nature genetics

PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans

Ingrid L Berg, Rita Neumann, Kwan-Wood G Lam, Shriparna Sarbajna, Linda Odenthal-Hesse, Celia A May & Alec J Jeffreys

© 2010 Nature America, Inc. All rights reserved.

PRDM9 has recently been identified as a likely *trans* regulator of meiotic recombination hot spots in humans and mice[1–3]. PRDM9 contains a zinc finger array that, in humans, can recognize a short sequence motif associated with hot spots[4], with binding to this motif possibly triggering hot-spot activity via chromatin remodeling[5]. We now report that human genetic variation at the *PRDM9* locus has a strong effect on sperm hot-spot activity, even at hot spots lacking the sequence motif. Subtle changes within the zinc finger array can create hot-spot nonactivating or enhancing variants and can even trigger the appearance of a new hot spot, suggesting that PRDM9 is a major global regulator of hot spots in humans. Variation at the *PRDM9* locus also influences aspects of genome instability—specifically, a megabase-scale rearrangement underlying two genomic disorders[6] as well as minisatellite instability[7]—implicating PRDM9 as a risk factor for some pathological genome rearrangements.

PR domain–containing 9 (PRDM9) is a meiosis-specific histone H3 methyltransferase with a C-terminal tandem-repeat C2H2 zinc finger (ZnF) domain encoded by a minisatellite[8]. Mouse subspecies hybrids show variation in their recombination profiles and crossover hot-spot activity depending on the *Prdm9* alleles they carry[3]. In humans, evidence that PRDM9 influences recombination is less direct. First, ~40% of human hot spots contain the motif CCNCCNTNNCCNC, which might serve as a binding site for a ZnF protein[4,9], and bioinformatic analysis has identified PRDM9 as the human ZnF protein most likely to recognize this motif[2]. Second, the most common PRDM9 variant coded by the A allele binds this motif *in vitro*[1]. Third, individuals in a single pedigree carrying the variant *PRDM9* allele I, encoding a ZnF array that cannot bind this motif, show a genome-wide shift in hot spot usage, as shown by linkage analysis[1].

We now test the influence of variation in the PRDM9 ZnF array on human recombination hot spots. The hypothesis is straightforward—hot spots containing a motif should show high activity in sperm from men carrying two motif-recognizing *PRDM9* alleles and lower activity in men lacking these alleles. Conversely, hot spots lacking the motif will presumably be regulated by some other mechanism and should not be influenced by *PRDM9* status. Limited *PRDM9* variability in Europeans[1,3] makes this test difficult. We therefore surveyed ZnF alleles in a panel of 74 African semen donors and 156 European donors, revealing a diversity of alleles having 8–18 ZnFs, including the known alleles A–E[1] and 24 novel alleles (L1–L24) (**Fig. 1a** and **Supplementary Fig. 1**). Allele I associated with a shift in hot-spot usage[1] was not seen. Diversity in Africans was high[3] (**Fig. 1b**), and classification of alleles according to the predicted ability of the resulting protein variants to recognize the hot-spot motif (**Fig. 1a** and **Supplementary Fig. 1**) showed that nearly half of these variants should result in impaired binding relative to the common A allele (**Fig. 1b**), thus providing a key resource for analyzing PRDM9 effects.

We selected ten very active crossover hot spots for study—five with a central motif and five lacking a clear motif[10,11] (**Supplementary Fig. 2**). For each hot spot, we identified semen donors amenable to crossover detection[12]. From these donors, we selected men carrying two common A alleles (A/A) and men carrying one non-A allele (A/N) or two non-A alleles (N/N). Samples from multiple men were analyzed to smooth out variation caused by factors such as variable DNA quality and possible *cis*-acting variants[11,13,14] that could influence recombination frequency estimates. Simple comparison of crossover frequencies in men with A/A, A/N and N/N genotypes showed a dramatic effect, with almost all testable N/N men showing heavy reduction of recombination to <5% of normal activity at all hot spots, even at those that do not contain an obvious hot-spot motif (**Fig. 1c**,**d** and **Supplementary Fig. 2**). This association between N/N status and very low recombination frequency was highly significant. Thus, ranking of 112 men by recombination frequency over all hot spots tested (Online Methods) showed that the 17 lowest-ranked individuals included 14 of the 15 men with N/N genotypes (**Supplementary Table 1**) ($P = 5 \times 10^{-16}$). This association remained in a subgroup of men of Southern East-African descent (11 N/N men from this population all ranked lowest in the 38 men tested; $P = 2 \times 10^{-9}$), ruling out differences between populations at other loci as significant confounding variables. We therefore conclude that PRDM9 is a major regulator of human hot spots.

The strong crossover suppression seen in most N/N men suggests that most *PRDM9* non-A alleles are incapable of triggering activity at the hot spots tested. However, activating N alleles do exist, as shown by one man who was fully active at hot spot K (**Fig. 1c**); this man carried two N alleles (L9 and L10) that are both predicted to bind the motif (**Supplementary Fig. 1**), though it is impossible

Department of Genetics, University of Leicester, Leicester, UK. Correspondence should be addressed to A.J.J. (ajj@le.ac.uk).

Received 19 May; accepted 11 August; published online 5 September 2010; doi:10.1038/ng.658

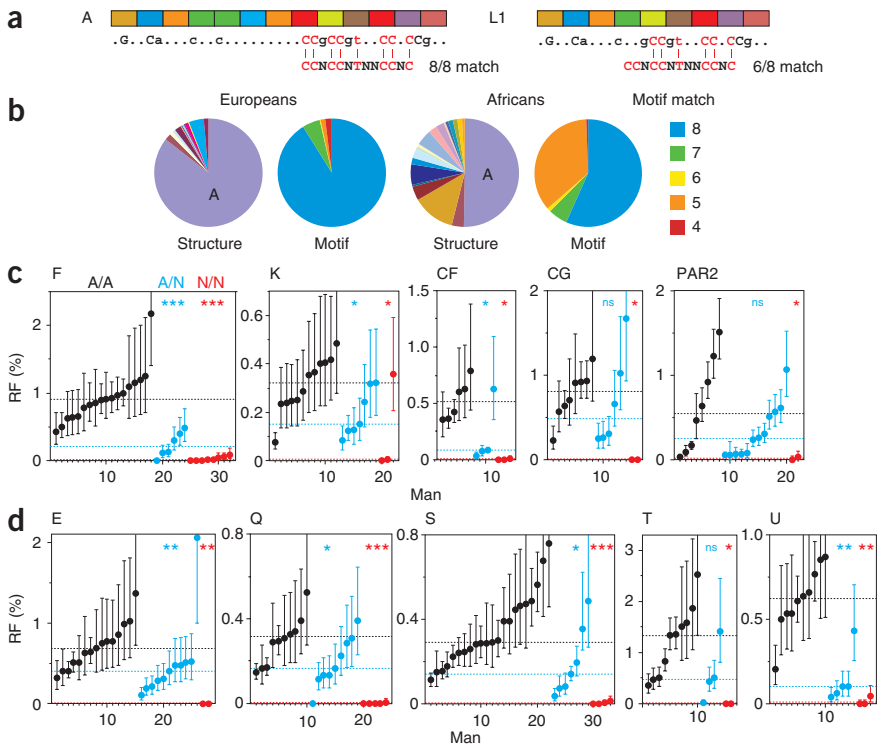

Figure 1 *PRDM9* ZnF variants and crossover hot-spot activity in sperm. (**a**) Examples of tandem repeats encoding the ZnF array with variant repeat units colored differently. The predicted DNA binding sequence is shown below, with dots indicating weakly predicted bases and uppercase letters indicating the most strongly predicted bases, and aligned with the hot-spot motif CCNCCNTNNCCNC (ref. 4). The binding sequence for allele A matches all eight specified bases in the motif, while allele L1 matches at best only six of the eight bases. (**b**) *PRDM9* ZnF array diversity in Europeans and Africans, with alleles classified either by structure or by the strength of the predicted match with the motif (details of alleles are provided in **Supplementary Fig. 1**). (**c**) Variation between men in sperm crossover activity in hot spots, named above each panel, containing a central hot-spot motif. Different sets of men informative for the SNPs required for crossover detection were analyzed at each hot spot (**Supplementary Table 1**). Men carrying two *PRDM9* A alleles (A/A, shown in black), one A allele (A/N, shown in blue) or two non-A alleles (N/N, shown in red) were grouped separately in ascending order. Confidence intervals for each estimate of recombination frequency are shown, and median recombination frequencies within each group are indicated by dotted lines. Mann-Whitney test results for the significance of differences between the A/A group and the A/N or N/N groups are given at the top right (ns, not significant, $P > 0.05$; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). (**d**) Corresponding analyses of hot spots lacking an obvious hot-spot motif (**Supplementary Fig. 2**).

to determine whether both alleles are activating. As expected, A/N men tended to show significantly lower recombination frequencies than A/A men over all ten hot spots ($P = 7 \times 10^{-18}$ for all hot spots combined) (**Fig. 1c,d**). Median recombination frequencies in A/N men were 41% ± 16% of those in A/A men, which is consistent with a simple additive model whereby hot-spot activity is proportional to the number of activating alleles present.

Fourteen different *PRDM9* alleles were identified as nonactivators in heavily suppressed N/N men who showed <5% of the median crossover frequency seen in A/A men; in each case, the allele failed to activate any testable hot spot (**Fig. 2**). Thus, allele C, which is common in Africans and encodes major changes in the ZnF array, could not activate any of the nine testable hot spots. Surprisingly, there was no obvious relationship between the inability of an allele to activate and its predicted binding motif, with even subtle changes in the ZnF array creating hot-spot–nonactivating alleles. For example, allele L20 is closely related to the recombination-promoting allele A, differing only by a single Asn→His substitution in the tenth ZnF. This substitution is located at one of the four key DNA-contact residues in C2H2 ZnFs (specifically, at position 3 of the −1, 2, 3 and 6 residues; **Supplementary Fig. 1**)[1,2,15] and should alter its DNA-binding characteristics. More remarkably, the nonactivating allele L13 is identical to the A allele except for a Ser→Arg substitution at the noncontact position −2 in the 11th ZnF, which is a change not predicted to alter

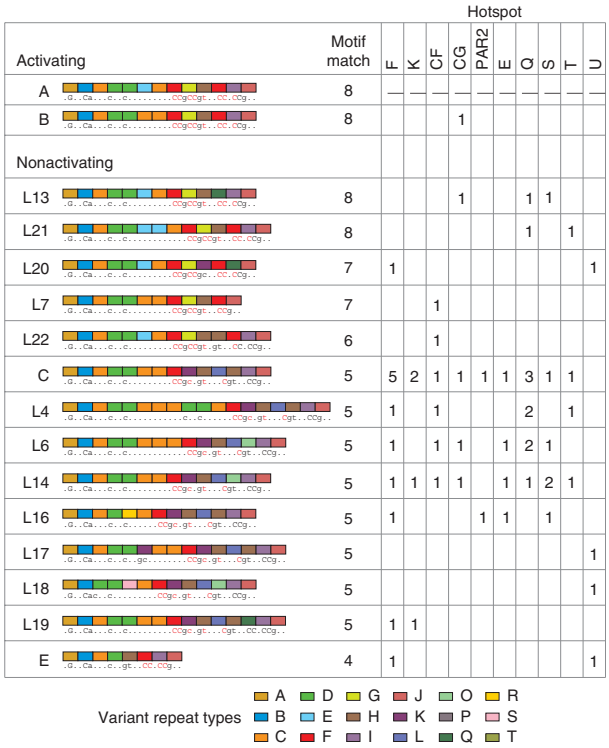

Figure 2 Activating and nonactivating *PRDM9* alleles. Nonactivating alleles at each hot spot were defined as alleles present in N/N men who showed <5% of the median recombination frequency seen in A/A men. Allele structures are coded as in **Supplementary Figure 1**, with predicted DNA binding sequences and best motif matches shown as in **Figure 1**. Data for each hot spot give the number of specific N alleles detected in suppressed men; for example, five such men typed at hot spot F carried the C allele. Evidence that the B allele is active is based on a B/L6 heterozygote assayable only at hot spot CG who showed crossovers at 40% of the median frequency seen in A/A homozygotes. Because allele L6 is a nonactivator at CG, this implies that allele B is similar in activity to allele A.

© 2010 Nature America, Inc. All rights reserved.

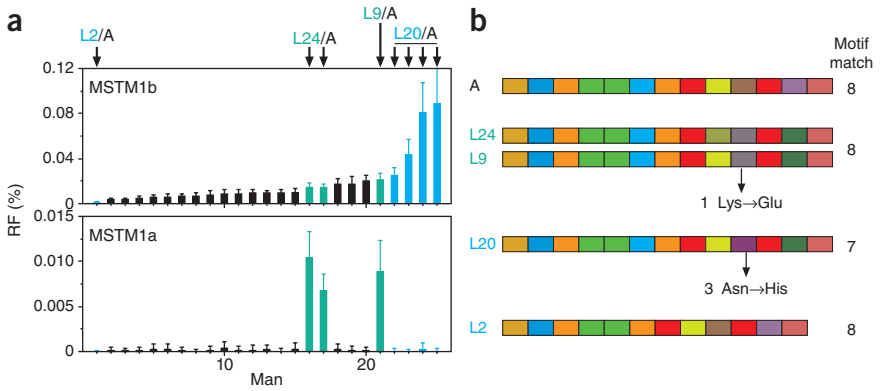

Figure 3 *PRDM9* variation influences crossover activity at hot spots MSTM1a and MSTM1b. (**a**) Variant alleles present in men previously typed for sperm crossovers at these hot spots[16]; all other men were A/A homozygotes. Recombination frequencies at each hot spot are shown with 95% CIs. Note that alleles L9 and L24 associate not only with MSTM1a activity but also apparently with elevated recombination frequency at MSTM1b ($P = 0.035$). (**b**) DNA sequence structures of these alleles, colored as in **Figure 2**, plus amino acid sequence changes relative to allele A, with locations given with respect to the main ZnF DNA-contact residues (−1, 2, 3 and 6)[15].

DNA binding[15]. This inability to activate hot spots might be due to allele L13 carrying an additional alteration that inactivates the gene, though such null alleles should be rare given that *Prdm9* disruption in mice causes sterility[8].

We extended the search for *PRDM9* influences to a pair of hot spots (MSTM1a and MSTM1b; **Supplementary Fig. 2**) previously shown to vary in sperm crossover activity independently of local DNA sequence[16]. MSTM1b showed a 75-fold variation in activity (**Fig. 3a**), and the lowest-ranked man we analyzed carried the non-A allele L2, an association of borderline significance ($P = 0.04$). Notably, the four men with the highest recombination frequencies all carried allele L20, a significant association ($P = 0.0002$) establishing that L20 is a recombination-enhancing allele for MSTM1b. Under a simple additive model, we estimated that this allele enhances recombination ~13-fold relative to the A allele. Remarkably, L20 cannot activate hot spots F and U (**Fig. 2**), indicating that *PRDM9* alleles can have different influences on different hot spots.

The neighboring hot spot MSTM1a was active in only 3 of the 25 men surveyed[16]. These three men carried the non-A alleles L9 or L24 (**Fig. 3a**), which differ from each other by a single synonymous base substitution and thus code for the same ZnF array. This association between the L9/L24 alleles and hot-spot activity was significant ($P = 0.0004$). Alleles L9 and L24 differ from allele A by a single Lys→Glu replacement not predicted to affect DNA binding (**Fig. 3b**). Thus, a very subtle change in the amino acid sequence can trigger the appearance of a new hot spot.

The hot-spot motif is also associated with genomic regions that are prone to rearrangement[4]. We therefore extended the *PRDM9* survey to modes of meiotic genome instability detectable as *de novo* events in sperm. Small-pool PCR analysis of three highly unstable minisatellites[17–19], which each show instability that is driven substantially by meiotic gene conversion-like events, revealed a major *PRDM9* effect with instability in N/N men reduced to 3–10% of the level seen in A/A men (**Fig. 4**). Rank-order analysis of men showed that low minisatellite instability is significantly associated with N/N status ($P = 0.000015$). PRDM9 clearly regulates minisatellite instability, though it is unclear whether this occurs by interaction of the ZnF array with hot-spot motifs present within the minisatellites themselves (**Fig. 4a**)[4] or by interaction with a flanking recombination hot spot that in turn drives repeat turnover[20,21].

The genomic disorders Charcot-Marie-Tooth type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP) result from unequal meiotic exchanges between CMT1A-REP sequences that are separated by 1.5 Mb[22], and *de novo* rearrangements in sperm can be detected by PCR amplification of the exchange junctions[6]. Exchanges cluster into a narrow hot spot that contains a hot-spot motif[4] and that may also function in allelic crossover[23]. Analysis of the frequency of these rearrangements in sperm again showed a major effect of *PRDM9*, with N/N men showing ~4% of the activity seen in A/A men and with A/N men showing reduced instability (**Fig. 5**).

We also tested whether *PRDM9* alleles influence the frequency of the recurrent translocation t(11;22)(q23;q11). Translocation junctions map within a short palindromic AT-rich repeat (PATRR), allowing *de novo* translocations to be detected by PCR[24]. As found previously[24], we saw these translocations in sperm but not in somatic DNA (**Fig. 6a**). Translocation frequencies in sperm varied widely between men (**Fig. 6b**), reflecting the sensitivity of the process to

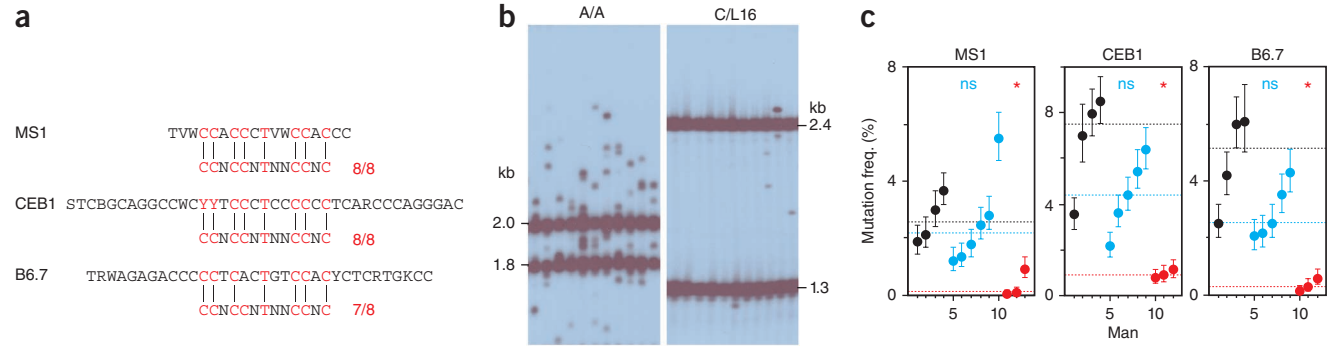

Figure 4 Influence of *PRDM9* variation on minisatellite instability in sperm. (**a**) Minisatellite repeat units, with the best matches to the hot-spot motif indicated. Minisatellites MS1 and CEB1 both contain perfect matches. Two contiguous repeats are shown for MS1. (**b**) Representative small-pool PCR results for minisatellite CEB1, for a *PRDM9* A/A homozygote and a C/L16 heterozygote. (**c**) Mutation frequencies in A/A (black), A/N (blue) and N/N (red) men, with CIs, median values per group, and Mann-Whitney test results indicated as in **Figure 1** (ns, not significant, $P > 0.05$; $*P < 0.05$).

© 2010 Nature America, Inc. All rights reserved.

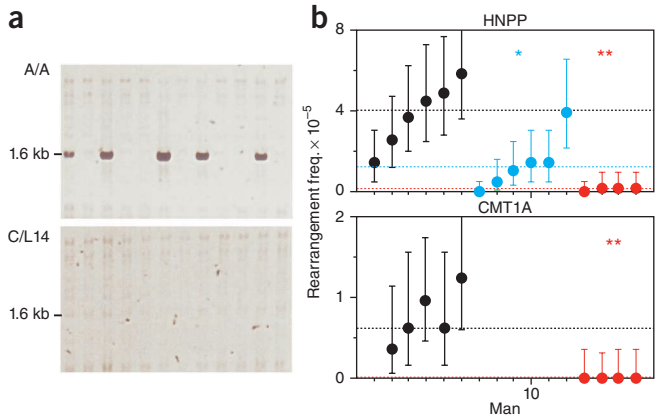

Figure 5 *PRDM9* variation and *de novo* HNPP/CMT1A rearrangements in sperm DNA. (**a**) Detection of HNPP deletion junctions in sperm DNA from a *PRDM9* A/A homozygote and a C/L14 heterozygote with 40 ng DNA input per PCR reaction. (**b**) Rearrangement frequencies in A/A (black), A/N (blue) and N/N (red) men for HNPP deletions and CMT1A duplications, with CIs, median values per group, and Mann-Whitney test results shown as in **Figure 1** (*$P < 0.05$; **$P < 0.01$). The same men were analyzed for both rearrangements, but not all men were typed for duplications, which arise at a lower frequency than deletions[6]. Rearrangements in blood were rare[6] (frequency of $< 2.2 \times 10^{-6}$ for deletions and $< 1.3 \times 10^{-6}$ for duplications; $P > 0.95$).

variation in the PATRRs[25]. However, comparison of *PRDM9* A/A, A/N and N/N men showed no significant influence of *PRDM9* on translocation frequency, with indistinguishable median frequencies, no evidence for suppression in N/N men and no indication of specific non-A alleles being associated with unusual translocation frequencies. This suggests that, although these translocations are probably meiotic in origin[24], they are not facilitated by PRDM9-mediated chromatin remodeling.

We therefore conclude that variation within the PRDM9 ZnF domain strongly influences hot-spot activity. The differential effects of allele L20 on different hot spots is consistent with PRDM9 directly interacting with hot spots. The independent variation seen at hot spots MSTM1a and MSTM1b, which have centers only 2 kb apart, suggests an interaction highly local to the hot spot and could argue against extended chromatin domains being opened by PRDM9. It is surprising how sensitive hot spots can be to minor variation within the ZnF domain, with variant alleles differing by a single amino acid substitution that can enhance or fail to activate hot spots or even trigger the appearance of a new hot spot. This raises a fascinating conundrum: how can PRDM9 variants that are predicted not to influence DNA binding have a major impact on hot-spot activity, while hot-spot–promoting alleles such as the A allele can activate hot spots that lack an obvious hot-spot motif? If PRDM9 does indeed regulate hot spots by binding to the motif, as suggested by ZnF bioinformatic surveys[2] and *in vitro* studies[1], then the rules governing interactions between PRDM9 and hot spots must be subtle and complex and are most likely only addressable by *in vivo* studies. In any event, these rules are evidently not well captured by consensus motifs established from large-scale comparisons of historical hot spots[4], nor do such motifs provide clear predictions of hot-spot locations (**Supplementary Fig. 2**).

PRDM9 status is a major risk factor for *de novo* CMT1A and HNPP rearrangements, with non-A alleles being strongly protective in N/N individuals, at least in males. Substantial shifts in *PRDM9* allele frequencies between populations imply corresponding shifts in the incidence of *de novo* rearrangements. They also suggest major changes in crossover hot spot usage in different populations. If variant *PRDM9* alleles trigger the appearance of new hot spots genome wide (as implied by the L9- and L24-specific appearance of hot-spot MSTM1a), then Africans are likely to use a relatively broad repertoire of hot spots, providing an explanation (in addition to that of greater population age) for the relatively short haplotype blocks and rapid decay of linkage disequilibrium (LD) seen in Africans[26]. One challenge moving forward will be to identify African-specific hot spots and to see whether they too are 'tuned' to a specific subset of *PRDM9* alleles. We note that almost all hot spots in the present study were initially identified from LD breakdown in Europeans[10,11] and were therefore likely to be activated by the A allele of *PRDM9*. The only exception to this was hot spot MSTM1a, which has not left a mark on LD in Europeans[16] and is not activated by this allele.

Are there other *trans* regulators, equivalent to PRDM9, of meiotic recombination hot spots in humans? Even though all hot spots analyzed to date are regulated by PRDM9, substantial variation in recombination frequency can remain between men of the same *PRDM9* status (for example among A/A men at hot spot S; **Fig. 1d**). Such variation might provide a lead to other *trans* regulators, such as loci already identified that associate with variation in linkage map lengths[27,28]. Finally, the influence of PRDM9 ZnF variants on minisatellite instability is intriguing, given that these variants themselves are encoded by a minisatellite. Could *PRDM9* influence its own evolution by generating alleles that affect instability at the *PRDM9* minisatellite to promote the generation of new alleles? If so, then bursts of new *PRDM9* alleles could be generated whenever an interacting allele appears in a population, implying a potentially chaotic mode of hot-spot evolution.

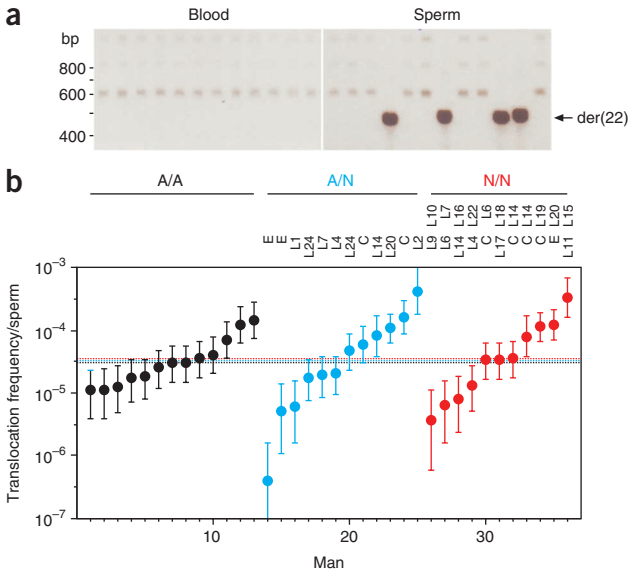

Figure 6 *PRDM9* variation and t(11;22) translocation frequencies. (**a**) Detection of *de novo* der(22) translocation junctions by nested PCR amplification[29] of multiple 150 ng aliquots of blood and sperm DNA. Minor variation in junction size resulted from differences in translocation breakpoint locations within the PATRR[24]. (**b**) der(22) translocation frequencies, with CIs, in sperm DNA in *PRDM9* A/A, A/N and N/N men, with identities of the N alleles shown above. The median frequency per group, similar to previously reported translocation frequencies[24], is indicated by a dotted line. There was no significant difference in translocation frequency between the groups (Kruskal-Wallis test, $P = 0.98$). No *de novo* translocations were seen in blood DNA tested from four different men (frequency $< 5 \times 10^{-7}$, $P > 0.95$).

© 2010 Nature America, Inc. All rights reserved.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Sequences of the *PRDM9* novel alleles L1–L24 have been deposited in GenBank with the accession numbers HM210983–HM211006.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J. Blower and volunteers for providing semen samples, colleagues for helpful discussions and the Medical Research Council, the Wellcome Trust (ref. 081227/Z/06/Z), the Boehringer Ingelheim Fonds, the Royal Society and the Louis-Jeantet Foundation for funding support.

AUTHOR CONTRIBUTIONS

I.L.B., R.N., K.-W.G.L., S.S., L.O.-H., C.A.M. and A.J.J. all contributed to designing aspects of the study. I.L.B., R.N. and A.J.J. characterized *PRDM9* alleles, I.L.B., S.S., L.O.-H. and A.J.J. analyzed hot spots, R.N. surveyed minisatellite instability, K.-W.G.L. characterized genome rearrangements and A.J.J. analyzed translocations. A.J.J. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
2. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).
3. Parvanov, E.D., Petkov, P.M. & Paigen, K. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
4. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* **40**, 1124–1129 (2008).
5. Paigen, K. & Petkov, P. Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet.* **11**, 221–233 (2010).
6. Turner, D.J. *et al.* Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders. *Nat. Genet.* **40**, 90–95 (2008).
7. Jeffreys, A.J. *et al.* Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6**, 136–145 (1994).
8. Hayashi, K., Yoshida, K. & Matsui, Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**, 374–378 (2005).
9. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
10. Webb, A.J., Berg, I.L. & Jeffreys, A. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proc. Natl. Acad. Sci. USA* **105**, 10471–10476 (2008).
11. Jeffreys, A.J. & Neumann, R. The rise and fall of a human recombination hot spot. *Nat. Genet.* **41**, 625–629 (2009).
12. Jeffreys, A.J., Ritchie, A. & Neumann, R. High resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.* **9**, 725–733 (2000).
13. Jeffreys, A.J. & Neumann, R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.* **31**, 267–271 (2002).
14. Jeffreys, A.J. & Neumann, R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* **14**, 2277–2287 (2005).
15. Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys_2His_2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
16. Neumann, R. & Jeffreys, A.J. Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Hum. Mol. Genet.* **15**, 1401–1411 (2006).
17. Buard, J., Bourdet, A., Yardley, J., Dubrova, Y. & Jeffreys, A.J. Influences of array size and homogeneity on minisatellite mutation. *EMBO J.* **17**, 3495–3502 (1998).
18. Tamaki, K., May, C.A., Dubrova, Y.E. & Jeffreys, A.J. Extremely complex repeat shuffling during germline mutation at human minisatellite B6.7. *Hum. Mol. Genet.* **8**, 879–888 (1999).
19. Berg, I., Neumann, R., Cederberg, H., Rannug, U. & Jeffreys, A.J. Two modes of germline instability at human minisatellite MS1 (locus *D1S7*): complex rearrangements and paradoxical hyperdeletion. *Am. J. Hum. Genet.* **72**, 1436–1447 (2003).
20. Jeffreys, A.J., Murray, J. & Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**, 267–273 (1998).
21. Conrad, D.F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
22. Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. & Lupski, J.R. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat. Genet.* **2**, 292–300 (1992).
23. Lindsay, S.J., Khajavi, M., Lupski, J.R. & Hurles, M.E. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.* **79**, 890–902 (2006).
24. Kurahashi, H. & Emanuel, B.S. Unexpectedly high rate of *de novo* constitutional t(11;22) translocations in sperm from normal males. *Nat. Genet.* **29**, 139–140 (2001).
25. Kato, T. *et al.* Genetic variation affects *de novo* translocation frequency. *Science* **311**, 971 (2006).
26. Reich, D.E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
27. Kong, A. *et al.* Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* **319**, 1398–1401 (2008).
28. Chowdhury, R. *et al.* Genetic analysis of variation in human meiotic recombination. *PLoS Genet.* **5**, e1000648 (2009).
29. Kurahashi, H. *et al.* Tightly clustered 11q23 and 22q11 breakpoints permit PCR-based detection of the recurrent constitutional t(11;22). *Am. J. Hum. Genet.* **67**, 763–768 (2000).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Characterizing *PRDM9* ZnF alleles. Semen samples were collected with approval from the Leicestershire Health Authority Research Ethics Committee and with informed consent. The collection of African samples has been described previously[30]. DNAs were extracted and manipulated under conditions designed to minimize the risk of contamination[20]. *PRDM9* minisatellite alleles were amplified from 230 men using the primers PN0.6F and PN2.5R (**Supplementary Table 2**), which were designed not to amplify the paralogous locus *PRDM7* (ref. 1). Alleles in men showing minisatellite alleles of two different lengths were separated by agarose gel electrophoresis and sequenced with the primers PN1.2F and PN2.4R (**Supplementary Table 2**) using BigDye Terminator v3.1 Cycle Sequencing (Applied Biosystems) on a 3730 DNA Analyzer (Applied Biosystems). All remaining length homozygotes who were subsequently subjected to crossover analysis (consisting almost exclusively of men carrying two 13-repeat alleles) were sequenced from diploid PCR products. In most instances, nonmixed sequence traces were obtained that matched the A allele of *PRDM9*, indicating A/A homozygosity. All remaining instances of mixed traces could be fully explained by heterozygosity for allele A and another already-characterized allele, allowing the allele status of these individuals to be deduced. In total, 300 *PRDM9* alleles were sequenced. The DNA binding sequence predicted for each allele was established as previously described[1] and simplified to show the most strongly predicted bases (in lowercase for >80% consensus for a specific base and in uppercase for >95% consensus across >60 observations). Analysis of sperm counts estimated from DNA yields across all Africans showed no significant differences between A/A, A/N and N/N men (Kruskal-Wallis test, 2 degrees of freedom, $P = 0.97$), suggesting no obvious effect of *PRDM9* ZnF variation on fertility.

Hot-spot analysis. A panel of donors including all 74 Africans in the study and 98 Europeans, including all individuals with atypical-length *PRDM9* alleles, was genotyped for SNPs across ten different hot spots. Hot spots E, F, K, Q and S have been described previously[10,11]. Hot spot S is a doublet in some Europeans, with the second hot spot at this location being triggered by a single base substitution in *cis*[11]; however, this substitution was absent from all additional men tested in the present study who should therefore contain only a single hot spot. Details of the remaining five hot spots (T, U, CF, CG and PAR2) are provided in **Supplementary Figure 2**. Men for crossover analysis were selected based on appropriate hot-spot genotypes and *PRDM9* status. Crossovers were analyzed as described previously[12] using nested repulsion-phase allele-specific long PCR to selectively amplify crossover molecules from sperm DNA. Shifts in SNP frequencies in Africans at some hot spots forced the development of additional allele-specific primers to maximize the number of analyzable men. A full list of primers is provided in **Supplementary Table 3**. For each hot spot and individual in the study, we typically analyzed 4 × 12 aliquots of sperm DNA estimated from recombination frequency data to each contain 0.5, 1.0, 2.0 and 4.5 crossover molecules per aliquot. This strategy provided reliable estimates of recombination frequency in active men and had sufficient power to detect major reduction in recombination frequency in inactive men to <3% of normal activity. All DNA concentrations were quantified on a NanoDrop 1000 spectrophotometer, and a single DNA molecule PCR efficiency of 50% was assumed throughout[10]. A full listing of recombination frequencies in all men tested at each hot spot is provided in **Supplementary Table 1**. In rank-order analysis of recombination frequencies in groups of men, the B allele of *PRDM9* was treated as an A allele because of its similarity to A and its lack of major hot spot–shifting effects in linkage maps[1], plus evidence that the B allele is fully active (**Fig. 2**). Further evidence for activity of the B allele came from analysis of A/B heterozygotes, assayed at hot spots F, CF, CG, S and U, who showed crossover activity at 80% of the median value seen in A/A homozygotes. To assess associations between *PRDM9* status and recombination frequency, we normalized recombination frequencies at each hot spot in each man with respect to the median value seen in all A/A men. We then averaged these normalized recombination frequencies over all hot spots tested in an individual. All individuals were then ranked using these mean recombination frequencies, and the association between low recombination frequency and N/N status was assessed by the Mann-Whitney test. This approach is conservative, as crossover suppression over multiple hot spots in an individual was treated as a single observation.

Additional analyses of crossovers. Crossover locations already established for hot spots MSTM1a and MSTM1b[16], plus locations analyzed in men typed at hot spot PAR2, showed no significant shifts in hot-spot position between men. PAR2 contains a central hot-spot motif with a motif-disrupting SNP (rs700442) that suppresses crossovers to ~30% of normal activity. This *cis*-acting influence was ignored in **Figure 1c**, though similar conclusions about the influence of *PRDM9* were drawn for men who were matched for allelic status at this SNP. A full description of this hot spot will be provided elsewhere. Possible *cis*-acting SNPs at hot spots E and F were also checked by resequencing all analyzed donors over a 1.5 kb interval spanning the center of the hot spot. No obvious *cis* effects were detected in these intervals.

Analysis of minisatellite instability. Men were typed for allele length at all three minisatellites[17–19], and individuals carrying two relatively short (0.6–5.2 kb) alleles were chosen for mutation analysis to facilitate mutant resolution. *De novo* mutants were analyzed by small-pool PCR of multiple aliquots of sperm DNA, with each aliquot containing 0.2–0.4 ng DNA (40–80 haploid genomes, assuming 60% single molecule PCR efficiency), followed by agarose gel electrophoresis and Southern blot hybridization with a ^{32}P-labeled minisatellite probe[17–19]. Mutation frequencies in all men tested are provided in **Supplementary Table 1**.

CMT1A and HNPP rearrangements. Exchange junctions within the CMT1A-REPs were the same as described previously[6] but were detected using different paralog-specific PCR primers (listed in **Supplementary Table 3**). Primer sites in both REPs were sequenced in all men selected for analysis, and individuals carrying a variant at any site were excluded. *De novo* exchanges were detected using nested paralog-specific PCR on typically 72 PCR reactions per man, each seeded with 40–80 ng sperm DNA (6,700–13,300 haploid genomes, assuming 50% single molecule PCR efficiency), in parallel with 12–24 PCR reactions seeded with 60–80 ng blood DNA as a negative control[6]. PCR cycling for HNPP deletions was at 96 °C for 1 min, followed by 28 or 34 cycles (for primary and secondary PCRs respectively) at 96 °C for 20 s, 49 °C/48 °C (for primary and secondary PCRs) for 30 s, and 66 °C for 2.5 min. Cycling for CMT1A duplications used this same procedure but with annealing at 46 °C/49 °C (for primary and secondary PCRs). Rearrangement frequencies in all men tested are provided in **Supplementary Table 1**.

Translocations. *De novo* der(22) junctions arising from t(11;22) translocations were detected by nested PCR using primers and amplification conditions described previously[29] but with the PCR extension temperature reduced to 60 °C to facilitate amplification across the AT-rich junction. Twenty-four PCR reactions, each seeded with 75–150 ng sperm DNA, were analyzed per man, with additional assays at lower DNA inputs for men showing the highest translocation frequencies. Assays for the reciprocal der(11) junction[29] performed on five men yielded translocation frequencies indistinguishable from der(22) frequencies. Translocation frequencies in all men tested are provided in **Supplementary Table 1**.

30. Monckton, D.G. *et al.* Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nat. Genet.* **8**, 162–170 (1994).

© 2010 Nature America, Inc. All rights reserved.

doi:10.1038/ng.658