Minireview

Hotspots of homologous recombination in the human genome: not all homologous sequences are equal
James R Lupski

Address: Departments of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. Current address (sabbatical until July 2005): Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. E-mail: j12@sanger.ac.uk

Abstract

Homologous recombination between alleles or non-allelic paralogous sequences does not occur uniformly but is concentrated in 'hotspots' with high recombination rates. Recent studies of these hotspots show that they do not share common sequence motifs, but they do have other features in common.

Homologous recombination is the process whereby two DNA sequence substrates that share a significant stretch of identity are brought together, in an enzyme-catalyzed reaction, and undergo strand exchange to give a product that is a novel amalgamation of the two substrates. It occurs during meiosis, leading to crossovers between alleles (allelic homologous recombination, AHR), and during repair of double-strand breaks in DNA and other processes, leading to recombination between paralogous sequences (non-allelic homologous recombination, NAHR, also known as ectopic recombination). The intermediates of NAHR can be resolved to give several products, including deletions, duplications, and inversion rearrangements or, as in the case of AHR, the replacement of one sequence by a homologous one (gene conversion). When NAHR results in a duplication in one product it is usually accompanied by a reciprocal deletion in the other. Low-copy repeats that can induce NAHR account for 5-10% of the human genome [1], and rearrangements between them can result in a class of diseases known as genomic disorders [2,3].

Finding hotspots
It might be thought that homologous recombination is driven only by shared sequence identity among substrates. If this was the case, strand exchange would be expected to occur with equal frequency all the way along a segment of homology. Experimental observations suggest, however, that this is not the case and have provided evidence for local 'hotspots' - short regions of the genome where strand exchanges are more common than elsewhere. These observations come from pedigree studies that examined the parent-to-offspring transmission of alleles, linkage disequilibrium (LD) studies and, more recently, direct DNA sequencing of the products of recombination using either sperm (which represent a large number of recombination products from a single meiosis) or junction fragments from ectopic recombination (NAHR) [4,5]. These recombination hotspots are a common feature of both AHR and NAHR. Such hotspots have important implications for how linked genes and other markers are inherited in haplotypes (their amount of LD [4,6-8]) and for studies of LD and haplotypes including the International HapMap project [9], as well as potentially for disease-association studies and susceptibility to rearrangements causing genomic disorders in different world populations.

The distribution of meiotic recombination events along chromosomes has been examined at several levels of resolution, from the megabase (Mb) scales of genetic mapping (1 Mb is approximately equal to 1 centiMorgan (cM) for average recombination rates) to the nucleotide levels of resolution afforded by sequencing of strand-exchange products. High-resolution examination, at the nucleotide sequence level, defines hotspots as localized sites of recombination and enables recombination hotspots to be examined for

common features. The mechanism underlying the formation of recombination hotspots remains obscure, but recent studies suggest that a 'punctate' distribution of recombination events (in other words, a hotspot-like pattern of recombination) occurs throughout the human genome [6,10]. Furthermore, the local positions of recombination hotspots may not be conserved among closely related primate species [11], and in some cases hotspots are characterized by signatures of concerted evolution [7], whereby duplicated sequences are more similar to one another than to their orthologs in a closely related species.

The distribution of AHR across the genome has been reviewed recently [4,8]. Initial high-resolution analysis of human crossover hotspots characterized using sperm DNA studies identified a 1.5 kb region adjacent to the MS32 minisatellite [12] and several 1-2 kb intervals containing hotspots across the 210 kb class II region of the major histocompatibility complex [13,14]. Sperm analysis also identified a hotspot initially inferred from the observed nonuniform distribution of recombination within the human β-globin gene cluster [15,16]. These and other AHR hotspots cluster within small regions (1-2 kb), with crossover breakpoints spread in a normal distribution within the narrow hotspot; they have no obvious sequence similarities with one another, and coincide with gene-conversion hotspots [4]. The location of AHR hotspots is not conserved across distantly related mammalian species (human and mouse) [4], consistent with the fact that hotspots do not reflect conserved primary sequence motifs.

Jeffreys and colleagues [4] have pointed out that the punctate distribution of human recombination hotspots is very similar to that of meiotic double-strand breaks in budding yeast [17]; the latter are sequence-nonspecific and occur at yeast recombination hotspots [18,19], suggesting that hotspots could reflect where recombination is initiated by double-strand breaks. Furthermore, the observation that a recombination reporter placed in different positions in the yeast genome acquires properties of its location is argued [4] to support a model in which higher-order chromatin structures and/or chromosome dynamics contribute to the control of the local frequency of recombination-initiation events.

Hotspots have also been observed in association with NAHR (reviewed in [5]). The recombination event can be readily ascertained because the rearrangement (deletion or duplication) conveys a phenotype or produces a genomic disorder. Also, as paralogous sequences are used in NAHR, rather than allelic homologous sequences as in AHR, paralogous sequence variations (also known as *cis*-morphisms [3]) can be used to map crossover sites precisely. NAHR hotspots were initially observed in diverse populations as the recombinations associated with duplication and deletion rearrangements responsible for two common dominant peripheral neuropathies [5,20-23]. DNA structures that have been shown to induce double-strand breaks (such as palindromes, minisatellites and DNA transposons) have often been reported near NAHR hotspots (reviewed in [5,23]). Sequence analyses of the NAHR hotspots [21,22] revealed proximity to some of these structures, suggesting a link between double-strand breaks and NAHR hotspots [24]. Hotspots were observed subsequently in all NAHR crossovers examined at the nucleotide sequence level [25-28]. Like AHR hotspots, common features shared among NAHR hotspots include clustering within small regions (under 1 kb), no obvious sequence similarities with one another, and coincidence with apparent gene conversion events. Interestingly, recombination hotspots associated with reciprocal deletion and duplication events coincide; those associated with either the deletion or duplication could be used to predict the position of the hotspot associated with the reciprocal event [20,26].

Studying hotspot distribution systematically

The fine-scale structure of recombination-rate variation throughout the human genome was reported recently [6,10]. Both studies used surveys of single-nucleotide polymorphisms (SNPs) in different populations, and both developed novel statistical methods to infer patterns of fine-scale variation in the recombination rate along the genome. One study [10] focused on a 10 Mb region of chromosome 20 in European (Caucasian) and African-American populations, whereas the other [6] examined 74 candidate genes to search for hotspots by resequencing DNA from 23 European-Americans and 24 African-Americans. Both studies [6,10] found evidence for recombination-rate variation, with hotspots occurring at least every 200 kb and potentially as frequently as every 50 kb, the latter value being the same as has been observed in yeast [29]. No single factor was consistently associated with the presence of hotspots - neither GC content, the frequency of CpG dinucleotides, the presence of $(AC)_n$ repeats, nor any primary DNA sequence motif that had previously been hypothesized to influence the existence of hotspots. Whereas one fine-scale study [6] found extensive recombination-rate variations both within and between genes, the other [10] suggested that recombination occurs preferentially outside genes. The degree to which SNPs residing within segmental duplications (paralogous sequence variations or *cis*-morphisms [3,30-32]) influence the interpretation of these analyses remains to be determined.

Both studies [6,10] provided some evidence for differences in recombination-rate variation among different populations, but to what extent this reflects differences in the genetic background of the populations is not clear. The absence in the chimpanzee of a hotspot in the region homologous to the human recombination hotspot in the major histocompatibility complex *TAP2* gene suggests that recombination rates can change between very closely related species and raises the

possibility that recombination rates may differ among human populations [11].

What is the origin of recombination hotspots in the human genome? One recent study [7] of NAHR between two paralogous sequences that mediate deletions causing male infertility - human endogenous retrovirus (HERV) proviral sequences flanking the Y-chromosome locus *Azoospermia factor a* (*AZFa)* - provided evidence that several hominid-specific gene-conversion events have rendered the associated hotspots better substrates for chromosomal rearrangements in humans than in chimpanzees or gorillas. But, as the authors state [7], because gene conversion and chromosomal rearrangement reflect the alternative products of a common intermediate, it may be that a recombinogenic sequence motif or structure underpins the association, and increased sequence identity may play only a minor role in determining the frequency of chromosomal rearrangement. Nevertheless, the coincidence of the signatures of concerted evolution and recurrent breakpoints of chromosomal rearrangements (mapped at the DNA sequence level) may enable the identification of putative rearrangement hotspots from analysis of comparative sequences from great apes.

What causes hotspots?

What is the signal for recombination hotspots in the human genome? Does it reflect only the positional preference of double-stranded breaks by the recombination machinery? If so, is this dictated by access to the DNA because of a unique chromatin structure or is the signal contained within the DNA itself? We do know that the signal is not likely to be a *cis*-acting primary sequence motif similar to the *chi* of *Escherichia coli*, which stimulates recombination [33], as no such common motif has been identified in the multitude of AHR [4] and NAHR [5,28] hotspots studied to date, and the position of hotspots does not appear to be conserved among closely related primate species (at least for the *TAP2* hotspot) [11]. Such a signal could be embedded in a configuration consisting of a non-B form of DNA (such as Z DNA) [34], however, or could reflect an epigenetic mark such as methylation or the absence thereof in the hotspot region.

Recombination hotspots are being revealed as a global feature of the human genome [6,10]. Such hotspots have implications for studies of LD [6-8], the International HapMap Project [9], and for disease association studies in different world populations, because meiotic recombination exerts a profound influence on genome diversity and evolution [4]. They may also potentially be responsible for susceptibility within a population for NAHR-induced rearrangements associated with genomic disorders. Thus, functional studies to delineate the precise molecular mechanisms responsible for hotspots in the human genome are essential and are likely to enable further insights into the most basic properties of homologous recombination.

References

1. Eichler EE: **Segmental duplications: what's missing, misassigned, and misassembled - and should we care?** *Genome Res* 2001, **11**:653-656.
2. Stankiewicz P, Lupski JR: **Genome architecture, rearrangements and genomic disorders.** *Trends Genet* 2002,**18**:74-82.
3. Lupski JR: **Genomic disorders recombination-based disease resulting from genomic architecture.** *Am J Hum Genet* 2003, **72**:246-252.
4. Kauppi L, Jeffreys AJ, Keeney S: **Where the crossovers are: recombination distributions in mammals.** *Nat Rev Genet* 2004, **5**:413-424.
5. Inoue K, Lupski JR: **Molecular mechanisms for genomic disorders.** *Annu Rev Genomics Hum Genet* 2002, **3**:199-242.
6. Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M: **Evidence for substantial fine-scale variation in recombination rates across the human genome**. *Nat Genet* 2004, **36**:700-706.
7. Hurles ME, Willey D, Matthews L, Hussain SS: **Origins of chromosomal rearrangement hotspots in the human genome: evidence from the *AZFa* deletion hotspots.** *Genome Biol* 2004, **5**:R55.
8. Arnheim N, Calabrese P, Nordborg M: **Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved.** *Am J Hum Genet* 2003, **73**:5-16.
9. The International HapMap Consortium: **The International HapMap Project.** *Nature* 2003, **426**:789-796.
10. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P: **The fine-scale structure of recombination rate variation in the human genome.** *Science* 2004, **304**:581-584.
11. Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M: **Absence of the TAP2 human recombination hotspot in chimpanzees.** *PLoS Biol* 2004, **2**:849-855.
12. Jeffreys AJ, Murray J, Neumann R: **High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot.** *Mol Cell* 1998, **2**:267-273.
13. Jeffreys AJ, Ritchie A, Neumann R: **High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot.** *Hum Mol Genet* 2000, **9**:725-733.
14. Jeffreys AJ, Kauppi L, Neumann R: **Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex.** *Nat Genet* 2001, **29**:217-222.
15. Schneider JA, Peto TE, Boone RA, Boyce AJ, Clegg JB: **Direct measurement of the male recombination fraction in the human beta-globin hot spot.** *Hum Mol Genet* 2002, **11**:207-215.
16. Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH: **Nonuniform recombination within the human beta-globin gene cluster.** *Am J Hum Genet* 1984, **36**:1239-1258.
17. Baudat F, Nicolas A: **Clustering of meiotic double-strand breaks on yeast chromosome III.** *Proc Natl Acad Sci USA* 1997, **94**:5213-5218.
18. Xu L, Kleckner N: **Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast.** *EMBO J* 1995, **14**:5115-5128.
19. Xu F, Petes TD: **Fine-structure mapping of meiosis-specific double-strand DNA breaks at a recombination hotspot associated with an insertion of telomeric sequences upstream of the HIS4 locus in yeast.** *Genetics* 1996, **143**:1115-1125.
20. Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, Lupski JR: **A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element.** *Nat Genet* 1996, **12**:288-297.
21. Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR: **Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients.** *Am J Hum Genet* 1998, **62**:1023-1033.
22. Lopes J, Tardieu S, Silander K, Blair I, Vandenberghe A, Palau F, Ruberg M, Brice A, LeGuern E: **Homologous DNA exchanges in humans can be explained by the yeast double-strand break repair model: a study of 17p11.2 rearrangements associated with CMT1A and HNPP.** *Hum Mol Genet* 1999, **8**:2285-2292.
23. Lupski JR, Garcia A: **Charcot-Marie-Tooth peripheral neuropathies and related disorders.** In *The Metabolic and Molecular*

Bases of Inherited Diseases. Edited by Scriver CR, Sly WS, Childs B, Beaudet AL, Valle D, Kinzler KW, Vogelstein B. New York: McGraw-Hill; 2001, 5759-5788.

24. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW: **The double-strand-break repair model for recombination.** *Cell* 1983, **33:**25-35.

25. Lopez-Correa C, Dorschner M, Brems H, Lazaro C, Clementi M, Upadhyaya M, Dooijes D, Moog U, Kehrer-Sawatzki H, Rutkowski JL, *et al.*: **Recombination hotspot in NF1 microdeletion patients.** *Hum Mol Genet* 2001, **10:**1387-1392.

26. Bi W, Park SS, Shaw CJ, Withers MA, Patel PI, Lupski JR: **Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2.** *Am J Hum Genet* 2003, **73:**1302-1315.

27. Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA: **Mutational mechanisms of Williams-Beuren syndrome deletions.** *Am J Hum Genet* 2003, **73:**131-151.

28. Shaw CJ, Withers MA, Lupski JR: **Uncommon deletion of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates.** *Am J Hum Genet* 2004, **75:**75-81.

29. Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD: **Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2000, **97:**11383-11390.

30. Estivill X, Cheung J, Pujana MA, Nakabayashi K, Scherer SW, Tsui LC: **Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome.** *Hum Mol Genet* 2002, **11:**1987-1995.

31. Hurles M: **Are 100,000 "SNPs" useless?** *Science* 2002, **298:**1509.

32. Fredman D, White SJ, Potter S, Eichler EE, Dunnen JT, Brookes AJ: **Complex SNP-related sequence variation in segmental genome duplications.** *Nat Genet* 2004, **36:**861-866.

33. Smith GR: **Chi Sites and their Consequences.** In *Bacterial Genomes*. Edited by de Bruijn FJ, Lupski JR, Weinstock GM. New York: Chapman & Hall; 1998, 49-66.

34. Bacolla A, Wells RD: **Non-*B* DNA conformations, genomic rearrangements, and human disease.** *J Biol Chem* 2004. doi: 10.1074/jbc.R400028200.