50th anniversary of the discovery of the structure of DNA by James Watson and Francis Crick and soon after the announcement of the completion of the human genome sequence to 99% accuracy. The meeting covered a vast range of topics, including medical genomics, pharmacogenomics, stem-cell genomics, comparative genomics, genome variation and diversity, model organisms, functional annotation of genomes, proteomics, transcriptomics, new technologies, bioinformatics, and ethical issues. Given the scope and the number of parallel sessions, it is possible to give only a partial illustration here of the ideas and results presented.

A special session was devoted to celebrating the history of the human genome and thinking about the future. Victor McKusick (Johns Hopkins University School of Medicine, Baltimore, USA) reminded us that it was also the 15th anniversary of HUGO and the 30th of the human genome mapping initiatives. 'Genome', as we learned, is a word coined in 1920 by Hans Winkler in Germany, putting together 'gene' with 'chromosome'. The precise number of human chromosomes - 46, not 48 as was previously thought - was established in a meeting in Copenhagen as recently as 1956. The human genome is giving the foundation for a neo-Vesalian basis for medicine, beyond the conventional modern medicine that started with the first complete textbook of human anatomy by Andreas Vesalius in 1543.

The sequencing of the human genome was an effort of 16 laboratories in 6 different countries. Francis Collins (National Human Genome Research Institute, National Institutes of Health (NIH), Bethesda, USA) acknowledged the people present at the meeting who had been involved in the sequencing, and then summarized future goals in genomics and its potential contributions to biology, health and society. Biology will benefit from genomics through the generation of haplotype maps (which show correlations between nearby polymorphisms), the sequencing of many more genomes, new expression technologies, the identification of all functional sequence elements, and the development of computational models of the cell. Reducing the cost of sequencing a mammalian genome to a few thousand dollars seems attainable. The potential benefits of genomics to human health that Collins put forth include no less than the identification of the genetic and environmental factors involved in all diseases, the development of 'sentinel' systems for early detection of disease, and the ability to perform high-throughput screens using specific small-molecule agonists and antagonists for every human protein. At the level of society, genomics raises ethical questions about privacy against genetic discrimination and the need for precise definitions for regulating reproductive cloning, and it also gives us a new understanding of the genetic differences between groups of human beings in history. The lesson that Collins drew from this summary was that if you want to get things done you should make big plans, not little ones.

Medical genomics

It was clear from the meeting that genomics has the potential to revolutionize medical knowledge and, eventually, medical practice, particularly through the identification of

genes conferring susceptibility to common diseases. We learned about the identification of novel gene products linked to common disorders and how they relate to biological processes leading to disease. In the case of diabetes, the calpain family of calcium-activated cysteine proteases have been found to be involved in a variety of cellular functions including regulation of intracellular signaling and apoptosis (Graeme Bell, University of Chicago, USA). For systemic lupus erythematosus (SLE), an intronic polymorphism in the immunoreceptor *PDCD1* was found to be associated with the disease (Ludmila Prokunina, Uppsala University, Sweden). This polymorphism alters the binding site of the transcription factor *Runx-1*, which could lead to aberrant regulation of the *PDCD1* gene, contributing to the deregulated self-tolerance characteristic of SLE.

Stephen W. Scherer (Hospital for Sick Children, Toronto, Canada) described the integration of functional and structural maps of human chromosome 7, leading to the discovery of genes responsible for diseases such as holoprosencephaly, distal tubular renal acidosis, type II citrullinemia, hereditary papillary renal carcinoma, and more. Holoprosencephaly is a genetically heterogeneous disorder causing brain and craniofacial defects; papillary renal carcinoma is a dominant hereditary form of renal carcinoma characterized by multiple bilateral papillary renal tumors; distal tubular renal acidosis is a recessive disorder causing kidney failure; and type II citrullinemia is associated with a decreased activity of the enzyme argininosuccinate synthetase in liver, which can cause disorientation or a coma.

The annotation of genes, including their splicing patterns, and the precise position of their promoters, has been made more successful by isolating full-length capped cDNAs, as explained by Sumio Sugano (Human Genome Center, University of Tokyo, Japan). By taking advantage of the common origin of most of the population living in Quebec, which means that people with the same genetic disease are likely, through shared ancestry, to carry the same allele, Thomas Hudson (McGill University, Toronto, Canada) is trying to identify haplotype blocks that are overrepresented in individuals with multifactorial diseases such as asthma and coronary heart disease. Although some large haplotype blocks have been found around alleles of recessive disease genes, these were not found for the multifactorial diseases. Gerardo Jiménez-Sánchez (Institute of Genetic Medicine, Baltimore, USA) described current progress on the annotation of the 'Human Disease Genome' and presented a list of 900 genes obtained so far.

Zhu Chen (Chinese National Human Genome Center, Shanghai, China) presented recent advances in parasite genomics. By sequencing expressed sequence tags and comparing them with genes from other species, he showed that the trematode parasite *Schistosoma japonicum* encodes receptors for signals from hosts such as insulin, progesterone, and others; this 'molecular mimicry' is likely to be a strategy used to escape the host immune system. This approach allowed the identification of possible targets for new drugs against schistosomiasis.

Functional genomics and bioinformatics

Mike Tyers (University of Toronto, Canada) illustrated the complexities involved in the assignment of functions to genes, especially when genetic perturbations cause weakly discernable phenotypes. Cell-size homeostasis in yeast gives one example of this problem - the genes involved are hard to find because cell-size defects often cause little or no growth defect. His group has applied multiple functional genomic approaches (proteome, transcriptome, and roboticized synthetic genetic arrays) to the study of this problem in yeast. He has found that deletions in 15 ribosome-biogenesis factors give mutant cells that are abnormally small. Many of the genes encoding these factors are regulated by a transcription factor (*Sfp1p*), which, when deleted, also gives very small cells. These approaches reveal a plausible link between ribosome-biogenesis factors and the cell-cycle machinery and also identify new genes that might regulate the triggering of cell division in late G1 phase.

The genetic and molecular basis for phenotypic variation is still largely unknown. In a talk by Leif Andersson (Uppsala University, Sweden) we heard about what was claimed to be the first identification of a single point mutation underlying a quantitative trait locus (QTL). By comparing wild boars and the Large White breed of domestic pig, Andersson and colleagues showed that a QTL that increases muscle growth, decreases fat deposition and increases the weight of the heart in domestic pigs is attributable to a regulatory mutation that increases postnatal expression of insulin-like growth factor 2 in skeletal muscle.

A very good overview of chromatin structure and its role in gene regulation was given by Carl Wu (National Cancer Institute, NIH, Bethesda, USA). He introduced us to a landscape of short linker regions between nucleosomes - of 20 to 60 base pairs - which are dynamically made available to the transcription machinery. Four different classes of ATP-dependent chromatin-remodeling machines catalyze the movement of nucleosomes. During chromatin remodeling, histone octamers remain on the DNA; their sliding is thus required in the regulated activation or repression of many genes.

Methylation of CpG islands - which are found upstream of many genes - is regulated by the methyl-CpG-binding domain (MBD) family of transcriptional regulators. In Rett syndrome, a disease unique to females, brain retardation has been found to correlate with lack of expression in the brain of *MECP2* (a member of the MBD family), even when the protein is expressed in other tissues. Tim Roloff (Max Planck

Institute for Molecular Genetics, Berlin, Germany) has found new members of the MBD family and proposed that, in tissues other than the brain, these proteins must compensate this lack of expression of MECP2 in Rett syndrome patients.

Alicia González (Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico) described detailed studies of co-regulated divergent promoters in yeast, a frequent type of organization in this organism. The glutamate synthase gene *GLT1* is expressed from a promoter that is divergent from that of *UGA3*, which encodes a transcriptional activator of gamma aminobutyric acid (GABA)-dependent genes, but *GLT1* is expressed at 10-fold higher levels than *UGA3*. This difference is partly due to a binding site for the activator ARS-binding factor (Abf1p); binding of Abf1p causes a peculiar chromatin organization that stimulates *GLT1* and represses *UGA3*. The directional effect on the expression of these two promoters is unique to this activator, even though both genes are subject to control by three more factors.

Alberto Kornblihtt (University of Buenos Aires, Argentina) described progress in understanding how alternative mRNA splicing is modified by alterations in the rate of transcriptional elongation by RNA polymerase II. Kornblihtt has previously shown that transcription and splicing are coupled; differential splicing of exons is affected by the elongation rate, which in turn depends on the transcriptional promoter. Some viral transcriptional activators, such as VP16, increase exon skipping, whereas others, such as large T antigen, decrease it. Kornblihtt explained that fast elongation with no pausing will naturally increase the likelihood that exons are skipped, but only if there are regulatory sites for alternative splicing in the gene; and a single mutation can lead to the inclusion of exons, independent of the rate of elongation. In a human cell line, a mutated RNA polymerase II was shown to lead to increased inclusion of exons.

Minoru Ko (National Institute on Aging, NIH, Bethesda, USA) has made a comprehensive catalog of genes expressed in mouse embryonic and adult stem cells as well as in unfertilized eggs and blastocysts [<http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html>]. This cDNA collection provides alternatively spliced forms for more than 3,400 genes that have previously been described, and at least 1,450 novel genes were found. A microarray made from about 22,000 of these sequences has been constructed, which may give insights into the fundamental biology of stem cells.

The problem of finding syntax and query languages with which to manage the various biological databases has been tackled by the community with the development of 'Gene Ontology' (GO), arising out of Monica Riley's functional classification of genes in *Escherichia coli*. Michael Ashburner (Cambridge University, UK) described how GO is gradually expanding to many other areas of biological interest, such as

anatomy, pathology, and phenotypic data. And John Matese (Stanford University, USA) presented a unified computational genomic resource, 'Source', that facilitates searching, analyzing and interpreting data of microarray experiments in yeast [<http://source.stanford.edu/>].

Rosa María Gutiérrez (UNAM, Morelos, Mexico) has analyzed the congruence between microarray data and current knowledge of regulatory networks in *E. coli*. She showed that the consistency ranges from 40% to 80%, depending on the type of information and the experimental conditions used. This analysis provides a new perspective, differing from that gained from clustering analyses, that will help us to understand how regulatory networks determine expression of the genes detected by microarray expression profiles.

Finally, Yoshiyuki Sakaki (RIKEN, Yokohama, Japan), the current president of HUGO, concluded that with the opportunity to compare other complete genomes (such as that of the chimpanzee) with the human genome sequence, as well as information on differential expression of genes in different tissues, a new research period of outstanding structural, functional and evolutionary discoveries lies ahead. Abstracts of the meeting are freely available online [<http://hgm2003.hgu.mrc.ac.uk>] and the next HUGO meeting will be held in Berlin, Germany, in April 2004.