Microsatellite markers for genome-wide association studies

The recent Opinion article by Jorgenson and Witte[1] revisits genome-wide association (GWA) studies, emphasizing the need to focus on intragenic variants. However, their postulate is based on a partial view of the available resources and expertise for GWA studies; they do not mention that, besides SNPs, there is another well-known, but considerably less widely exploited resource of polymorphic variants — microsatellites.

Previous articles in your journal[2], and primary papers elsewhere[3], have acknowledged the use of microsatellites for GWA studies. However, they also pointed out the clear limitations to this approach that existed at the time of their publication, which mainly involved inadequate numbers of informative microsatellites. This hurdle has since been overcome. The human genome is now charted by approximately 30,000 highly polymorphic microsatellites, and this has led to two published GWA studies[4,5] and the description of a general framework for using this 30K set in such studies[4].

The rationale behind the 30K number is that the average length of linkage disequilibrium (LD) for microsatellites is ~100 kb, which is considerably higher than that of SNPs. Therefore, a single microsatellite captures a larger genomic region than does a single SNP. Microsatellites also provide several other advantages, such as a higher information content (6–10 alleles as compared with 2 alleles for SNPs), and a smaller interpopulation variability[6,7].

Furthermore, with regard to SNP haplotypes, it is important to remember that these are virtual products, based on statistical hypotheses that depend on several conditions (for example, $r^2 > 0.8$, minor allele frequency (MAF) > 0.1, >90% capture). In addition, in certain cases recombination events cannot be included in these haplotypes, due to the lack of a heterozygous SNP at the relevant position. This results in the generation of incorrect haplotype blocks. Finally, the definition of a SNP haplotype block can vary, depending on the 'deepness' of the sequencing effort.

Because of these factors, a considerable number of susceptibility loci can be missed in GWA studies that are based on SNPs only (whether genic or non-genic), and others might be wrongly identified and therefore not replicable, as has been documented elsewhere[8]. Therefore, in order to conduct reliable GWA studies using SNP haplotypes, the total number of tested haplotypes might reach that of the individual SNPs themselves. By contrast, microsatellites do not suffer from these problems, and could be used as a stand-alone source of markers in GWA studies (see above and REF. 9). Furthermore, microsatellites have important intrinsic functional relevance[10].

Returning to the core proposal of Jorgenson and Witte, using microsatellite LD structures will overcome the need to specifically genotype coding SNPs (cSNPs) as, given the arguments that are presented above, the odds are likely to be much higher that these causal SNPs will be picked up by surrounding microsatellites than by SNP haplotypes. Indeed, the ability of microsatellite markers to capture information on cSNPs has been tested in what could arguably be considered the most extreme case — the human major histocompatibility complex (MHC, also called HLA), which in humans contains well over 2,000 HLA alleles in a segment of <4 Mb[11]. Even in this setting, microsatellites have been used successfully to define haplotypes that encompass and reliably predict HLA alleles (that is, cSNPs)[12,13]. These same microsatellites have been used to successfully map causal cSNPs in GWAs within this 'sea of polymorphism.'[14,15]

Seiamak Bahram and *Hidetoshi Inoko*

Human Molecular Immunogenetics, Centre de Recherche dImmunologie et d Hématologie, 4 rue Kirschleger, 67085 Strasbourg Cedex, France.

Department of Genetic Information, Division of Molecular Life Science, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan.

Correspondence to H.I. e-mail: hinoko@is.icc.u-tokai.ac.jp

1. Jorgenson, E. & Witte, J. S. A gene-centric approach to genome-wide association studies. *Nature Rev. Genet.* **7**, 885–891 (2006).
2. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).

3.   Bailey-Wilson, J. E. *et al.* Genetic Analysis Workshop 14: microsatellite and single-nucleotide polymorphism marker loci for genome-wide scans. *BMC Genet.* **6**, S1 (2005).

4.   Tamiya, G. *et al.* Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Hum. Mol. Genet.* **14**, 2305–2321 (2005).

5.   Kawashima, M. *et al.* Genomewide association analysis of human narcolepsy and a new resistance gene. *Am. J. Hum. Genet.* **79**, 252–263 (2006).

6.   Jorde, L. B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).

7.   Sawyer, S. L. *et al.* Linkage disequilibrium patterns vary substantially among populations. *Eur. J. Hum. Genet.* **13**, 677–686 (2005).

8.   Myers, R. Considerations for genomewide association studies in Parkinson disease. *Am. J. Hum. Genet.* **78**, 1081–1082 (2006).

9.   Xiong, M. & Jin, L. Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. *Am. J. Hum. Genet.* **64**, 629–640 (1999).

10.   Kashi, Y. & King, D. G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253–9 (2006).

11.   Schreuder, G. M. *et al.* The HLA Dictionary 2004: a summary of HLA-A, -B, -C, -DRB1/3/4/5 and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and -DQ antigens. *Tissue Antigens* **65**, 1–55 (2005).

12.   Malkki, M., Single, R., Carrington, M., Thomson, G. & Petersdorf, E. MHC microsatellite diversity and linkage disequilibrium among common *HLA-A*, *HLA-B*, *DRB1* haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies. *Tissue Antigens* **66**, 114–124 (2005).

13.   Foissac, A. *et al.* Microsatellites in the HLA region: HLA prediction and strategies for bone marrow donor registries. *Transplant Proc.* **33**, 491–492 (2001).

14.   Ota, M. *et al.* A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. *Genomics* **71**, 263–270 (2001).

15.   Okamoto, K. *et al.* Identification of IκBL as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis. *Am. J. Hum. Genet.* **72**, 303–312 (2003).