Reply: Microsatellite markers for genome-wide association studies

The widespread use of microsatellite markers in genome-wide association (GWA) studies, as supported by Bahram and Inoko, currently faces several obstacles. These include determining genome-wide coverage, developing high-throughput genotyping platforms and analysis issues.

Coverage has been well studied for SNP sets because it determines the power of a GWA study[1]. Bahram and Inoko state that "…the average length of linkage disequilibrium (LD) for microsatellites is ~100 kb." This does not adequately address the coverage issue for microsatellite markers. First, the ~30,000 available microsatellites are unlikely to be evenly distributed across the genome; therefore, simply by chance, some areas will not be covered. Second, LD varies across the genome, necessitating additional markers in low LD regions. Third, the measure of LD that is used by Bahram and Inoko in their estimate of 30,000 markers, its expected magnitude, and the population in which it was measured are unclear; all are needed to determine coverage and power. Fourth, it is unknown how well LD between microsatellites reflects LD among SNPs. Microsatellites have considerably higher mutation rates compared with SNPs, and could therefore represent younger alleles that might not capture older SNP alleles. To determine the ability of the ~30,000 available microsatellite markers to comprehensively cover the genome, they should be genotyped on the existing HapMap samples.

Another concern is the availability of high-throughput genotyping platforms. Bahram and Inoko cite two studies that have used more than 20,000 microsatellite markers to conduct association testing, both of which used pooled DNA to facilitate the genotyping of such a large number of markers. Pooling reduces the effort and cost of genotyping, but can introduce additional errors in estimating allele frequencies. This approach reduces statistical power and limits the size of detectable allele-frequency differences[2]. Many SNP-based GWA studies have instead decreased genotyping costs through a multistage approach[3,4]. Cheaper and more reliable high-throughput microsatellite genotyping would be required to make such a multistage approach equally attractive for microsatellite markers. Furthermore, microsatellite markers are not immune to interference from copy-number polymorphisms[5], a problem that is only now being addressed on large-scale SNP genotyping platforms.

Bahram and Inoko correctly point out that testing SNP haplotypes in GWA studies has several drawbacks, including the incorrect assignment of haplotype blocks. Multimarker tests can be conducted without defining blocks, and the first published GWAs reported the results of SNPs that were tested individually[3,6,7]. Although there is a significant multiple-testing burden when testing hundreds of thousands of SNPs or their haplotypes, testing each of the 6–10 alleles of 30,000 microsatellites will require large numbers of both tests and degrees of freedom for each test. Assuming that a set of microsatellite markers provides the same level of coverage as a dense SNP set, gains in statistical power or reductions in sample size are likely to be limited[8].

Finally, extended LD complicates the analysis and fine-scale localization of an association signal[9–12]. A study that utilizes a sparse set of microsatellite markers instead of SNPs in the initial genotyping phase will require a larger number of SNPs to be genotyped for fine-scale mapping compared with studies that use a dense SNP platform.

Eric Jorgenson and John S. Witte

Department of Epidemiology & Biostatistics and Center for Human Genetics, University of California, San Francisco CA 94143-0794, USA.

Correspondence to E.J. e-mail: Eric.Jorgenson@ucsf.edu

1. Jorgenson, E. & Witte, J. S. Coverage and Power in Genomewide Association Studies. *Am. J. Hum. Genet.* **78**, 884–889 (2006).

2. Carlson, C. S., Eberle, M. A., Kruglyak, L. & Nickerson, D. A. Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446–452 (2004).
3. Maraganore, D. M. *et al.* High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77**, 685–693 (2005).
4. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genet.* **38**, 209–213 (2006).
5. Jorgenson, E. *et al.* Ethnicity and human genetic linkage maps. *Am. J. Hum. Genet.* **76**, 276–290 (2005).
6. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
7. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genet.* **32**, 650–654 (2002).
8. Witte, J. S., Elston, R. C. & Cardon, L. R. On the relative sample size required for multiple comparisons. *Stat. Med.* **19**, 369–372 (2000).
9. Hughes, A. E. *et al.* A common CFH haplotype, with deletion of *CFHR1* and *CFHR3*, is associated with lower risk of age-related macular degeneration. *Nature Genet.* **38**, 1173–1177 (2006).
10. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103**, 14068–14073 (2006).
11. Li, M. *et al.* CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nature Genet.* **38**, 1049–1054 (2006).
12. Maller, J. *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genet.* **38**, 1055–1059 (2006).