European Journal of Human Genetics (2001) 9, 316–318

www.nature.com/ejhg

MEETING REPORT

3rd International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis: SNPs: 'Some Notable Progress'

P Scott White*[,1], Pui-Yan Kwok[2], Peter Oefner[3] and Anthony J Brookes[4]

[1]*Center for Human Genome Studies and Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA;* [2]*Division of Dermatology, Washington University School of Medicine, St. Louis, MO, USA;* [3]*Stanford DNA Sequencing & Technology Center, Palo Alto, CA, USA;* [4]*Center for Genomics Research, Karolinska Institute, Stockholm, Sweden*

Fervent activities for the collection and exploitation of single nucleotide polymorphism (SNP) data continue, amid concerns about their real utility. The desire to understand complex disease aetiology remains a key driving force for this activity. Recent developments provided a level of cautious optimism not seen in previous International Meetings on Single Nucleotide Polymorphism and Complex Genome Analysis. The 3rd such meeting, held 8–11 September 2000 in Taos, New Mexico, covered research on technologies for SNP scoring, analytical tools for using SNPs to map disease genes, examples from researchers using SNPs for specific disease studies, and databases and tools for facilitating these activities. Studies of human history, and a range of studies upon model organisms were also represented. Whilst the transition from technology oriented work (methods, discovery, etc.) to successful biological application is occurring relatively slowly, a clear trend in this direction is now apparent, and it will surely gain momentum in future months and years. Many fundamental properties of SNPs remain unknown, and many other basic questions are still unanswered, but the field is moving forward on all necessary fronts, promising exciting advances just around the corner.
European Journal of Human Genetics (2001) **9**, 316–318.

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

Introduction

Much of the early excitement about the potential of SNP data continues to be viewed cautiously by many researchers attempting to locate causative genes for a variety of genetic disorders. Optimal clinical sampling, availability of ideal SNPs, unknown population genetics parameters, technological limitations, and adequate data analysis methods, present imposing barriers for researchers involved in common disease gene hunting. These issues formed the main focus of attention for almost 150 scientists representing 16 countries at the 3rd International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis, in Taos New Mexico, between 8–11 September, 2000. Abstracts from all the meeting presentations are available at http://snp2000.cgr.ki.se

Even before the announcement of the completed draft of the human genome, efforts have been shifting from producing a single genome sequence towards functional genomics questions, not least the analysis of genetic variation and the aetiology of complex disease. An overview of the potential of this field, plus its challenges, was presented by Ray White (DNA Sciences, Mountain View, CA, USA). Nir Navot (GamidaGen Ltd., Ashdod, Israel) then gave examples of real-world practical application of genetic

*Correspondence: P Scott White, PhD, B-N1, Mailstop M888, Genomics, and Center for Human Genome Studies, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
Tel: +1 505 665 3860; Fax: +1 505 665 3024;
E-mail: swhite@telomere.lanl.gov
Received 13 November 2000; revised 24 November 2000; accepted 28 November 2000

screening, now widely accepted in different settings in Israel.

Available SNPs and genotyping tools

The first technological step towards success has essentially been taken, in that large collections of SNPs and other variations data do now exist in the private and public domains. However, discovery of new SNPs remains necessary for those with highly targeted approaches where SNP coverage is not adequate in the specific genes or populations under analysis. To be useful, the crude lists of SNPs must now be endowed with various properties, not least descriptions of their allele frequencies in various populations, and their haplotype and Linkage Disequilibrium (LD) arrangements. Valuable efforts towards establishing a range of population allele frequency estimates for over 50 000 SNPs were summarised by Michael Phillips (Orchid BioSciences, Inc., Princeton, NJ, USA). Databases (HGBASE and dbSNP) and processing tools for handling the masses of emerging SNP data were summarised by Heikki Lehväslaiho (EMBL, Cambridge, UK) and Stephen Sherry (NCBI, Bethesda, MD, USA), and useful assay ontology development work was detailed by Francisco De La Vega (PE Biosystems, Foster City, CA, USA).

A second technological step that must be taken is the establishment of genotyping platforms. Impressive tools are now available, but none can be said to provide the functionality needed to underpin truly exhaustive exploitation of SNPs. Scott White (Los Alamos National Laboratory, Los Alamos, NM, USA) presented an overview of technologies, breaking them down into assays, platforms, and formats. The number of samples, and the number of sites to be scored per individual, will impact which approaches make the most sense to implement. There is no single globally 'optimal' technology.

In focus at the meeting were hybridisation and single base extension methods, although poster presentations represented essentially the entire spectrum of scoring assays. Magnus Jobs (Karolinska Institute, Stockholm, Sweden) presented an update on Dynamic Allele Specific Hybridisation (DASH), a simple and robust system, which achieves high discriminatory power by scanning all hybridisation stringencies. The same 'dynamic stringency' concept was apparent in electronically-addressable chips presented by Ron Sosnowski (Nanogen, Inc., San Diego, CA, USA), and temperature regulated chips reported by Ali Arjomand (CombiMatrix Corp., Snoqualmie, WA, USA). Multiplexed single-base extension reactions carried out in a single tube were described by Penny Dong (PE Biosystems, Foster City, CA, USA) and others, and Vincent Phillips (Quantum Dot Corp., Palo Alto, CA, USA) presented several new assays implemented on beads labelled by 'quantum dot nanocrystals' – multi-colour fluorophores with high sensitivity, low cross-talk, photostability and single wavelength excitation. Flow cytometry-based systems are receiving considerable interest, for example SNP Genomic Analysis using Multiplexed Microsphere Arrays (SNP-GAMMArrays) were presented by Scott White (Los Alamos National Laboratory, Los Alamos, NM, USA), as were similar developments by Michael Wagner (Glaxo Wellcome, Inc, Research Triangle Park, NC, USA). Finally, some revolutionary ideas that were notable included immobilised PCR-colony based (Polony) sequencing from Robi Mitra (Harvard Medical School, Boston, MA, USA), and whole genome direct cloning of identical-by-descent regions from Vivian Cheung (Univ. of Pennsylvania, Philadelphia, PA, USA).

Strategy and analytical methods

Single SNP analyses in association searches are widely used. One feature of this strategy is that it almost always depends to some degree upon LD. Evidence presented by amongst others William Cookson (Wellcome Trust Centre for Human Genetics, Oxford, UK), Deborah Nickerson (Univ. of Washington, Seattle, WA, USA), Pui-Yan Kwok (Washington Univ. School of Medicine, St. Louis, MO, USA), Andrew Collins (Univ. of Southampton, UK), and David Nelson (Baylor College of Medicine, Houston, TX, USA), showed that LD can extend for many tens or even hundreds of kb, while 'usefully strong' LD may sometimes only cover some 50 kb at best, or less than the length of many individual genes. A potentially more meaningful distance perspective (4Nc), based upon the recombination scale and the effective population size, was put forward by Tony Long (Univ. of California, Irvine, Orange, CA, USA). Although populations with distinctly different histories were similar in the general LD patterns, immense variance in LD span was reported to exist throughout the genome, prompting an urgent call from Pui-Yan Kwok (Washington Univ. School of Medicine, St. Louis, MO, USA) for the creation of a genome-wide LD map. Where LD is high, the association studies could be enhanced by an analysis of haplotypes. However, current tools for predicting haplotypes from genotypic information are of limited accuracy, according to Colin Dykes (Variagenics, Inc., Cambridge, MA, USA). Li Jin (Univ. of Texas-Houston, TX, USA/Shanghai, China) and Klaus Rohde (Max-Delbrueck-Centrum, Berlin-Buch, Germany) are therefore each developing new and improved algorithms.

Designing association studies holds many pitfalls. One repeated message, delivered most eloquently by Dahlia Nielsen (North Carolina State Univ., Raleigh, NC, USA), was that for indirect association analysis to work, pathogenic alleles and tested markers need to not only be in high LD, but also to have similar allele frequencies. Guidance about which SNPs might be more likely to be functionally important, is starting to emerge from independent efforts reported by John Moult (Univ. of Maryland, Rockville, MD, USA), Shamil Sunyaev (EMBL, Heidelberg, Gemany), and Gerald Wyckoff (Univ. of Chicago, Chicago, IL, USA). By various means, these teams are categorising non-conservative cSNPs or amino-acid

replacements generally, with respect to their predicted functional impact.

Many questions still remain about the range of analysis strategies one might aim to employ. Andrew Collins (Univ. of Southampton, UK) argued that the parameter rho was superior to alternatives, and he and many others are now using data randomisation strategies to evade the need for crude Bonferroni corrections. Jurg Ott (Rockefeller Univ., New York, NY, USA) and Pamela Flodman (Univ. of California, Irvine , Orange, CA, USA) are working on analysis systems that will extend association studies to multiple loci, and therein hopefully reveal interactions. Richard Spielman (Univ. of Pennsylvania School of Medicine, Philadelphia, PA, USA) made a case that the transmission/disequilibrium test (TDT) was more robust to population structure, background association and heterogeneity than simpler and cheaper case-controls designs.

Applications and biological insights

Using a range of available tools and ideas, many efforts to use SNPs are now ongoing, and these studies span many organisms. Work was reported by Kerry Kornfeld (Washington Univ. School of Medicine, St. Louis, MO, USA) summarising positional cloning advances in *C. elegans*, and by Lars Steinmetz and Peter Oefner (Stanford Univ., Palo Alto, CA, USA) showing how DHPLC and array hybridisations can be employed together to rapidly positionally clone phenotypic loci in both *S. cerevisiae* and *A. thaliana*. The genetic basis of bristle number variation in *Drosophila* is under investigation by a candidate gene SNP association approach, as summarised by Charles Langley (Univ. of California, Davis, CA, USA). In addition, various interesting posters reported SNP-based analyses in situations as diverse as mice, maize, and the livestock industry.

In the human situation, reported studies touched upon questions of human history and the nature of complex disease aetiology. Surprisingly (disappointingly) few new advances in biological knowledge were reported at the meeting, despite the range of technology and strategic developments that have emerged over the last several years. However, improved understanding of the nature of the global spread of modern man based upon mitochondrial DNA analyses was summarised by Douglas Wallace (Emory Univ. School of Medicine, Atlanta, GA, USA), and Joanna Mountain (Stanford Univ., Stanford, CA, USA) reviewed evidence from RFLP data that there really is very little genetic differentiation between various geographical groups, with typical SNPs probably being up to seven times more ancient in origin than the time of the earliest isolating migrations. On a more recent time-scale, Levy Ulanovsky (Los Alamos National Laboratory, Los Alamos, NM, USA) reported that surprisingly few Ashkenazi haplotypes exist, suggesting greater rates of genetic drift in this population than previously suspected.

Finally, turning to the holy grail of the SNP field – improved understanding of the aetiologic role of SNPs in the predisposition of complex disease – a handful of interesting data were presented, though one might have hoped for more. Targeted candidate gene SNP searches for direct or indirect associations with a (disease) phenotype, appear to be the most promising with present day technologies. Reported 'success' stories, yet to be widely replicated, include evidence presented by Anthony Brookes (Karolinska Institute, Stockholm, Sweden) from a screen of over 50 candidate genes identifying a potential role for apoptosis (TNFRSF6 promoter variation) in Alzheimer's Disease, and data from David McDermott (NIH, Bethesda, MD, USA) showing that RANTES promoter variations may affect the natural history of HIV infection and AIDS. The commonality of promoter variations in these two reports should perhaps not be ignored.

Interactions between the environment and genetic polymorphisms were suggested by two reports. Specifically, Scott Diehl (NIH, Bethesda, MD, USA) argued that alcohol exposure and ADH gene SNPs interacted in the predisposition of nonsyndromic cleft lip and palate defects, and Gerome Breen (Univ. of Aberdeen, UK) reported highly significant data suggesting an interaction between malnutrition and APOE alleles in the development of schizophrenia in Chinese clinical materials.