

Medical applications of haplotype-based SNP maps: learning to walk before we run

Ever since Risch and Merikangas¹ introduced the concept of whole-genome association using single-nucleotide polymorphisms (SNPs) in 1996, the question of how many SNPs would be required for whole-genome association studies has been extensively debated^{2,3}. The initial prediction of one million SNPs has recently been reduced to approximately 300,000 (ref. 4). Furthermore, it is clear that chromosomally mapped and ordered SNP variants can be grouped into bins/blocks as distinct 'haplotypes'^{5,6}. Recently, the National Human Genome Research Institute proposed the development of a map of the common haplotype patterns in at least three ethnic populations.

The scientific community is currently divided regarding the perceived need for genome-wide maps of common haplotype blocks. Most support the proposal of constructing haplotype maps to maximize information content and to minimize the number of SNPs required for whole-genome genotyping. The scope of its utility is, however, bounded. Its applicability may not extend beyond the ethnic groups chosen, and is limited by the representativeness of the few samples studied. Furthermore, rare and sometimes detrimental variants are the ones most likely to be missed entirely from the derived SNP set. Pharmacogenetics focuses on the prediction of safety and efficacy following pharmaceutical treatment and clearly stands to benefit from the increased efficiency of a complete haplotype map. A practical question is whether common haplotype maps for several ethnic populations are essential, or even necessary, for immediate pharmacogenetic applications to advance.

The short history of SNP mapping reveals a wide gap between theoretical estimates based on mathematical models and actual experimental results. Estimation of study design features using assumed mathematical models may be viewed, by some, as a purely academic issue, but for pharmacogenetics, there are potentially huge medical and economical implications. Many existing molecules have been found to treat unmet medical needs quite effectively, but in clinical trials or post-marketing surveillance are associated with serious, or too frequent, adverse

events. The central point is that pharmacogenetics can contribute to less expensive clinical trials and that pharmacogenetic post-marketing surveillance can result in the continued availability of effective medicines by identifying individuals at highest risk for adverse events.

In 1999, a proof of principle experiment for the application of SNP mapping to pharmacogenetics was launched using abacavir, a marketed drug that is highly effective for treating HIV infection. Approximately 5% of abacavir-treated patients experience a hypersensitivity reaction (HSR), usually in the first six weeks of treatment.

One hundred fourteen polymorphisms from twelve candidate gene families were selected for analysis. The study identified two genes with alleles that were highly significantly associated with HSR. They were *TNF* (encoding TNF α) and *HLA-B*, both of which fall in the HLA region of chromosome 6. Hetherington *et al.*⁷ have reported the clinical aspects of these data, and an independent confirmation implicating the same genomic region was reported by Mallal *et al.*⁸. The fact that both genes are involved with immune response is consistent with the symptoms of the adverse event. However, the fact that they were approximately 200 kilobases apart on the same chromosome opened the question of an underlying haplotype effect and the possibility of neither one being the true causal variant⁹.

To determine whether the risk-associated alleles were in linkage disequilibrium with each other or contributing independently to risk of HSR, a closer analysis of individual patients' data was undertaken. Of the 25 patients with at least one copy of TNF α -238A, 92% (23) were HLA-B57-positive, and 82% (23 of 28) HLA-B57-positive cases were also TNF α -238A carriers. Thus, there seems to be considerable allelic association between the two markers ($D' > 0.95$), providing support for the principle of using haplotype blocks for association analyses.

To evaluate the haplotype structure of the region, 24 additional SNPs located between HLA-B57 and TNF α -238A were genotyped. This analysis revealed that although HLA-B57 and TNF α -238A were

in high linkage disequilibrium with each other, the intervening region between the SNPs was divided into approximately 3 haplotype bin/blocks. There are three practical lessons from this study. First, extended linkage disequilibrium (beyond the confines of a block) can be detected readily in pharmacogenetic association experiments sometimes, across a large region and several haplotype blocks. Association is not necessarily limited to small regions, suggesting that causality cannot always be inferred from association results that define a small chromosomal segment. Second, because the frequencies of most adverse events are low, the use of common haplotypes may overlook important genetic associations. Third, although complex disease theory suggests that large numbers of patients are required for gene discovery, the abacavir example is of sufficient magnitude to be detectable with modest outlay.

Examining several ethnically defined populations may be scientifically interesting; but the establishment of ethnically defined haplotype maps may not be necessary in pharmacogenetic practice. In fact, for clinically important pharmacogenetic analyses, haplotype-based maps consisting of only common haplotypes may be insufficient or unnecessary to detect adverse event or efficacy association regions. The increased knowledge generated by the haplotype map will be a welcome stride forward, far surpassing any structural estimates derived by mathematical simulation. For applying pharmacogenetics in the short term, it will provide not a panacea, but rather a valued addition to the analysis toolkit. For pharmacogenetic applications that do not select ethnic groups, it might be more logical to inform the debate with empirical data from population experiments involving real adverse events.

Eric Lai, Clive Bowman, Aruna Bansal, Arlene Hughes, Michael Mosteller & Allen D. Roses

Genetics Research, GlaxoSmithKline, Research Triangle Park, North Carolina 27709, USA. Correspondence should be addressed to A.D.R. (e-mail: allen.d.roses@gsk.com).

1. Risch, N. & Merikangas, K. *Science* **273**, 1516–1517 (1996).
2. Kruglyak, L. *Nat. Genet.* **22**, 139–144 (1999).
3. Collins, A., Lonjou, C. & Morton, N.E. *Proc. Natl. Acad. Sci. USA* **96**, 15173–15177 (1999).
4. Patil, N. *et al. Science* **294**, 1719–1723 (2001).
5. Johnson, G. *et al. Nat. Genet.* **29**, 233–237 (2001).
6. Daly, M. *et al. Nat. Genet.* **29**, 229–232 (2001).
7. Hetherington, S. *et al. Lancet* **359**, 1121–1122 (2002).
8. Mallal, S. *et al. Lancet* **359**, 727–732 (2002).
9. Roses, A.D. *Nature Rev. Drug Disc.* **1**, 541–549 (2002).