

- 3 Williams, E.J.B. and Hurst, L.D. (2000) The proteins of linked genes evolve at similar rates. *Nature* 407, 900–903
- 4 Lercher, M.J. *et al.* (2001) Local similarity in evolutionary rates extends over whole chromosomes in human–rodent and mouse–rat comparisons: implications for understanding the mechanistic basis of the male mutation bias. *Mol. Biol. Evol.* 18, 2032–2039
- 5 Goldstein, D.B. (2001) Islands of linkage disequilibrium. *Nat. Genet.* 29, 109–111
- 6 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- 7 Bernardi, G. (1995) The human genome: organization and evolutionary history. *Annu. Rev. Genet.* 29, 445–476
- 8 Smith, N.G.C. and Eyre-Walker, A. (2001) Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. *Mol. Biol. Evol.* 18, 982–986
- 9 Fryxell, K.J. and Zuckerkandl, E. (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* 17, 1371–1383

Nick G.C. Smith*

Dept of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.

*e-mail: nick.smith@ebc.uu.se

Martin J. Lercher

Dept of Biology and Biochemistry, University of Bath, Bath, UK BA2 7AY.

Meeting Report

Genotype to phenotype: associations, errors and complexity

Laura C. Lazzeroni and Chris A. Karlovich

The Keystone Symposium on Genotype to Phenotype: Focus on Disease was held in Santa Fe, New Mexico, USA, from 19 to 24 February 2002.

There is a lot of ground between genotype and phenotype, but most of it was touched on at some point during this conference. The unusually wide range of topics provided everyone with at least some new and challenging material. Early sessions presented a potpourri of disease phenotypes and underlying disease mechanisms. Other sessions addressed a variety of developments in methodology, technology and information resources.

Mouse models, human disease

We saw pictures of hairless mice, fat mice, fat-free mice, cloned mice, mice with a Marfan syndrome analog, and many more. This proliferation of mice demonstrates the importance to human genetics of efficient research tools related to model organisms. Edward Rubin (LBNL, Berkeley, CA, USA) opened the conference on this note by presenting a suite of comparative genomics software called VISTA [1], which includes the ability to visualize sequence alignments (see <http://www-gsd.lbl.gov/VISTA/index.html>). VISTA has been used to identify regulatory elements around genes and to find a new gene (*APOAV*) involved in lipid metabolism by comparing mouse and human genomic sequences.

Nathaniel Heintz (Rockefeller University, NY, USA) talked about

GENSAT, a collaboration with Mary Beth Hatten and Alexandra Joyner that aims to generate large numbers of transgenic mice expressing tagged versions of central nervous system genes using bacterial artificial chromosomes (BACs) [2]. As was demonstrated for *Zipr1*, a BAC transgene is more likely to recapitulate gene expression faithfully than a conventional transgene because its size can insulate it from positional effects, and more regulatory sequences can be included in the vector. Because the GENSAT BACs are being engineered to express enhanced green-fluorescent protein (EGFP), they should reveal very specific spatial and temporal patterns of gene expression in the brain and will be valuable tools in the study of neurodevelopment and of neurodegenerative disorders (see <http://www.hhmi.org/research/investigators/heinz.html>).

Genotypes

Several speakers presented new high-throughput assays for individual genotyping and for allele frequency estimation in DNA pools (Andreas Braun, Sequenom, San Diego, USA; Craig Gelfand, Orchid Biosciences, Princeton, NJ, USA; Wolfgang Pusch, Bruker Daltonik, Bremen, Germany). DNA pooling is an increasingly common technique whereby the DNA of hundreds of individuals is extracted from blood or tissue and pooled together. Pools can be used to screen large numbers of single nucleotide polymorphisms (SNPs)

rapidly for allele frequency differences between the samples from patients and from controls, but they are unsuitable for detailed analyses involving individual haplotypes or clinical data. Russ Higuchi (Roche Molecular Systems, Alameda, CA, USA) warned that if the level of measurement error from pooling is large, it could conceivably overwhelm the subtle differences expected for genes with low relative risks. However, Ray White (DNA Sciences, Fremont, CA, USA) pointed out that genotyping failures could systematically under-represent particular genotypes (e.g. heterozygotes) and potentially create false disease–genotype associations or remove real ones.

Lack of replication in disease-association studies

Joel Hirschhorn (MIT, Cambridge, MA, USA) addressed the persistent concern that initial reports of disease–genotype associations in human populations are often not confirmed in subsequent studies. He examined 166 associations reported in case-control studies that were later studied in at least three populations, and found that most initial findings were confirmed less than 75% of the time. However, there was substantially more positive replication of initial results than could reasonably be explained by chance or publication bias had none of the reported associations been real. For 25 of these inconsistently replicated results, Hirschhorn performed a meta-analysis that combined data from all follow-up

studies, confirming the existence of weak associations (odds ratios of 1.1 to 1.5) in seven of the cases. This could indicate that when an initial study reports a real, but modest, effect of a gene on disease risk, subsequent studies can fail to corroborate the original result consistently if they lack adequate power.

Haplotypes

Other speakers also talked about association studies, a hot topic right now in light of the proposed haplotype map of the human genome [3,4]. The rationale for the map is based on short regions of the genome, known as haplotype blocks. Linkage disequilibrium between genetic variants in the same block is high, whereas disequilibrium between variants in different blocks is low. It has been suggested that populations share the same haplotype block boundaries, but not necessarily the same common haplotypes within those boundaries. If this is true, genetic studies could be able to retrieve most available genetic information by typing a small number of SNPs, a sort of haplotype tag, within each block.

J. Claiborne Stephens (Genaissance Pharmaceuticals, New Haven, CT, USA) and Hirschhorn (in separate work with colleagues Stacey Gabriel and David Altshuler) added to the growing empirical evidence about whether haplotype blocks exist in the human genome [5]. These and previous studies have identified haplotype blocks in which four or five common haplotypes account for as many as 90% of the chromosomes in their samples, although block size and

coverage depends on how the blocks are defined. But if haplotype tags from these blocks are used in lieu of more extensive genotyping, this subset of haplotypes is likely to predict some untyped alleles better than others. Moreover, most genetic disease studies will type haplotype tags using methods that yield multilocus genotypes, not unambiguous haplotypes. Consequently, analytical methods for the map will need to work with genotype data at the observed SNPs. This could introduce additional biases, because it is easier to resolve haplotypes for some multilocus genotypes than others. As in single-locus genotyping methods, stochastic or systematic errors have the potential to overwhelm weak disease associations or to create associations where none, in fact, exist. Careful analysis will be needed to understand and adjust for such potential biases.

Contextual genetics

One day, all of this might lead to an era of personalized medicine. But first, we must understand how complex phenotypes depend upon a full complement of genetic and environmental factors. Two speakers talked about examples from the pharmacogenetics of oncology. William Evans (St Jude Children's Research Hospital, Memphis, TN, USA) showed that gene expression profiles can be used to stratify childhood acute lymphoblastic leukemia (ALL) patients to predict responses to drug therapy [6]. Nicholas Dracopoli (Bristol-Myers Squibb, Princeton, NJ, USA) reported on microarray studies aimed at identifying

a small set of gene markers that predict cancer disease outcome and response to treatment [7]. However, useful complexity has its limits. As Dracopoli said, pharmacogenetics tools will be useful to the physician only under certain circumstances. The genetic variants need to be relatively common, and strong clinical and analytical validation is essential.

References

- 1 Dubchak, I. *et al.* (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* 10, 1304–1306
- 2 Heintz, N. (2001) Bac to the future: the use of BAC transgenic mice for neuroscience research. *Nat. Rev. Neurosci.* 2, 861–870
- 3 Weiss, K.M. and Clark, A.G. (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* 18, 19–24
- 4 Helmuth, L. (2001) Genome research: map of the human genome 3.0. *Science* 293, 583–585
- 5 Stephens, J.C. *et al.* (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489–493
- 6 Evans, W.E. and Relling, M.V. (1999) Pharmacogenetics: translating functional genomics into rational therapeutics. *Science* 286, 487–491
- 7 Cockett, M. *et al.* (2000) Applied genomics: integration of the technology within pharmaceutical research and development. *Curr. Opin. Biotechnol.* 11, 602–609

Laura C. Lazzeroni*

Dept of Health Research and Policy, Stanford University School of Medicine, Stanford, CA 94305, USA.

*e-mail: Lazzeroni@stanford.edu

Chris A. Karlovich

Roche Molecular Systems, Inc., Berkeley, CA 94710, USA.

Free journals for developing countries

The WHO and six medical journal publishers have launched the Access to Research initiative, which enables ~70 developing countries to gain free access to biomedical literature through the Internet.

The science publishers, Blackwell, Elsevier Science, the Harcourt Worldwide STM group, Wolters Kluwer International Health and Science, Springer-Verlag and John Wiley, were approached by the WHO and the British Medical Journal in 2001. Initially, >1000 journals will be available for free or at significantly reduced prices to universities, medical schools, research and public institutions in developing countries. The second stage involves extending this initiative to institutions in other countries.

Gro Harlem Brundtland, director-general for the WHO, said that this initiative was 'perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries'.

See <http://www.healthinternetwork.net> for more information.