

Editorial

In 1980, the seminal article by Botstein *et al.* laid out a rosy future for mapping genes underlying the susceptibility and/or resistance to inheritable diseases.¹ It led to a rush to identify polymorphic markers in the human genome, although only a handful of biallelic nucleotide polymorphisms were discovered due to technical difficulty. Since 1989, microsatellites have been the markers of choice, as they are much more informative. At the Annual Meeting for the American Society of Human Genetics in 1995, the introduction of a new technology for mutation detection—denaturing high-performance liquid chromatography (DHPLC)—stirred up some excitement, and the identification of a large number of single nucleotide polymorphisms was suddenly within reach. The first systematic effort to identify single nucleotide polymorphisms, known as SNPs by then, was achieved by Lander's group from the Whitehead Institute, using brute force sequencing rather than DHPLC.² As of early 2004, dbSNP (build 119) at the National Center for Biotechnology Information hosts over 7 million SNPs, with 3.4 million of them validated. It should not be long before SNPs replace microsatellites in many applications, if not altogether, in human genetics.

In 2001, a groundbreaking discovery of the block-like structure of linkage disequilibrium (LD) in the human genome led to the illumination of the genomic structure of SNPs.^{3–7} Within each block, extremely high LD was evident, together with very little trace of recombination. More importantly, within each block, the number of haplotypes (ie, a combination of variants in the block) was generally few. The significance of this discovery is that only a few SNPs are needed to represent an entire block. This significantly reduces the genotyping load, one of the primary obstacles that have hindered genome-wide association studies. It is, therefore, not surprising that there has been a surge in the number of studies on haplotypes and haplotype blocks of SNPs, as evident in this issue of *Human Genomics*.

To characterise SNP haplotype structure, Clark and Dean compare the haplotypes determined by pedigrees and haplotypes estimated using the EM algorithm in two segments of the human genome: a 150 kb region containing four chemokine receptor genes on 3p21 and a region containing six chemokine genes on 17q11–12. A nearly perfect concordance between the two estimations was observed for the 3p21 region, while, conversely, the two haplotype estimations for the 17q11–12 region were less consistent with each

other. These results suggest that, while estimations of haplotype frequency and LD may be relatively simple in some genomic regions with population samples, a higher resolution haplotype analysis is required for the others—such as chromosome 17q11–12—which are characterised by more complex environment.

It has been recognised that bias in ascertaining the SNPs in the human genome may have a significant impact on the subsequent population genetic analysis. In this issue of *Human Genomics*, Nielsen takes a very careful look into such ascertainment bias in frequency spectrum, inferences of demographic parameters and LD. Several recently developed methods for correcting for the ascertainment bias are also discussed. Furthermore, the inference of haplotype block structure will also be affected by the sample size and the SNPs selected when only a subset of variations is used initially. Sun *et al.* conduct a detailed empirical study to examine such an impact by analysing three representative autosomal regions from a large genome-wide study of haplotypes. The results of this study raise considerable concerns on the density of the markers initially selected and advocate a relatively large sample size and a very dense marker panel in characterising the haplotype structure in human populations.

SNPs are not equal: some SNPs are more difficult to type than others. Probe and primer design for SNP typing can be challenging for those located in A–T-rich or G–C-rich regions. Belousov *et al.* describe a new approach using the MGB Eclipse™ System to introduce modified bases into the probes and primers. The combination of MGB Eclipse probes and primers enriched with the MGB ligand and modified bases has allowed the analysis of refractory SNPs where other methods have failed.

By the time this issue reaches our readers, the data from the first phase of the HapMap project is expected to be available. Much analysis will be done and will add significantly to our knowledge of the genomic structure of the variations in human populations. *Human Genomics* would like to be the forum for publishing your contributions to this very exciting area of research.

References

1. Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) 'Construction of a genetic linkage map in man using restriction fragment length polymorphisms', *Am. J. Hum. Genet.* Vol. 32, pp. 314–331.

2. Wang, D.G., Fan, J., Siao, C., *et al.* (1998) 'Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome', *Science* Vol. 280, pp. 1077–1082.
3. Daly, M.J., Rioux, J.D. and Schaffner, S.F. (2001) 'High-resolution haplotype structure in human genome', *Nat. Genet.* Vol. 29, pp. 229–232.
4. Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001) 'Intensely punctuate meiotic recombination in the class II region of the major histocompatibility complex', *Nat. Genet.* Vol. 29, pp. 217–222.
5. Rioux, J.D., Daly, M.J., Silverberg, M.S., *et al.* (2001) 'Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn's disease', *Nat. Genet.* Vol. 29, pp. 223–228.
6. Johnson, G.C.L., Esposito, L., Barratt, B.J. *et al.* (2001) 'Haplotype tagging for the identification of common disease genes', *Nat. Genet.* Vol. 29, pp. 233–237.
7. Goldstein, D.B. (2001) 'Islands of linkage disequilibrium', *Nat. Genet.* Vol. 29, pp. 109–111.

Li Jin
Managing Editor
Human Genomics

A supplementary paper by Jianying Gu and Xun Gu is available online — see page 239 for how to access the online version of the Journal.