Hot spots unglued

Andrew G Clark

Single-sperm genotyping has provided evidence for hot spots of locally intense recombination in humans. Previous efforts have shown that statistical methods can identify these hot spots as local deficits of linkage disequilibrium, but a report in this issue shows that some recombination hot spots leave no signature of reduced linkage disequilibrium.

Localized regions of exceptionally high recombination rate in the human genome attracted a great deal of attention when it was suggested that they might punctuate the genome into blocks of high linkage disequilibrium (LD)[1]. If most human recombination were punctate, it could mean that fewer SNPs would be needed for whole-genome association testing, making this approach for gene mapping much more feasible. The existence of local hot spots of recombination is unequivocally supported by single-sperm genotyping (**Fig. 1a**), which shows that recombination events in a small number of regions show sharply defined clustering[2,3]. Indirect inference of local spikes in the 'population recombination rate' (**Fig. 1b**), a parameter that arises in the population genetics theory of LD, also suggested that recombination might be localized[4]. It was especially heartening when these two approaches seemed to converge in identifying the same hot spots[5,6], as it implied that the costly and time-consuming method of single-sperm genotyping, which allows direct testing of meiotic crossover rates, could be replaced by indirect estimates based on SNP genotype counts. On page 601 of this issue, we see the first exception to this pattern, as Jeffreys *et al.*[7] show that there can be true recombination hot spots in regions where LD remains high.

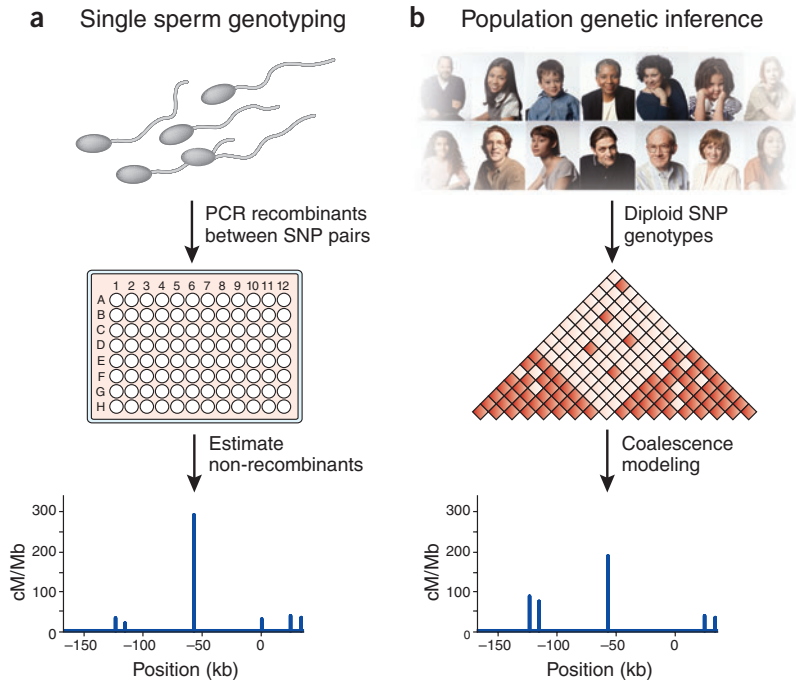

Figure 1 Methods for inferring recombination hot spots. (**a**) Sperm genotyping. By diluting sperm to the level of single cells, whole-genome amplification will yield products from single haploid cells. These single-sperm amplicons can then by genotyped by a variety of PCR-based methods, allowing one to score recombination rates between each pair of SNPs by directly counting recombinants. (**b**) Population recombination rate can be estimated indirectly from the pattern of LD derived from genotypes of diploid individuals. Statistical methods infer estimates of $\rho = 4Nr$ that are most compatible with the data.

Sperm genotyping

Single-sperm genotyping can yield recombination rate estimates at very fine scales because it allows for sample sizes large enough to detect rare events reliably[2,3]. But the high expense and effort of genotyping single sperm to infer recombination at short intervals has resulted in only a few regions being studied. This is unfortunate because the sperm typing data provides the only unambiguous and direct scoring of recombinants at the fine scale needed to identify hot spots. Sperm typing has also shown that there is extensive variation in the location and intensity of hot spots[8]. The issue of variability is especially important because it could shed light on the mechanism for hot-spot formation. In addition, such variability could contribute to population-specific differences in LD[9].

At present, we have not defined the extent of population variability in recombination hot-spot location or intensity as assessed by sperm typing. A technology that allows large-

Andrew Clark is in the Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.
e-mail: ac347@cornell.edu

scale sperm genotyping across many genomic positions in multiple individuals from several population groups would help to clarify several issues about the stability and generality of recombination hot spots. But even with such a technology, single-sperm genotyping has the shortcomings of not accounting for female recombination and not being able to identify hot spots on the nonpseudoautosomal portions of the X chromosome. Sperm typing also involves a bit of statistical inference, as only recombinants are scored by PCR and the nonrecombinant count must be estimated. Even more fundamentally, if recombination rates vary among individuals and over time, then an instantaneous snapshot of recombination from a few individuals will fail to capture the past historical landscape of recombination.

Recombination and LD

A population in drift-recombination balance for neutral polymorphism will have an inverse relation between the steady state level of LD and the rate of recombination[10]. This theory shows that the signature of recombination in population sample data is always confounded with the population size (because random drift generates LD). In fact, many factors besides recombination will impact the degree of LD between SNPs, including mutation, selection, migration, bottlenecks and random drift. But at a megabase scale, there is very good correspondence between the inferred rate of population recombination and local intensity of meiotic recombination (in cM Mb^{-1}), presumably because these other factors average out[5]. Rate estimation is based on an approximate method known as composite likelihood[11], which seems to work unexpectedly well given its underlying assumption of independence across site-pairs. At a finer scale, hot spots of recombination have been inferred by contrasting the likelihood of the data under a neutral coalescent model assuming either homogeneous recombination or the presence of a hot spot[4–6,12]. The statistical methods have mostly been tested by their ability to identify hot spots introduced artificially in simulated data. Thus, apart from the TAP2 recombination hot spot[3], the new work of Jeffreys *et al.*[7] provides the first real test of statistical hot-spot detection.

By scoring 200 SNPs across a 206-kb region of chromosome 1 in 80 European men, Jeffreys *et al.*[7] identified seven hot spots by statistical inference. They then tested these hot spots by single-sperm genotyping in the men whose SNP genotypes allowed it and confirmed that all seven had elevated recombination. But sperm typing identified an additional recombination hot spot that showed no statistical signature. This is a small sample from which to infer generalities about rates of false positive and false negative errors, but the one exception is enough to make one wonder why LD could be so high across a recombination hot spot.

The suggestion that hot spots may be transient raises the question of whether hot spots are stable in comparisons of closely related species. Recent analyses of polymorphisms in chimpanzees have shown that LD-based inferences of hot spots in humans and chimpanzees are widely divergent[13,14]. One might be tempted to question the statistical inference of the hot spots, as single-sperm typing has not been done in chimpanzees. But even the sharpest critic would agree that the LD landscapes of humans and chimps differ markedly. The data suggest that hot spots are rather transitory in both genomes and that despite the mere 1% divergence in DNA sequence, there are radical differences in the signals that dictate recombination rates. This becomes more plausible when one notes that hot-spot activity also differs widely among yeast strains[15]. Perhaps in this light, it is not surprising to find evidence for variation and transience in hot spots in humans, as suggested by Jeffreys *et al.*[7]

The ideas that recombination hot spots are moving about at such a high rate that there is virtually no shared ancestral positioning between humans and chimps and that they are so fluid that even in humans there is great variability in their position and degree raise a series of questions that are fundamental to our understanding of recombination and the evolution of genomes. At the root of these questions is the mechanism: what determines the position and intensity of a hot spot? Is it heritable? Should we replace our view of the human genetic map with a model more akin to statistical mechanics, where in each interval there is a probability cloud for recombination hot spots, and each individual is an epigenetic realization of a draw from this probability density? If so, how does this impact our use of linkage for association studies? Are differences in patterns of LD across human populations driven more by hot-spot variation or by local demographic differences? Are the hot spots detected by statistical inference simply the ones that have stayed in one position long enough and in enough individuals to have left a signature in eroded LD?

At the coarse level of pedigree analysis, averaging out the wild variability in recombination over many small intervals gives a well-behaved mean rate (just as Newtonian physics is not rendered useless by quantum mechanics). But when we consider using statistically inferred recombination as a tool for association mapping at a finer level, perhaps we need to understand more thoroughly the root causes for fine-scale variation in recombination rates. On the other hand, from the point of view of complex trait mapping, the statistically inferred recombination is of more direct relevance, because it quantifies an average strength of statistical associations across sites. Ultimately, it seems that the greatest gaps in our understanding of hot spots will best be filled by understanding the mechanism behind their formation. And, if the past offers any guidance, experiments with model organisms will continue to pave the way.

1. Jeffreys, A.J., Kauppi, L. & Neumann, R. *Nat. Genet.* **29**, 217–222 (2001).
2. Cui, X.F. *et al. Proc. Natl. Acad. Sci. USA* **86**, 9389–9393 (1989).
3. Jeffreys, A.J., Ritchie, A. & Neumann, R. *Hum. Mol. Genet.* **9**, 725–733 (2000).
4. Li, N. & Stephens, M. *Genetics* **165**, 2213–2233 (2003).
5. McVean, G.A. *et al. Science* **304**, 581–584 (2004).
6. Crawford, D.C. *et al. Nat. Genet.* **36**, 700–706 (2004).
7. Jeffreys, A.J., Neumann, R., Panayi, M. & Donnelly, P. *Nat. Genet.* **37**, 601–606 (2005).
8. Yu, J. *et al. Am. J. Hum. Genet.* **59**, 1186–1192 (1996).
9. Sawyer, S.L. *et al. Eur. J. Hum. Genet.* **13**, 677–686 (2005).
10. Ohta, T. & Kimura, M. *Genetics* **68**, 571–580 (1971).
11. Hudson, R.R. *Genetics* **159**, 1805–1817 (2001).
12. Fearnhead, P., Harding, R.M., Schneider, J.A., Myers, S. & Donnelly P. *Genetics* **167**, 2067–2081 (2004).
13. Ptak, S.E. *et al. Nat. Genet.* **37**, 429–434 (2005).
14. Winckler, W. *et al. Science* **308**, 107–111 (2005).
15. Gerton, J.L. *et al. Proc. Natl. Acad. Sci. USA* **97**, 11383–11390 (2000).