

MEETING REPORT

4th International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis

Various Uses for DNA Variations

Anthony J Brookes*,¹

¹Center for Genomics and Bioinformatics, Karolinska Institute, Stockholm, Sweden

At the 4th International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis (Stockholm, Sweden, 10th–14th October 2001), approximately 100 scientists from more than 20 nations undertook a probing review of latest developments in the field. Despite impressive and still ongoing activities towards SNP discovery and validation, plus efforts towards haplotype exploitation, it was clear that supporting technologies for genotyping are way behind where they need to be. Innate complexity and large variances in aspects of genome function together pose immense challenges that are difficult to surmount in the human situation. In contrast, studies in simpler organisms and population/evolutionary genetics studies are yielding important new insights. Breakthroughs that are being made in understanding the genetic etiology of complex disease tend to involve genes of larger effect or extremely well merited candidates. Linkage studies and proximal phenotypes are being recommended, though the best way forward is still hotly debated. Consequently, many diverse and ambitious projects are underway, from which the data itself will eventually show what is and is not possible.

European Journal of Human Genetics (2002) 10, 153–155. DOI: 10.1038/sj/ejhg/5200774

Keywords: population genetics; variation; SNP genotyping; complex disease

Despite concerns about international travel safety and escalating global conflict, a group of approximately 100 scientists from more than 20 nations made the effort to come together in Stockholm between 10th–14th October 2001 to attend the 4th International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis. The meeting was organised by Anthony Brookes, Barbara Cohen, Svante Pääbo, and Claes Wahlestedt, and it provided another in the yearly series of opportunities to undertake a frank and insightful review of ongoing developments in the field, without undue commercial bias. To enable this, attendance was competitive based upon reviewed scientific contribution, and sponsorship was provided without rights to influence

any aspect of the event. Abstracts from all the meeting presentations are available at <http://snp2001.cgr.ki.se/>

Keynote presentations by Eric Lander (Whitehead Institute) and Craig Venter (Celera Genomics), summarised the state of the art of using SNPs to analyse phenotypic variance, explained the continued value of large scale sequencing practice, and emphasised the role of major international collaborative efforts to map genome polymorphism and haplotype structures. This progress was then placed in a real-world perspective by enlightening ethics presentations delivered by Bernadette Modell (University College Medical School) and Bartha-Maria Knoppers (Université de Montréal). They made it clear that the most immediate application of new genetic knowledge will be for diagnosis and prevention, especially in the developing world, but that at the present time the supporting ethical systems (established around monogenic disease) are not ready to handle complex disease etiologies. Relevant here is the fact that we all carry many disease risk variants, as emphasized by Leena Peltonen (UCLA School of Medicine).

*Correspondence: Professor Anthony J Brookes, Center for Genomics and Bioinformatics, Karolinska Institute, Berzelius väg 35, S-171 77 Stockholm, Sweden. Tel: +46 (0)8 7286630; Fax: +46 (0)8 324826; E-mail: Anthony.Brookes@cgb.ki.se
Received 2 December 2001; revised 7 December 2001; accepted 14 December 2001

Population and evolutionary genetics studies continue to be areas of fascinating activity, endeavouring to explain patterns of genome sequence variance and thereby relate this to function. By using coalescent theory and considering genetic drift and recombination, Gabor Marth (NCBI) computationally examined 500 000 predicted genomic SNPs and uncovered evidence for a significant human population size collapse several hundred generations ago. Investigating genetic features that define humans as a species, Svante Pääbo (Max-Planck-Institute for evolutionary Anthropology) observed that between humans and chimpanzees the rate of over-all transcriptome change in the brain has accelerated about threefold over that in other organs. Additionally, distinct footprints of positive selection are being revealed by generally comparing non-synonymous and synonymous alterations in coding regions both between and within species, as reported by David Liberles (Stockholm University) and Justin Fay (Berkeley). Jeffrey Long (University of Michigan) reached similar conclusions, specifically for functional variants of the ADH and ALDH2 genes within Asian populations.

As in previous years, many presentations focused upon trying to understand genome function in the human. However, many new insights about genome function are actually emerging from studies in other species. For example, Thomas Mitchell-Olds (Max-Planck Institute of Chemical Ecology) showed convincing SNP-based evidence that balancing selection was occurring in *Arabidopsis thaliana*, and Peter Oefner (Stanford Genome Technology Center) reported how heterosis plus interactions between alleles in several neighbouring genes together explained the ability of clinical yeast strains to grow at high temperature.

For the last several years, great efforts have been put into human SNP collection, database development, and genotyping technologies. Lincoln Stein (Cold Spring Harbor Laboratory) updated delegates on The SNP Consortium (TSC) (snp.cshl.org) which has predicted >2 000 000 SNPs and, with Orchid BioSciences, will now characterize allele frequencies for some 60 000 of these whilst also defining haplotypes. Ray Miller (Washington University School of Medicine) also reported validation efforts upon TSC SNPs. Generally, high levels of validation are seen once repeat sequences and other problematic loci are discounted. However, there seems to be less than 50% chance that any random predicted genomic SNP will exist in a researchers own population of interest, a fact which retards research progress. EST-based SNP prediction efforts were summarised by Christopher Lee (UCLA) and Kenneth Buetow (NCI Center for Bioinformatics), the latter team also working towards comprehensive validation studies. Similar cSNP validation activities are ongoing in a partnership between Sequenome and Incyte, as presented by Dirk Walther (Incyte Genomics). David Fredman (Karolinska Institute) and Stephen Sherry (NCBI) presented database efforts to store such data, respectively the HGVbase (<http://hgvbase.cgb.ki.se>) project

in Europe (recently renamed from HGBASE, reflecting its goal of establishing a unified summary of all known clinical mutations and Human Genome Variation) and the dbSNP effort (<http://www.ncbi.nlm.nih.gov/SNP>) in the US (an archival polymorphism database with information tightly integrated to a range of other NCBI resources). A point has thus now been reached where lists of predicted SNPs have grown from thousands to millions, and the characterisation of these continues with haste.

To use these SNP resources effectively requires suitable genotyping technologies. Many alternatives exist, and the better of these can generate several thousand genotypes per day. Sample pooling helps in a way, but destroys all opportunities for useful downstream processing such as haplotype or clastic analysis. At the meeting, glimpses into several possible next-generation technologies were presented. Anthony J Brookes (Karolinska Institute) showed proof-of-principle data for how the simplicity of hybridization and the robustness of dynamic temperature sweeps (as in the DASH assay) can be combined into a system for producing around 1 000 000 genotypes per day at a total cost of around USD 1000. Erica J Beilharz (Lynx Therapeutics) presented the Megatype technology, wherein whole-genome sets of restriction sites having differential allele frequencies in two test populations may be selected upon microbeads, thus achieving single-pass genome association scanning. Peter Brooks (Centre National de Genotypage) reported a greatly improved system for direct identification of chromosomal regions that are identical by descent between two related individuals. Most ambitiously, Colin Barnes (Solexa Ltd) summarised bold efforts towards total human genotyping and re-sequencing, based upon solid-phase sequencing of a huge number of isolated single DNA molecules in parallel.

Regarding complex disease studies, there were disappointingly few breakthroughs reported. Many delegates considered that slow progress on relating genotypes to phenotypes likely reflects an under-estimation of the level of complexity inherent in that relationship. To deal with this, François Cambien (INSERM) suggested that 'proximal' phenotypes (features in biological pathways closer than crude 'case-control' status to the genetic alteration) might be more reasonable targets for study. Similarly, Anthony Brookes (Karolinska Institute) compared positional-candidate genes (from regions suggested by comprehensive linkage findings) to biological-candidate genes (targets biased by what we think we know of disease processes), and proposed that association analysis upon the former would be more likely to succeed with present-day resources. Candidate gene approaches may be viable for particularly well-recommended candidates, such as the known appetite modulator Neuropeptide Y, within which sequence variants were reported to influence bodyweight by Claes Wahlestedt (Karolinska Institute). But such studies seem to fail all too often. It is therefore interesting to note that the majority of the few presently established 'complex disease' loci actually involve

effect sizes big enough to be revealed by linkage scans. An excellent example here would be Gilles Thomas (CEPH) discovery of risk alleles for Crohn's disease in the gene NOD2/CARD15. Other strong arguments for why family linkage studies should be prioritized were delivered by Joseph Terwilliger (Columbia University). The alternative of whole genome association scanning is not yet technically feasible with SNPs, but was predicted to be viable in principle by Chun-Fang Xu (GlaxoSmithKline) who showed that a scan of approximately 1 Mb around the CYP2D6 would give strong positive signals for the poor drug metaboliser phenotype.

Delivering another perspective on the slow progress of complex disease studies, a small contingent at the meeting suggested that the field may be generally trying to do the impossible. Significantly, this group includes some leading thinkers such as Andrew Clark (Pennsylvania State University) and Joseph Terwilliger (Columbia University). The message they present is that whilst a few common variants of large effect may exist and will be uncovered, the majority of complex phenotypes would be expected to involve enormous levels of heterogeneity, epistasis, environmental influence, and phenocopies – placing them beyond the reach of any technologies or strategies likely to be established in the foreseeable future.

Assuming that the bleakest view above is too pessimistic, a number of delegates reported on efforts to enhance research strategies by moving beyond the study of individual SNPs, towards the examination of haplotypes (patterns of observed alleles along portions of individual chromosomes). The range of haplotypes that exist in a population will be influenced by many factors including population demographics and history, mutational processes, selective pressures, genetic drift, and patterns of recombination – but critically they will not be random. Accordingly, data was presented showing that chromosomal regions were often dominated by only a

few common haplotypes arranged in 'blocks' (somewhat discrete and independently occurring units). This has inspired the concept that haplotype blocks (characterisable by merely a few component SNPs) could replace individual SNPs as the 'marker' one would test for disease association – enabling much faster, cheaper, and simpler genome association scans. But for this to work, haplotype blocks must truly exist as units, they must be standard across populations, and they must harbour risk-modifying alleles non-randomly. Showing faith that these things will (at least sometimes) be true, efforts are now being initiated to create a human 'haplotype map' to support haplotype-disease association studies. Initial data from Patricia Taillon-Miller (Division of Dermatology) indicates extensive haplotype blocks of hundreds of kb or more, but this is contradicted by the findings of Kenneth Kidd (Yale University School of Medicine), Clay Stephens (Genaissance Pharmaceuticals), Peidong Shen/Peter Oefner (Stanford Genome Technology Center), and Augustine Kong (deCODE Genetics) who each reported that marker frequencies, linkage disequilibrium and haplotype patterns are often highly variable both between populations and between genomic regions. These conflicting data are yet to be explained, but raise concerns that there may be difficulties in creating and using 'a' human haplotype map. Such uncertainty and debate thus continues to epitomise this new research field, and it will be interesting to see how things look at next years meeting.

Acknowledgments

The meeting was sponsored principally by the Swedish Wenner-Gren Foundations, who also included it in their own Symposium Series and hosted the whole event at their conference center. Additionally, Orchid BioSciences volunteered significant additional funds to cover all extra costs.