

MEETING REPORT

First International SNP Meeting at Skokloster, Sweden, August 1998

Enthusiasm mixed with scepticism about single-nucleotide polymorphism markers for dissecting complex disorders

Ann-Christine Syvänen¹, Ulf Landegren², Anders Isaksson², Ulf Gyllensten² and Anthony Brookes²

¹*Department of Medical Sciences;*

²*Department of Genetics and Pathology; Uppsala University, Uppsala, Sweden*

Single nucleotide polymorphisms (SNP) are frequent in our genomes, occurring on average once every thousand nucleotides. They are useful as genetic markers because SNPs evolve slowly and because they can be scored by technically simple methods. Moreover, a great deal of the functional variation that explains human diversity probably results from SNPs. In recent years, SNP-based strategies to explain common disease have received enormous attention from the scientific community as well as from pharmaceutical and biotechnology companies. The enthusiasm stems largely from the hope that large-scale association studies using SNPs will help identify genes that contribute to common disorders with complex genetic causation, and allow formulation of improved therapeutic drugs.

The *1st International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis* was held at Skokloster in the heart of Sweden, 29 August–1 September, 1998, with the support of the Swedish Foundation for Strategic Research, Cell Factory and Genome Research Programmes, and by the Human Genome Organization (HUGO), through a contract (BMH4-CT97-2031) between HUGO Europe and the European Commission. The meeting convened leading scientists in fields related to SNP research. Altogether around 100 scientists, representing both academic institutions and companies in 17 different countries, were accommodated at the conference facilities at the Skokloster castle on lake Mälaren, out of twice that number of applicants. The remote and peaceful rural setting of the meeting, together with the relatively small

number of participants, created a relaxed atmosphere that stimulated many informal scientific discussions.

Following an introductory lecture from Gert-Jan van Ommen (Leiden, NL), President of HUGO, there was a lively discussion on the role of HUGO in the genome project, revolving around its part in linking the diverse activities in the field, ranging from facilitating mapping and mutation analysis to nomenclature supervision, providing guidance in genome diversity and population studies and more generally in the ethical and intellectual property areas. During the two and a half days of the meeting advances in technology for discovering and scoring SNPs, SNP-based approaches to dissect complex diseases and the utility of SNPs in the study of human populations as well as in bioinformatics, computational, ethical and commercial issues related to SNPs were discussed.

SNP Discovery and Screening Methods

Many large-scale efforts are in progress to discover novel SNPs by resequencing genomic regions from several individuals. Genset (Martha Blumenfeld, France) expects to find by 1999 three polymorphisms from each of 20000 ordered BAC clones, spanning the human genome. Similarly, the whole-genome sequencing effort at Celera (Mark Adams, MD, USA) will generate hundreds of thousands of polymorphisms, although it is at present unclear whether these will be made publicly available. Andrew Clark (Penn State

University, PA, USA) and his collaborators have sequenced the LPL gene locus from 24 individuals representing three different populations, and they intend to sequence another 15 candidates to find SNPs that may be informative in the search for cardiovascular disease genes. Peter Underhill (Stanford University, CA, USA) described the usefulness of denaturing HPLC for identifying previously unknown SNPs through heteroduplex analysis, and he also presented the technique as a tool for scoring known SNPs. Several other approaches to accurate and cost-effective scoring of SNPs were presented. Microplate array diagonal gel electrophoresis (MADGE) represents a simple but efficient multiplex genotyping assay (Ian Day, Southampton University, UK). Multiplex typing on DNA arrays, either through minisequencing (Ann-Christine Syvänen, Uppsala University, Sweden, and Andres Metspalu, Tartu University, Estonia), or via a ligase assay (Ed Southern, Oxford University, UK) were shown to distinguish SNPs in heterozygote and homozygote form more accurately than hybridisation-based methods. Nonetheless, hybridisation to high density DNA chips from Affymetrix was shown to be useful for analysing HLA polymorphism (Henry Erlich, Roche Molecular Systems, CA, USA), as well as for discovering SNPs and typing thousands of them after multiplex PCR (Jian-Bing Fan, Affymetrix, CA, and Michelle Cargill, Whitehead Institute, MA, USA).

As alternatives to the DNA microarray formats, SNPs can be efficiently scored in individual amplification reactions by enzyme-assisted homogeneous assays based on fluorescence resonance energy transfer (FRET) or fluorescence polarisation (Pui-Yan Kwok, Washington University, MO, USA), using the homogeneous, FRET-based TaqMan assay (Ken Livak, Perkin-Elmer/Applied Biosystems, CA, USA), by pyrosequencing (PDI NyrJn, Royal Institute of Technology, Sweden), or by dynamic allele-specific hybridisation (DASH) (Anthony Brookes, Uppsala University, Sweden). Throughput in currently available scoring methods is limited by the difficulty of multiplexing PCR, necessitating large numbers of amplification reactions. These contribute significantly to the costs of current genotyping assays, as pointed out by James Weber (Marshfield Medical Research Foundation, WI, USA). Intramolecular ligation probes – padlockprobes – combined with rolling circle replication for signal amplification offer the promise of SNP scoring in genomic DNA without prior target amplification (Ulf Landegren, Uppsala University, Sweden). Work is in progress to

array padlock probes on planar supports, and to target multiple SNPs along DNA fibres for direct detection of haplotypes of the markers.

SNP Databases

Improved bioinformatics tools and public databases cataloguing SNP variation are of central importance to the field. The HUGO mutation database initiative was presented by Richard Cotton (Mutation Research Centre, Melbourne, Australia), and Charles Scriver (McGill University, Canada) described the database for the phenylalanine hydroxylase (PAH), as a model of a database serving both the scientific community, clinicians, and phenylketonuria patients. Elements needed fully to describe sequence variants are being worked out in conjunction with the sequence variation database at European Bioinformatics Institute (EBI), with the aim of harmonising database design and ensure accessibility (Heikki Lehmäslaiho, EMBL-EBI, UK).

Four major public SNP databases were presented or referred to:

- (i) HGBASE (<http://hgbase.interactiva.de>) from Uppsala University and Interactiva Biotechnologie GmbH links more than 2400 SNPs to genes and is designed as a purpose-built tool to facilitate association studies;
- (ii) dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/snp.how—to—submit.html>), a joint effort by the NHGRI and the NCBI, is currently accepting submissions. It can be searched via sequence-tagged site (STS) number and will be integrated with Genbank;
- (iii) the SNP Databases from Washington University and the Stanford Genome Centre (<http://www.ibr.wustl.edu/SNP/>) contain several hundred SNPs, which are being integrated into dbSNP;
- (iv) The Whitehead SNP Database (<http://www.genome.wi.mit.edu/SNP/human/index.html>) contains over 3000 SNPs, most of which are chromosomally mapped. The database is searchable by genomic region or STS number.

Owing to the anticipated large commercial use of SNPs, most discovered SNPs are currently stored in private databases owned by genomics companies or the pharmaceutical industry. In principle, SNPs can be

patented, especially if their relevance to a disease is disclosed, and many patent applications on SNPs have been filed, but as yet no patents have been granted (Christian Stein, Patent and Licensing Agency for the German human genome project). Without patent protection, a large part of the identified SNPs will no doubt remain in commercial databases that are not publicly accessible.

SNPs and Complex Disorders

So far successful association studies for identifying genes underlying complex traits have exploited SNPs within candidate genes, selected on the basis of functional clues. The clear association of the variant apolipoprotein allele E4 with Alzheimer disease exemplifies successful application of this approach, and other well known examples are the role of certain HLA alleles in diabetes or other autoimmune diseases, as discussed by Henry Erlich (Roche Molecular Systems, CA, USA). In other cases, such as in Parkinson disease, genetic associations of borderline significance have been found (Nobutaka Hattori, Juntendo University School of Medicine, Japan). The chances of finding genes underlying complex diseases may be increased by investigating SNPs in genes that encode the components of complete metabolic pathways, selected because of suspected relevance to the disease, as exemplified by genes involved in oxidative phosphorylation in Alzheimer's disease (Anthony Brookes, Uppsala University, Sweden). Nancy Pedersen (Karolinska Institute, Sweden) recommended combining association studies using SNPs with established epidemiological practices, for example by analysing samples from the extensive twin registries available in the Scandinavian countries.

An important explanation for the intense interest in SNPs is the expectation that genome-wide association studies will serve to identify previously unknown genetic variants of importance in complex disorders. Nicholas Schork (Case Western Reserve University, OH) described a positional SNP strategy to locate genes associated with complex traits that is based on calculating the regional strength of linkage disequilibrium and identification of haplotypes. As reviewed by Nigel Spurr (SmithKline Beecham, UK) the pharmaceutical industry has initiated pharmacogenomics programmes to identify SNPs, develop statistical tools, and

collect patient samples in order to use the SNPs for identifying new drug targets and to define in greater detail the genetic variation that determines the responses to drugs. So far the polymorphisms of the drug metabolising enzymes of the *CYP* gene family are the most thoroughly studied example of pharmacogenetic variation (Magnus Ingelman-Sundberg, Karolinska Institute, Sweden).

In the most intensely debated contribution at the meeting, Joe Terwilliger (Columbia University, NY, USA) presented his sceptical view of the general utility of SNPs in association studies for finding genes behind complex disorders. He cautioned that over-enthusiastic SNP researchers underestimate the complexity of the human genome and assume far too low allelic heterogeneity for genes causing complex disorder. His view received support from Rosalind Harding (John Radcliffe Hospital, Oxford University, UK) who failed to identify the mutant allele causing sickle cell anaemia using SNP data alone in a model study involving a random population sample, and from Andrew Clark (Penn State University, PA, USA), who illustrated the complex pattern of haplotype variability in the lipoprotein lipase (*LPL*) gene, due to frequent recombination within the *LPL* locus.

Population Genetics

A rapidly growing number of SNPs are being identified in the non-recombining part of the Y chromosome (Peter Underhill, Stanford University, CA, USA). Haplotypes composed of SNPs and microsatellite markers on the paternally transmitted Y chromosome complement the maternally transmitted mitochondrial DNA variation and increase the accuracy of studies on human evolution and population structures (Chris Tyler-Smith, University of Oxford, UK). Population genetic studies also provide information on the age, frequency, and distribution of SNPs in a population. This information could be used in the design and statistical interpretation of the results of SNP-based disease association studies, to reduce the problems of allelic heterogeneity. An example of this possibility was given by Svante Pääbo (University of Munich, Germany) who suggested that linkage disequilibrium created by genetic drift, rather than by a population bottleneck, could be of value to define genomic regions containing complex disease genes.

Ethical Considerations

Sandy Thomas (Nuffield Council on Bioethics, UK) drew attention to the fact that most ethical questions related to genetic information discussed so far, such as confidentiality, genetic testing and ownership issues, deal with single gene disorders. The complex genetic background of multifactorial diseases will raise new ethical issues, for which single gene diseases will not provide useful models. It is urgent that satisfactory ethical guidelines are worked out also for the common complex disorders, so that the results of the intense research on human genetic variation using SNPs can be fully used for the benefit of more targeted and preventive health care in the future.

Conclusions

The meeting left little doubt that knowledge about human genetic variation will provide useful tools for dissecting the genetics behind complex diseases. A number of more specific questions remain contentious, however, including how to ensure that information about human genetic variation is made widely available, how best to ascertain such markers in large patient cohorts and, figuring prominently at this conference, how best to design studies in order to avoid them generating overwhelming statistical noise, from which it will be hard or impossible to identify significant genetic variants.