PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2005) **360**, 1605–1607
doi:10.1098/rstb.2005.1691
Published online 25 July 2005

Will admixture mapping work to find disease genes?

David Reich[1,2,*] **and Nick Patterson**[1,2]

[1]*Department of Genetics, Harvard Medical School, Boston, MA, USA*
[2]*The Broad Institute of Harvard and MIT, Cambridge, MA, USA*

Admixture mapping is the first experimentally practical method for carrying out a whole-genome association scan, and is thus a promising method for detecting risk factors for common disease. The goal of the community should now be to aggressively test whether the method is useful in practice for localizing disease genes, by carrying out at least three high-powered studies. We also propose a stringent criteria we believe the community should adopt before declaring a statistically significant admixture association to disease.

Keywords: admixture mapping; genetics; ancestry

1. ADMIXTURE MAPPING: A PROMISING NEW DISEASE GENE MAPPING METHOD

At present there are two approaches to interrogating the genome for disease genes: linkage and association. Linkage mapping has been remarkably successful at finding genes for rare Mendelian diseases. Whole-genome association has much more power in principle to discover the genes of weak effect that may contribute to common disease. However, so far it has eluded researchers because the most commonly discussed strategy requires studying 300 000–1 000 000 markers, which would be prohibitively expensive.

'Admixture mapping' is an approach to whole-genome association mapping that is practical today because it requires 200- to 500-fold fewer markers. The idea is that people with disease who happen to descend from the recent mixing of ethnic groups should have an increased probability of inheriting DNA from the ancestral populations with higher disease risk. For example, African Americans (20% European ancestry on average), might tend to have an unusually high level of European ancestry near a multiple sclerosis gene, since the disease is one that occurs more often in people of European ancestry. There has been progress in developing experimental resources and methodologies for admixture mapping in the last few years. In other publications, we and others have described the following:

(i) We developed the first practical resource for admixture mapping in African Americans: 2154 genetic markers validated in our laboratory as being very different in frequency in west African as compared with European Americans (Smith *et al.* 2004).

(ii) We and others have developed methodologies that use a Hidden Markov Model nested within a Markov Chain Monte Carlo to analyse data from high density maps to find disease genes (Hoggart *et al.* 2004; Patterson *et al.* 2004; Montana & Pritchard 2004).

(iii) We applied these resources to 992 patients with prostate cancer and 748 with multiple sclerosis. The MS scan identifies an excellent candidate for disease risk, showing that the method may be a practical way to map genes. The prostate cancer scan is negative, but follow-up in more samples and a higher density of markers may reveal risk genes. These results are currently in preparation for publication elsewhere.

(iv) Zhu and colleagues (2005) published a whole-genome admixture scan of 737 individuals with hypertension and 573 controls, studied at 269 microsatellite markers. They found evidence for association on chromosomes 6q24 and 21q21, the first suggestion that admixture mapping may be identifying new risk factors for complex disease.

The technical details of admixture mapping are reviewed elsewhere (Halder & Shriver 2004; Patterson *et al.* 2004; Smith *et al.* 2004; McKeigue 2005). Here we provide an opinionated view about the future of admixture mapping:

(i) We forecast what will be necessary to provide convincing evidence about whether or not it is a practical way to find disease genes, and suggest that the community should focus on assessing this over the next few years.

(ii) We suggest a rigorous standard for declaring a statistically significant association by admixture mapping. The published paper on hypertension by Zhu *et al.* (2005) does not meet this standard, although their result is intriguing.

* Author for correspondence (reich@genetics.med.harvard.edu).

One contribution of 12 to a Discussion Meeting Issue 'Genetic variation and human health'.

Box 1 Criteria for declaring a genome-wide significant admixture association to disease

(i) The Bayesian statistic for detecting genome-wide significant association, suggested by Patterson *et al.* (2004), should be greater than 2 (or a similar criterion for the methods of Hoggart *et al.* 2004 or Montana & Pritchard 2004).
(ii) The deviation of European ancestry compared to the genome-average should be seen in cases only, and not controls.
(iii) The signal should remain when the marker contributing most strongly to disease is removed.
(iv) Markers in linkage disequilibrium with each other in the ancestral European and west African populations should be rigorously excluded from the scanning set.
(v) The region of association should appear statistically significant based on two different Markov Chain Monte Carlo analysis software packages.
(vi) The *P*-values for case–control association should be obtained by carrying out permutation testing, making sure that the statistic at the disease locus is more extreme than anything seen anywhere in the genome in 100 random permutations of the case and control labels.
(vii) The statistic for association should increase in significance when marker density at the locus is increased or more affected samples are added to the study.

2. HOW WILL WE KNOW WHETHER ADMIXTURE MAPPING IS USEFUL FOR FINDING DISEASE GENES?

Admixture mapping is an intellectually attractive approach to finding disease genes, but is still unproven. So far, most published papers on admixture mapping have focused on building resources and methodologies. The question of whether admixture mapping is likely to be useful in fact, however, can only be answered by carrying out real studies.

Proof that admixture mapping is viable will only come when it results in the identification of a previously unknown section of the genome modulating disease risk. Neither the admixture associations with hypertension published by Zhu and colleagues (2005), nor our own unpublished results on multiple sclerosis, nor any other study of which we are aware, have yet resulted in the actual fine mapping of new disease genes. Thus, it should be a high priority for researchers interested in admixture mapping to pursue such associations.

The alternative possibility is that admixture mapping might not be useful despite its theoretical attractiveness: there may be too few disease alleles that are differentiated enough in frequency across populations to be identifiable. We recently surveyed 14 previously published associations to obtain an empirical assessment of the fraction of common variants causing common disease that would be identifiable by the method. Only 29% of these published associations would have been detectable by admixture mapping with 2000 samples or fewer, while 43% would have required greater than 15 000 samples (Patterson *et al.* 2004). It is often suggested that admixture mapping focuses on those diseases that are known to be very different in prevalence between African and European Americans—e.g. prostate cancer, multiple sclerosis, lupus, hypertension and Type II diabetes—to optimize the chance of there being highly differentiated alleles that will be accessible to the method. However, highly differentiated alleles that are accessible to admixture mapping are entirely consistent with diseases that are similar in prevalence across populations, and so if admixture mapping works, it may work well for these diseases as well (Patterson *et al.* 2004; McKeigue 2005).

At present we are agnostic about whether admixture mapping will be a useful disease gene mapping method. The major effort of the community in the next few years should thus be to move away from building resources (which have developed greatly in the past few years and now make admixture mapping practical), and to formally test whether this promising method works.

We propose that a proper test of whether or not admixture mapping works will only require at least three very high-powered studies, each with more than 2000 samples and scanned in a map of markers as informative about African versus European ancestry as the one we recently published (Smith *et al.* 2004). If admixture mapping fails for several diseases *a priori* known to have high prevalence differences between populations, the method is not likely to be a practical way for localizing disease genes.

3. STANDARDS FOR DECLARING A SIGNIFICANT ASSOCIATION TO DISEASE BY ADMIXTURE MAPPING

We now have extensive experience applying admixture mapping to real disease studies (Patterson 2004; Reich & Patterson, unpublished data), which has taught us that there are many ways to obtain false-positive associations using these methods. Indeed, in a large proportion of the real data sets we studied, we initially obtained genome-wide significant evidence of association, which disappeared upon closer examination of the data. Given the complexity of the methods that need to be used to analyse admixture mapping data sets, we believe it is important for a signal to meet several stringent criteria (presented in Box 1) before it can be published as a genome-wide significant association to disease. A high standard of proof should be a criterion for publishing these studies, especially in the early days before we know that they are a reliable way to localize real disease risk genes.

4. CONCLUSION

Admixture mapping is the first experimentally practical method for carrying out a whole-genome association scan, and is thus a promising method for detecting risk factors for common disease. The goal of the community should now be to aggressively test whether the method is useful in practice for localizing disease genes,

by carrying out at least three high-powered studies. We also propose stringent criteria we believe the community should adopt before declaring a statistically significant admixture association to disease.

REFERENCES

Halder, I. & Shriver, M. D. 2004 Measuring and using admixture to study the genetics of complex diseases. *Hum. Genomics* **1**, 52–62.

Hoggart, C. J., Shriver, M. D., Kittles, R. A., Clayton, D. G. & McKeigue, P. M. 2004 Design and analysis of admixture mapping studies. *Am. J. Hum. Genet.* **74**, 965–978.

McKeigue, P. M. 2005 Prospects for admixture mapping of complex traits. *Am. J. Hum. Genet.* **76**, 1–7.

Montana, G. & Pritchard, J. K. 2004 Statistical tests for admixture mapping with case-control and cases-only data. *Am. J. Hum. Genet.* **75**, 771–789.

Patterson, N. *et al.* 2004 Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000.

Smith, M. W. *et al.* 2004 Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000.

Zhu, X. *et al.* 2005 Admixture mapping for hypertension loci with genome-scan markers. *Nature Genet.* **37**, 177–181.