

Dominique Rocha · Ivo Gut · Alec J. Jeffreys
Pui-Yan Kwok · Anthony J. Brookes
Stephen J. Chanock

Seventh international meeting on single nucleotide polymorphism and complex genome analysis: ‘ever bigger scans and an increasingly variable genome’

Received: 28 January 2006 / Accepted: 30 January 2006 / Published online: 24 February 2006
© Springer-Verlag 2006

Abstract In September 2005, the seventh international meeting on single nucleotide polymorphism (SNP) and complex genome analysis was held in Hinckley, near Leicester, UK and the meeting was organised by Anthony Brookes, Stephen Chanock, Ivo Gut, Alec Jeffreys and Pui-Yan Kwok. Similar to prior meetings, the 3-day meeting focused on new trends and methods in the analysis of SNPs and complex human disease. A substantial portion of the meeting was devoted to preliminary analyses of data emerging from the International HapMap Consortium and addressed key issues in patterns of recombination, linkage disequilibrium and population genetics. Of great interest were the sessions that addressed SNP analysis in other species and the emerging field of copy number variation. Overall, there have been a number of recent advances in genomics that

promise to accelerate the pace of dissecting the genetic basis of many complex diseases in humans—and perhaps in other species.

Introduction

The meeting was organised into nine oral sessions with ample time for questions and discussion, plus two extended poster sessions. Spirited discussion was evident throughout the meeting as many presenters offered preliminary or unpublished data. The major themes of the meeting are summarised by the sessions below.

Copy number and structural variation

The program began with a session on structural variation in the human genome. Charles Lee (Harvard Medical School) reviewed the emerging literature on copy number variation and presented data using array comparative genomic hybridisation. It is evident that there is extensive heteromorphology in the genome; the gain or loss of both small and large regions of the genome is greater than previously anticipated, and these are known as copy number variants (CNVs). CNVs likely contribute substantially to phenotypic heterogeneity in the human genome and it is estimated that each individual harbours perhaps hundreds of CNVs. To address this issue, speakers in the session discussed a new database for annotation of these types of polymorphisms (<http://www.projects.tcag.ca/variation>). Furthermore, it was announced that a new, international consortium, including Harvard Medical School, the Wellcome Trust Sanger Institute, the University of Toronto and Affymetrix have begun to comprehensively identify CNVs in the 270 HapMap samples (Altshuler et al. 2005) using reasonably high-resolution, genome-wide array platforms. Long-term plans include further anal-

Anthony J. Brookes and Stephen J. Chanock contributed equally to this work.

D. Rocha (✉)
Genus Cambridge Research Laboratory, University of Cambridge,
Department of Pathology, Genus plc, Tennis Court Road,
CB2 1QP, Cambridge, UK
E-mail: dr219@hermes.cam.ac.uk
Tel.: +44-1223-333645
Fax: +44-1223-333345

I. Gut
Centre National de Genotypage, Evry, France

A. J. Jeffreys · A. J. Brookes
Department of Genetics, University of Leicester,
University Road, LE1 7RH, Leicester, UK
E-mail: ajb97@leicester.ac.uk

P.-Y. Kwok
Cardiovascular Research Institute, University of California,
San Francisco, CA, USA

S. J. Chanock
Section on Genomic Variation, Pediatric Oncology Branch,
Center for Cancer Research & Core Genotyping Facility, Division
of Cancer Epidemiology and Genetics, National Cancer Institute,
Bethesda, MD, USA

ysis of approximately 1,300 individuals from the CEPH diversity panel. The importance of CNVs was underscored by Anthony Brookes (University of Leicester) who discussed the analytical challenges of testing copy number variation in large numbers of individuals at high resolution. He further discussed the utility of the technical platform, DASH, which has been useful in the identification of low-frequency copy number variation and from this estimated that perhaps one quarter of the genome could be involved in copy number variation, if one includes uncommon alleles. He also discussed the possibility that structural variation could be important in mapping complex human diseases. Nancy Cox (University of Chicago) continued this theme and offered analytical approaches towards the identification and application of CNVs in disease association studies. Specifically, she discussed an approach that utilises loss of heterozygosity data or data previously rejected in large-scale genotype studies, such as the International HapMap Consortium.

Nonhuman single nucleotide polymorphisms

Elaine Ostrander (National Human Genome Research Institute) presented an update on the mapping of complex traits in different breeds of the domestic dog. A phylogenetic study of common single nucleotide polymorphisms (SNPs) in 125 dog breeds was presented and examples were discussed that illustrate the utility of candidate gene association studies. There was an extensive discussion of the characterisation of candidate genes and regions for skeletal traits in Portuguese Water Dogs. Claire Wade (Massachusetts General Hospital and The Broad Institute of Harvard and MIT) presented a 20 kb genome-wide haplotype map of 50 inbred strains of mice. In an analysis of the initial draft sequence of the mouse genome, there was evidence of long haplotype blocks (1–2 Mb) using SNPs at an interval of one every 20 kb. To further address this issue, Perlegen Sciences has re-sequenced 15 inbred mice strains (11 lab strains and 4 wild strains) and this is being analysed in collaboration with The Broad Institute of Harvard and MIT. Affymetrix arrays were used to analyse 155,000 SNPs in 50 strains, in order to build a mouse HapMap. This HapMap is now being used for phylogeny studies, the identification of genes involved in coat colour and other complex traits and integration with variation patterns of expression data. Dr. Wade presented data on the haplotype structure of the dog genome on behalf of Kerstin Lindblad-Toh (The Broad Institute of Harvard and MIT). In the primary analysis of the canine genome project, over 1,500,000 SNPs were identified, which together with SNPs identified through the comparison of the Boxer sequence to the Standard Poodle sequence generated by The Institute for Genomic Research result in a database of over 2.5 million canine SNPs. The re-sequencing of ten 15 Mb long genomic regions (each one from a different chromosome) of 24 different dog

breeds plus 20 representatives from a further 9 breeds indicates that LD in the dog is biphasic. The first phase corresponds to a bottleneck estimated to have occurred approximately 27,000 years ago and was associated with the domestication of the wolf. The second phase results from a recent second bottleneck that occurred roughly 200 years ago with the creation of specific dog breeds. It was proposed that breeding history could account for the observed short linkage disequilibrium blocks between breeds but considerably longer LD blocks observed within a breed. Moreover, this could be exploited for mapping complex diseases across individual breeds. Already, a 20,000 SNP chip has been developed by the Broad Institute to accelerate mapping complex diseases in canines. Jonathan Flint (University of Oxford) completed the session with a presentation of how to use the mouse SNP data to map quantitative trait loci (QTL) in outbred mice. Claire Wade had described work on inbred mice; however, such mice have greater LD and are not so useful for fine mapping. The group in Oxford has developed a heterogeneous stock (HS) of mice, carrying ‘mosaic’ chromosomes, derived from eight different inbred mouse strains. Using an Illumina mouse SNP panel of more than 15,000 SNPs in ~3,000 HS mice (<http://www.well.ox.ac.uk/mouse/snp.selector>), 11,000 SNPs were determined to be polymorphic in the HS set (one marker every 200 kb). LD decay was observed to be less than 3 Mb. A genetic map containing 11,558 SNPs is currently under construction.

Linkage disequilibrium, recombination and HapMap

Peter Donnelly (University of Oxford) discussed the generation of a fine-scale map of recombination rates and the identification of ~30,000 recombination hotspots across the human genome using the 1.6M SNP data set generated by Perlegen Sciences and the HapMap resource (Altshuler et al. 2005). Recombination is evident throughout the genome and specifically dominated by hotspots. It is estimated that there is one hotspot every 50 kb of the genome, and the human recombination landscape appears similar in male and female meioses. The analysis of the sequence in these hotspots identified, for the first time, several sequence features associated with human hotspots. In two existing examples of mutations which disrupt hotspot activity, from Sir Alec Jeffreys’ group, it was shown that the disrupting mutation involves a change to the newly identified motifs. Recent studies in humans and chimpanzees indicate that hotspot locations, and recombination rates over small (kilobase) scales, are evolving quite quickly on evolutionary timescales, but there is evidence that rates are conserved over megabase scales. Stacey Gabriel (The Broad Institute) discussed technical platforms available for conducting genome-wide association studies and the investigation of patterns of LD and recombination. Between 2000 and 2005, the number of common human SNPs (with MAF greater

than 5%) identified has grown enormously from less than 1 million to more than 9 million. During the same time the multiplex level has increased (from 5 to 10-plex to 250,000-plex) and the cost per genotype diminished (from \$0.2–0.5 to \$0.01–0.1). In order to analyse the denser data sets, Web-based tools have been developed, such as the haplotype tagging program as well as programs for detection of large-scale chromosomal alterations, loss of heterozygosity and whole genome association studies. The presentation addressed issues of genome coverage and technical capabilities. Specifically, to conduct dense SNP studies, a genotyping platform should overcome signal/noise issues, be highly parallel and automated, accurate and affordable. It was recommended that all markers found positive during association studies be re-genotyped, to make sure that the association was not due to an error in genotype calling. Jonathan Marchini (University of Oxford) presented a general statistical framework for the localisation and detection of disease genes. For association studies, genotypes will only be available at a subset of SNPs across a region and many SNPs will effectively consist of missing data. A method was described that uses the fine-scale recombination map (derived from the HapMap data) to estimate the missing genotypes and carry out a statistical test for association at all observed and unobserved SNP locations. The software will soon be available from the Website (<http://www.stats.ox.ac.uk/~marchini/software.html>). Sarah Murray (Illumina) described the algorithms used to develop new panels for the Illumina Infinium II genotyping product. Efforts have focused on SNP panels of nonsynonymous SNPs and extensive coverage of the genome using a pairwise r^2 approach that captures bins of SNPs. Mitali Mukerji (Institute of Genomics and Integrative Biology, Delhi) described the Indian Genome Variation project. India with roughly one-sixth of the world population, its multitude of linguistically and ethnically diverse populations and communities has set up the Indian Genome Variation Consortium of six research institutes focusing on common diseases, pharmacogenomics and population diversity. The consortium plans to collect 15,000 samples from distinct populations and analyse 1,000 genes of medical importance. The consortium has initiated the project by re-sequencing high-profile candidate genes and found that ~12% of SNPs are novel, many of which are private to populations of the Indian subcontinent. A portal for the database and tools has already been developed. Francisco De La Vega (Applied Biosystems), in collaboration with SharDNA, presented data on the patterns of LD between SNPs in an isolated Sardinian population and the implications for the selection of markers for disease association studies. DNA samples from 101 “unrelated” individuals (second cousins or beyond) from the village of Talana (isolated since the seventeenth century) were typed for 771 SNPs from an 8 Mb region from HSA22 with the SNPlex™ System and the results compared with data from 45

Caucasian Coriell samples. They found that over 90% of the SNPs were polymorphic, that LD extends for longer distances and, although most common haplotypes were present in both samples, haplotype diversity is more restricted in this isolated population. These observations suggest that association studies in this isolate would require over 20% fewer SNPs for equivalent power compared to an outbred sample.

Long-term issues in genetic variation

Lisa Brooks (National Human Genome Research Institute) discussed the developments in the International HapMap Consortium and reviewed the preliminary results of HapMap1 published in October 2005 (Altshuler et al. 2005). With many groups around the world preparing to conduct whole genome scans using dense SNP platforms, it was proposed that the results of the association studies should be released into the public domain after 6 months to encourage others to utilise data sets and to develop new tools for complex analytical approaches. Moreover, there is a commitment to support the discovery and application of structural variation, specifically as it pertains to association studies. Generation of comparative data across species will also be supported, in order to identify not only conserved regions, but also regions that differ within populations. Anne Cambon-Thomsen (INSERM U558, Toulouse) discussed the question ‘do new scientific and technical dimensions in genomics challenge ethical issues or is it the other way around?’. Her presentation addressed the complex dynamic between the scientific potential associated with SNP studies and the possible deleterious perception of widespread genetic testing. In anticipation of the development of biobanks, or what Prof. Cambon-Thomsen called ‘biobank-omics’, it was argued that it will be important to assess the use and utility of these biobanks. She suggested putting in place a bioresource impact factor, known as a BRIF, one that could serve as a surrogate for measuring the value of a project or data collection. In this regard, the case was made that science is not only challenging the limits of ethics, but also that the ethical issues are, in turn, shaping scientific choices and research design. Paul Burton (University of Leicester) concluded the session with the case for the scientific utility of the UK Biobank project, a population-based prospective cohort study that plans to follow 500,000 people between the ages of 40 and 69 years for at least 30 years. Though this project has been questioned by some experts, the value of a population-based prospective cohort study was stressed, both in its ability to accumulate sufficiently large sample sets and the opportunity to study more complex interactions between diseases, biomarkers, genetics and the environment.

Expression variation

A session was devoted to the emerging field of analysing microarray expression data with dense SNP markers to determine the contribution of both *cis*- and *trans*-acting factors on gene expression. David Barker (Illumina) presented recent data on characterisation of human embryonic stem cells by gene expression and SNP genotyping based on an analysis of 18,000 genes assayed by expression profile. Interestingly, stem cells expressed more genes compared to other cell types studied in this project. Stem cells from different passages as well as lab origins were genotyped with 100,000 SNPs to identify copy number polymorphisms in late passages. He also described a new assay for determining DNA methylation at 1,536 sites simultaneously and demonstrated dramatic differences in the methylation patterns among human ES cells, normal adult cells and cancer cells. Emmanouil Dermitzakis (Wellcome Trust Sanger Institute) presented data on the association between common genetic variants and gene expression. The analysis of the expression of ~700 genes in 60 Caucasian samples included publicly available genotypes from the HapMap project using an approach known as 'expression' quantitative traits (eQTLs). A number of SNPs were observed to have significant associations and interestingly included SNPs located *cis* and others located megabases away from genic regions of altered gene expression. These latter *trans*-acting SNPs might be noise but further study will address these issues.

Molecular signatures of evolution

Gil McVean (University of Oxford) discussed the scale-specific effects of selection, mutation and recombination on human genetic diversity. Since the HapMap project has provided extensive data on common genetic variation, it is possible to analyse these features at the genomic scale level as well as at a fine regional scale. To address the issue of biological significance of differences in the scope of analysis, the approach of wavelet transformation was presented. This method allows identification of the proportion of variation acting at different scale (broad or fine scale). Using the wavelet-based method on genotyping data in three different populations, it was reported that the mutation rate influences genetic variation on a broad scale, whereas recombination hotspots have a local effect on the frequency and nature of genetic variation. But how does recombination promote polymorphism? It is possible that recombination hotspots mediate genetic variation, via a biased gene conversion favouring GC mutations over AT mutations. Mark Shriver (Penn State University) presented data on the evolutionary genetics of normal variation in human skin pigmentation, which could be under selection pressure. For example, high pigment content confers sunburn protection. To identify genes

involved in skin colour, he presented data on 11,078 SNPs (plus 313 X-linked SNPs) in six different human populations. Using pairwise F_{st} statistics, the study has identified SNPs (in *ASIP*, *OCA2*, *TYR*, *MATP* and *SLC243A5* genes) that are associated with skin pigmentation. The allelic distribution across 52 world populations for SNPs in *TYR*, *MATP* and *SLC243A5* genes showed strong differences between Europeans and non-European populations and also differences between East Asians and Europeans, suggesting a 'dark-skinned' ancestor origin followed by different 'lightening' events. Christopher Carlson (Fred Hutchinson Cancer Research Center) talked about identifying functional variation using evolutionary analysis. In an analysis of 1.6 million SNPs across 71 samples (23 African American, 24 European American and 24 Chinese) generated by Perlegen (Hinds et al. 2005), he identified large population-specific contiguous regions with decreased nucleotide diversity using the Tajima's D statistic. In select regions known as CRTRs, there was an observed reduction in Tajima's D statistic (Carlson et al. 2005). Genes in 'craters' are currently being re-sequenced, in order to analyse the full extent of differences in nucleotide diversity. In a region of 800 kb containing ~12 genes, one gene (*CLSPN*) has been completely analysed and showed one allele fixed in Europeans whereas the other allele is fixed in Chinese, indicating nonneutral evolution. Chris Spencer (University of Oxford) finished the session with a discussion of the evidence for and against neutral evolution. The detection of neutral evolution depends on a number of factors, including the strength of selection, the local recombination landscape, study design and the test statistic used. The speaker used simulations and empirical data to illustrate why caution must be taken when rejecting neutral evolution. By pairing genic with nongenic SNPs on human chromosome 20, he concluded that recent adaptive evolution is either rare, does not predominately occur near known exons or is hidden in the diversity of genetic variation generated by neutral evolution.

Association studies

Inês Barroso (Wellcome Trust Sanger Institute) presented preliminary data on genetic studies of type 2 diabetes (T2D) and obesity. In the past years, the Sanger Metabolic Disease Group, as part of the GEM (Genes for Energy Balance and Metabolism) Consortium, has employed two approaches to the identification of genetic variants involved in T2D. For re-sequence analysis, candidate genes have been analysed in cohorts with extreme phenotypes (>1,200 samples from a morbid childhood obesity cohort or ~200 samples from a cohort of extreme insulin-resistant subjects). So far, re-sequence analysis of ~100 genes identified new mutations in genes such as *LEPR*. More recently they started to perform linkage-disequilibrium mapping on 20q as well as a genome-wide association study. Previous linkage analyses

have identified that the 20q13 region is involved in T2D and, in particular, polymorphisms in the promoter of *HNF4A* have been suggested to play a role. Using 4,608 SNPs from 10 Mb of 20q and more than 2,500 case/control samples ~11.6 million genotypes were generated. After clean up, data for ~4,000 SNPs remained. A total of 768 SNPs with nominal significant association have been identified and are being genotyped across 2,700 new samples in a replication study. Alison Dunning (University of Cambridge) discussed the strategy to search for low-penetrance breast cancer genes, using a large population-based study in East Anglia, UK. Mutations in *BRCA1*, *BRCA2*, *TP53* and *ATM* account for less than 5% of total breast cancer cases, a common disease. In order to identify low-penetrance, high-frequency alleles, a large East Anglian population-based case/control study (with 2,300 cases and 2,300 controls) has been set up. From pathways possibly involved in breast cancer, 521 tagged SNPs in 101 genes have been selected and genotyped; 38 genes were found to be associated with breast cancer susceptibility. Using a larger sample set, only four SNPs appear to have been confirmed. A Breast Cancer Consortium has been organised to address the central issue of replication of findings emanating from studies with small or adequate power. In parallel, a strategy using dense whole genome SNP scans has been initiated with Perlegen Sciences. In the first stage, more than 266,000 SNPs were genotyped across 400 cases and 400 controls, and approximately 13,000 SNPs showed a possible association, though the vast majority are expected to be false positives. The best 12,000 SNPs will then be genotyped across 4,200 cases and 4,200 controls. SNPs that are still associated with breast cancer after the first replication study will be analysed across more than 30,000 samples collected by the Breast Cancer Consortium. Stephen Chanock (National Cancer Institute) then discussed the use of the candidate gene approach to identify genes involved in cancer. Even the whole genome scan strategy will eventually come back to candidate gene regions to be analysed in detail. He described the work underway by the InterLymph Consortium which analysed SNPs in 13 candidate immune genes in 3,500 cases and 3,500 controls and determined in a pooled analysis that functional variants in *TNF* and *IL10* are associated with the risk of non-Hodgkin's lymphoma (Rothman et al. 2006). The importance of replication strategies was then discussed, especially for low-penetrance, high-frequency alleles in the study of common diseases. Moreover, he reported plans for conducting whole genome scans in prostate and post-menopausal breast cancer with large sequential replication studies, known as cancer genetic markers of susceptibility. He proposed that whole genome scan or microarray gene expression studies are excellent tools for generating new candidate genes for study. Andreas Braun (Sequenom) presented data on a genome-wide SNP association study for genetic modifiers of beta thalassemia/HbE disease. Beta thalassemia/HbE disease is one of the main causes of childhood chronic disease in South East Asia. α -Globin and β -globin gene

clusters are involved but additional modifier genes exist. In order to find these genes, Sequenom performed a whole genome scan with ~110,000 SNPs and DNA pools (made up from affected/nonaffected people). More than 600 SNPs with variation in allelic frequency between the pools have been selected for genotyping the individual patient DNA. Some associated SNPs are located in known QTL regions for haemoglobin levels.

Population genetics

Bruce Weir (North Carolina State University) discussed the issue of genomic heterogeneity as a measure of population structure. Using the whole genome SNP survey generated by Perlegen Sciences (1.5 million genotypes across three populations) and the HapMap project (600,000 SNPs across four populations), he estimated F_{st} as a tool to analyse genetic population structure. This genome-wide F_{st} scan shows heterogeneity between segments within chromosomes but also similarities between the two datasets. Dennis Drayna (National Institute on Deafness and Other Communication Disorders) presented data on the diversity of coding sequence variation in human taste receptor genes across the world. Taste perception in human is mediated by receptors on the surface of taste cells on the tongue that recognise sweet, sour, salty, bitter or umami (glutamate) taste. The classical receptors are seven transmembrane G protein-coupled receptors, which are divided into two classes, the *T1R* and *T2R* families. The *T1R* family has three members (*T1R1*, *T1R2* and *T1R3*) working as dimers. *T1R1/T1R3* recognises umami taste, whereas *T1R2/T1R3* recognises the sweet taste. The *T2R* (bitter receptor) family has 25 members for which most of the ligands are unknown. *T1R1*, *T1R2*, *T1R3* and several *T2R* genes were re-sequenced in 55 individuals from different world populations. K_a/K_s ratio and Tajima's D statistics were used to identify genes under selection. Variation in *T1R* genes suggested that perception of sweet and umami substances has been influenced by selection. In addition, variation in the *T2R38* gene and other *T2R* genes, between tasters and nontasters, is present in all populations but with differences in haplotype frequency between populations; this latter observation suggests local adaptation (possibly matching local plant diversity). Arne Pfeufer (Institute of Human Genetics, Munich) discussed genome-wide LD mapping of the QT interval. Heart disease can be due to an arrhythmia associated with prolonged QT interval. Electrocardiogram can be used to detect patients with long QT interval but a better biomarker is needed. Measurements of the QT interval follow a normal distribution and have a moderate heritability. In order to identify markers for long QT interval, a genome-wide LD mapping approach was initiated. More than 2,000 women were studied and approximately 100 individuals for each extreme (high and low QT interval) were genotyped with 100,000 SNPs. In a second stage 150

SNPs were analysed across the most extreme 600 women (300/300). The most strongly associated SNP was identified in the 5' region of the *CAPON* gene (a regulator of *nNOS*). In a final stage ~4,000 individuals were genotyped with 14 SNPs in the vicinity of the *CAPON* gene. Two haplotypes (with seven SNPs) were identified, with the rare haplotype significantly associated with long QT interval.

Genotype and haplotype analysis

Lon Cardon (Wellcome Trust Centre for Human Genetics, Oxford) discussed testing for interactions in genome-wide association scans. In anticipation of whole genome scans being applied to many different diseases, he discussed the opportunity to test beyond the main effect of each variant. It is likely that sets of genes interact and modify effects but the computational requirements are enormous. For example, in a project expected to analyse 500,000 SNPs, $\sim 1.25 \times 10^{11}$ tests will be required to analyse two-way interactions. Is it worthwhile (and possible) to search for interactions? Three strategies exist: a locus by locus strategy, all interactions strategy and a two-stage strategy (all single tests followed by a full interaction analysis for associated loci identified during the first phase). Dr. Cardon fitted different interaction models and showed that the optimal approach is to test all possible interactions, but this strategy brings computational challenges that need to be overcome.

The session ended with three talks on new genotyping technologies. Ivo Gut (Centre National de Genotypage, Evry) presented a micro-haplotyping-based method for HLA typing. Current hybridisation-based HLA typing methods are relatively expensive and difficult to analyse in high throughput settings. A relatively inexpensive alternative was proposed which uses pools of oligonucleotides overlapping different haplotypes for primer extension. Chatarina Larsson (University of Uppsala) described a method for genotyping single DNA molecules in situ. Using padlock probes and rolling circle amplification, she showed that it was possible to detect mitochondrial SNPs directly on tissue sections. This technique is being applied for transcript genotyping. Pui-Yan Kwok (University of California, San Francisco) presented two new approaches for molecular haplotyping. The first method barcodes long individual DNA molecules (currently long PCR products) with multiple padlock allele-specific probes. After specific labelling, the fluorescent barcode is read using a flow cell system measuring fragment size. The current resolution is poor (with less than 10 kb resolution). The second novel method combines optical mapping (imaging of

stretched single molecule) with allele-specific labelling. It appears that the system can be expanded to include five colours (for nucleotide labelling and DNA backbone labelling) in an automated process.

Conclusion

The conference was well attended and offered an opportunity for senior, junior and in-training investigators to share results and develop new collaborations. Though the majority of the discussion centred on human SNPs and CNVs, there were many discussions of SNPs in not only mouse, dog and *Arabidopsis* but also a few 'exotic' species such as maize, Pacific oyster and pig, a testament to the pervasive value of analysing SNPs. A major theme of the conference was the emerging significance of copy number variation in the genome. It appears that we have only scratched the tip of the iceberg but clearly extensive work, including database annotation and new analytical approaches, is needed to effectively use CNVs to map complex human diseases. In contrast, the immediate utility of the International HapMap was evident throughout the meeting. With the development of new genotyping tools, scanning across the entire genome with a dense map of SNPs has now begun for many diseases. The outcome of this is very much anticipated for viewing at the autumn 2006 meeting in this series—renamed Human Genome Variation 2006 (HGV2006) and due to be held in Hong Kong.

References

- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) International HapMap Consortium. A haplotype map of the human genome. *Nature* 437:1299–1320
- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res* 15:1553–1565
- Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307:1072–1079
- Rothman N, Skibola CF, Wang SS, Morgan G, Lan Q, Smith MT, Spinelli JJ, Willet E, De Sanjose S, Cocco P, Berndt SI, Brennan P, Brooks-Wilson A, Wacholder S, Becker N, Hartge P, Zheng T, Roman E, Holly EA, Boffetta P, Armstrong B, Cozn W, Linet M, Xavier Bosch F, Ennas MG, Holford TR, Gallagher RP, Rollinson S, Bracci PM, Cerhan JR, Whitby D, Moore PS, Leaderer B, Lai A, Spink C, Davis S, Bosch R, Scarpa A, Zhang Y, Severson RK, Yeager-Jeffery M, Chanock S, Nieters A (2006) Genetic variation in *TNF* and *IL10* and risk of non-Hodgkin lymphoma: a report from the interlymph consortium. *Lancet Oncol* 7:27–38