

poorly understood, and the elucidation of such mechanisms could lead to the discovery of novel therapeutics.

References and Notes

1. B. Moss, in *Fields Virology*, D. M. Knipe, Ed. (Lippincott Williams & Wilkins, Philadelphia, 2007), vol. 2, pp. 2905–2946.
2. R. C. Condit, N. Moussatche, P. Traktman, *Adv. Virus Res.* **66**, 31 (2006).
3. G. L. Smith, A. Vanderplasschen, M. Law, *J. Gen. Virol.* **83**, 2915 (2002).
4. K. L. Roberts, G. L. Smith, *Trends Microbiol.* **16**, 472 (2008).
5. C. M. Sanderson, M. Way, G. L. Smith, *J. Virol.* **72**, 1235 (1998).
6. R. Blasco, N. B. Cole, B. Moss, *J. Virol.* **65**, 4598 (1991).
7. G. Hiller, K. Weber, L. Schneider, C. Parajsz, C. Jungwirth, *Virology* **98**, 142 (1979).
8. G. V. Stokes, *J. Virol.* **18**, 636 (1976).
9. S. Cudmore, P. Cossart, G. Griffiths, M. Way, *Nature* **378**, 636 (1995).
10. G. C. Carter *et al.*, *J. Gen. Virol.* **84**, 2443 (2003).
11. Materials and methods are available as supporting material on *Science* online.
12. L. G. Payne, K. Kristenson, *J. Virol.* **32**, 614 (1979).
13. P. C. Turner, R. W. Moyer, *Virology* **380**, 226 (2008).
14. T. R. Wagenaar, B. Moss, *J. Virol.* **81**, 6286 (2007).
15. T. R. Wagenaar, B. Moss, *J. Virol.* **83**, 1546 (2009).
16. K. M. Law, G. L. Smith, *J. Gen. Virol.* **73**, 549 (1992).
17. M. Law, R. Hollinshead, G. L. Smith, *J. Gen. Virol.* **83**, 209 (2002).
18. M. Engelstad, G. L. Smith, *Virology* **194**, 627 (1993).
19. R. L. Roper, E. J. Wolffe, A. Weisberg, B. Moss, *J. Virol.* **72**, 4192 (1998).
20. J. E. Parkinson, G. L. Smith, *Virology* **204**, 376 (1994).
21. J. Rosel, B. Moss, *J. Virol.* **56**, 830 (1985).
22. M. M. Lorenzo, I. Galindo, G. Griffiths, R. Blasco, *J. Virol.* **74**, 10535 (2000).
23. R. L. Roper, L. G. Payne, B. Moss, *J. Virol.* **70**, 3753 (1996).
24. H. van Eijl, M. Hollinshead, G. L. Smith, *Virology* **271**, 26 (2000).
25. B. Perdiguerro, R. Blasco, *J. Virol.* **80**, 8763 (2006).
26. S. Röttger, F. Frischknecht, I. Reckmann, G. L. Smith, M. Way, *J. Virol.* **73**, 2863 (1999).
27. E. J. Wolffe, A. S. Weisberg, B. Moss, *J. Virol.* **75**, 303 (2001).
28. A. A. McIntosh, G. L. Smith, *J. Virol.* **70**, 272 (1996).
29. W. H. Zhang, D. Wilcock, G. L. Smith, *J. Virol.* **74**, 11654 (2000).
30. R. Dulbecco, *Proc. Natl. Acad. Sci. U.S.A.* **38**, 747 (1952).
31. G. Elliott, P. O'Hare, *J. Virol.* **73**, 4110 (1999).
32. We thank Professor Rick E. Randall, University of St. Andrews for the lentivirus vectors. GLS is a Wellcome Trust Principal Research Fellow. This work was supported by the UK Medical Research Council.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1183173/DC1
Materials and Methods
Figs. S1 to S6
References
Movies S1 to S7

12 October 2009; accepted 8 January 2010
Published online 21 January 2010;
10.1126/science.1183173
Include this information when citing this paper.

Drive Against Hotspot Motifs in Primates Implicates the *PRDM9* Gene in Meiotic Recombination

Simon Myers,^{1,2*†} Rory Bowden,^{1,2*} Afidalina Tumian,¹ Ronald E. Bontrop,³ Colin Freeman,² Tammie S. MacFie,^{4‡} Gil McVean,^{1,2§} Peter Donnelly^{1,2§}

Although present in both humans and chimpanzees, recombination hotspots, at which meiotic crossover events cluster, differ markedly in their genomic location between the species. We report that a 13–base pair sequence motif previously associated with the activity of 40% of human hotspots does not function in chimpanzees and is being removed by self-destructive drive in the human lineage. Multiple lines of evidence suggest that the rapidly evolving zinc-finger protein *PRDM9* binds to this motif and that sequence changes in the protein may be responsible for hotspot differences between species. The involvement of *PRDM9*, which causes histone H3 lysine 4 trimethylation, implies that there is a common mechanism for recombination hotspots in eukaryotes but raises questions about what forces have driven such rapid change.

In humans and most other eukaryotes, meiotic crossover events typically cluster within narrow regions termed hotspots (1–5). Previously (6), we identified a degenerate 13–base pair (bp) motif, CCNCCNTNNCCNC, that is overrepresented in human hotspots. Both linkage disequilibrium (LD)–based analysis (6) and sperm typing at currently active hotspots (7)

implicated this motif in the activity of 40% of hotspots.

Despite nearly 99% identity at aligned bases, humans and chimpanzees show little if any sharing of hotspot locations (4, 5), although it has remained undetermined whether the recently identified hotspot motif is also active in the chimpanzee. To resolve this question, we collected chimpanzee genetic variation data at 22 loci where there is both an inferred hotspot at the orthologous location in humans and human-chimpanzee sequence conservation of the 13-nucleotide oligomer: 16 motifs within THE1 elements and 6 within L2 elements, chosen for their high activity of a particular “core” version of the motif in humans (fig. S1). We used the statistical software LDhat to estimate recombination rates separately in each region in different populations of both species (8). For humans, we used the Haplotype Map (HapMap) Phase II data. For chimpanzees, we genotyped 36 Western, 20 Central, and 17 Vel-lorosos chimpanzees at a total of 694 chimpanzee

single-nucleotide polymorphisms (SNPs), an average of 31.5 per region.

Because these regions are inferred human hotspots, the average estimated recombination rate surrounding the motif in humans showed a strong peak for both L2 and THE1 elements (Fig. 1A). In contrast, chimpanzees showed no evidence of increased recombination rates for either background. In Western chimpanzees, the THE1 estimated recombination rate around the motif was similar to the regional average, whereas a weak peak in mean rate for the L2 elements was produced solely by a single potential hotspot in one of the six regions (Fig. 1B). Results for the other chimpanzee subspecies were less informative (fig. S2) (8) but did not reveal a different pattern. To ensure that unknown haplotypic phase, smaller sample size, less dense data, and SNP ascertainment in chimpanzees had not compromised the ability to detect hotspots, we repeatedly sampled from the Centre d'Etude du Polymorphisme Humain (CEPH) from Utah (CEU) HapMap population data to produce human data sets comparable with those from chimpanzees in terms of these features (8). We conditioned only on the presence of the 13-nucleotide oligomer in THE1 and L2 elements and not the presence of a hotspot. This bootstrap technique revealed that the differences between human and chimpanzee rates cannot be explained by differences in power ($P = 0.00052$), although the signal was only significant for THE1 elements when analyzed separately ($P = 0.00012$) (fig. S3). These results provide evidence that the 13-nucleotide oligomer motif does not recruit hotspots in chimpanzees, implying changes in recombination machinery between humans and chimpanzees. The existence of factors capable of such changes in recombination genome-wide has been demonstrated in *Caenorhabditis elegans* (9) and by the mapping in mice of a trans-acting factor responsible for differences in hotspot location among inbred strain crosses (10, 11).

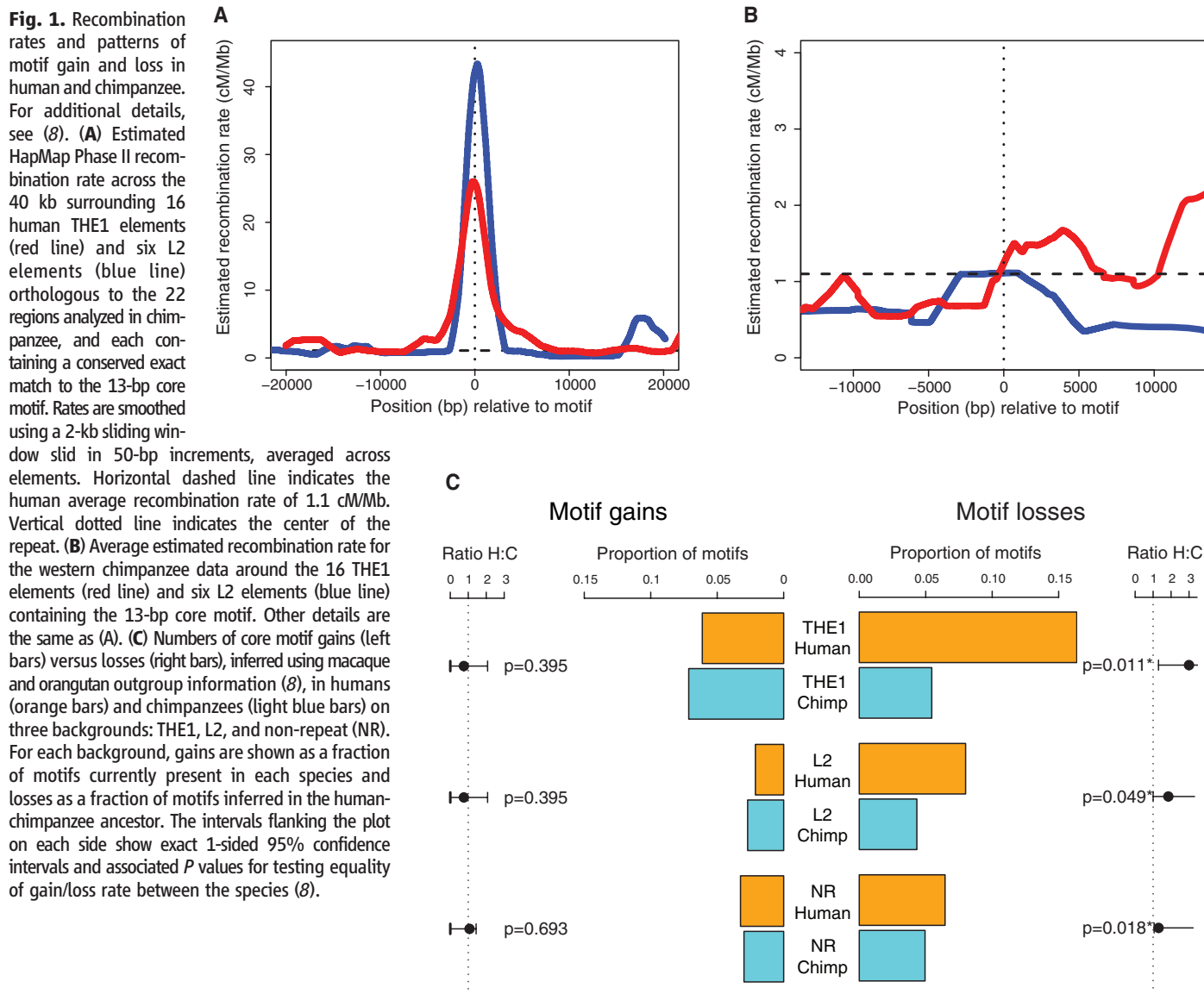
¹Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK. ²Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK. ³Department of Comparative Genetics and Refinement, Biomedical Primate Research Center, Lange Kleiweg 139 2288 GJ, Rijswijk, Netherlands. ⁴Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: myers@stats.ox.ac.uk

‡Present address: Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, 4 Newark Street, London E1 2AT, UK.

§These authors contributed equally to this work.



A separate process, predicted to cause a rapid evolution of individual hotspots, is the self-destructive drive inherent in double-strand break (DSB) formation, known as biased gene conversion (BGC) (12). Mutations reducing DSB formation in cis at recombination hotspots are preferentially transmitted as a consequence of repair of DSBs initiated on the other more recombinogenic strand in heterozygotes and are thus favored in a manner mimicking natural selection (13). This phenomenon could lead to rapid hotspot loss (14, 15). Direct evidence from sperm typing (16) has shown BGC at one polymorphic point mutation disrupting an occurrence of the 13-bp motif. More generally, BGC is predicted to eliminate copies of any recombination-promoting motif from the genome. The species-specific recombination activity of the 13-bp human hotspot motif suggests that losses of this motif should have occurred preferentially on the human lineage rather than that leading to chimpanzees.

To examine the evidence for BGC-driven motif loss, we therefore characterized rates and patterns of molecular evolution for the degenerate 13-nucleotide oligomer and the “core” version of the motif on specific backgrounds: THE1 elements, L2 elements, AluY/Sc/Sg elements (degenerate motif only), other repeats, and unique nonrepeat DNA (Table 1). We found a consistent substitution pattern imbalance, with chimpanzees having more copies of the motif than humans [empirical $P = 0.003$ for the most active form, with three of four independent backgrounds showing $P < 0.05$; $P = 0.002$ for the degenerate 13-nucleotide oligomer motif, with $P < 0.05$ for three of five individual backgrounds (8)]. As predicted by theoretical considerations of BGC [supporting online material (SOM) text and table S1] (14, 15), the magnitude of the imbalance was strongest for cases in which the motif has greatest activity. To assess whether motifs have been gained in chimpanzees or lost in humans, we

used the published macaque (17) and draft orangutan (18) genome sequences to infer ancestral sequence. For THE1 elements, L2 elements, and nonrepeat DNA, we observed an excess of human losses of the most active motif relative to chimpanzee ($P < 0.05$ in each case) (Fig. 1C and table S2) and similar results for the degenerate 13-nucleotide oligomer motif (table S3). The effect strength again correlates with hotspot activity. In contrast, there are no significant differences between species in motif gains ($P > 0.3$). Alu elements were not analyzed because of a high rate of uncertainty in inferring the ancestral base.

To determine whether motif activity has been lost on the chimpanzee lineage or gained on the human lineage, we compared our observations with a population-genetics model (SOM text) (14, 15). On the human lineage, approximately 16% of motifs on the THE1 and 8% on the L2 background have been lost in humans since human-chimpanzee divergence (Fig. 1C). If the

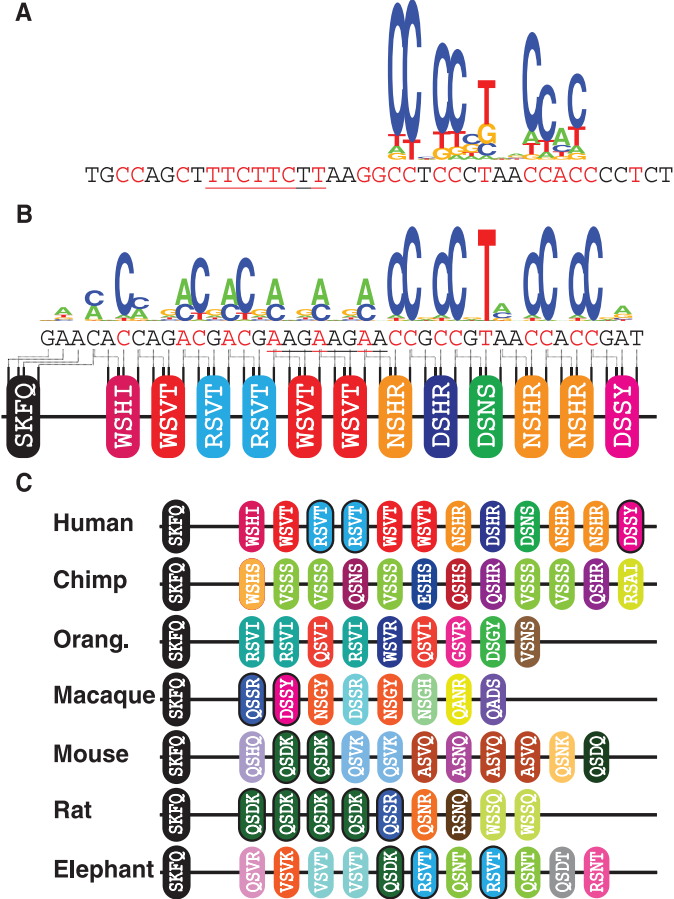
Table 1. Motif imbalance between human and chimpanzee. For the core motif and the degenerate motif, we analyzed cases in which the motif occurs in exactly one of human and chimpanzee. Results are shown for the full set of nonshared motifs and stratified into five backgrounds that differ in average human recombination activity. Significance levels are calculated in two ways: *P* values for

ratios are based on a one-sided exact binomial test of fewer human-only cases because the motif is known to be active in humans. Empirical *P* values are one-sided and obtained through comparisons of counts for the core or degenerate motif with counts observed for motifs of the same length and GC content on the same backgrounds (8). Dashes indicate zero counts in both species.

Sequence background	Core motif CCTCCCTNNCCAC				Degenerate motif CCNCCNTNNCCNC			
	Human only	Chimp only	Ratio (<i>P</i> value)	Empirical <i>P</i> value	Human only	Chimp only	Ratio (<i>P</i> value)	Empirical <i>P</i> value
All	425	515	1.21 (0.0018**)	0.0033**	19448	20245	1.04 (3.2e-05**)	0.0020**
THE1	20	39	1.95 (0.0092**)	0.0050**	50	76	1.52 (0.0128*)	0.0093**
L2	30	47	1.57 (0.0338*)	0.0307*	432	496	1.15 (0.0193*)	0.0219*
AluY,Sc,Sg	—	—	—	—	3642	3924	1.08 (0.0006**)	0.1119
Other repeats	99	131	1.32 (0.0204*)	0.0346*	10126	10254	1.01 (0.1868)	0.4373
Non-repeats	276	298	1.08 (0.2135)	0.2206	5198	5495	1.06 (0.0021**)	0.0215*

P* < 0.05. *P* < 0.01.

Fig. 2. (A) Previously estimated degeneracy of the 13-bp hotspot motif (logo plot; relative letter height proportional to estimated probability of hotspot activity and total letter height determined by degree of base specificity) (6) as well as an extended ~39-bp motif [text below logo, with influential positions (*P* < 0.01) shown in red]. (B) In silico prediction of the binding consensus for PRDM9, aligned with the 13-nucleotide oligomer, with more influential positions shown in red. Underlined in both (A) and (B) is an additional 8-bp matching sequence. The logo shows predicted degeneracy within this consensus (8). Below the text is the sequence of four predicted DNA-contacting amino acids for the 13 successive human PRDM9 zinc fingers (one oval per finger, differing colors for differing fingers, and the separated finger is gapped N-terminal from others) and their predicted base contacts within the motif. (C) Sequence of four predicted DNA-contacting amino acids for the PRDM9 zinc fingers in seven mammalian species, presented as in (B). Distinct fingers are given different colors; fingers present in at least two species have a black border.



zinc-finger proteins, the 13-nucleotide oligomer motif was present within the predicted binding sequence of five (fig. S4). Binding specificity was then further explored in silico by comparing predicted motif degeneracy for each candidate (inferred by calculating the relative binding score for every 1-bp mutation relative to the consensus) with empirical degeneracy patterns in the 13-bp motif (Fig. 2A). Predictions for one of the candidates, PRDM9, exactly matched the observed degeneracy at positions 3, 6, 8, 9, and 12 within the 13-bp motif (Fig. 2B) and lack of degeneracy at the other eight positions. Predictions for the other four candidates showed features inconsistent with the observed degeneracy (fig. S4). The predicted binding sequence for PRDM9 also contains an exact match on the opposite strand for an 8-bp region of the extended motif, upstream of the 13-bp degenerate motif, perhaps suggesting that PRDM9 zinc fingers might contact both DNA strands. Finally, the number of zinc fingers (13) in this protein, the positioning of the match to the 13-bp motif within the longer predicted binding sequence, and strong influence of this 13-bp region on specificity all match our previous predictions (6). The lack of activity of the 13-bp motif in chimpanzees demonstrated above suggests that in addition to having the predicted binding specificity, any motif-binding protein candidate should also show differences between humans and chimpanzees. For four of the five candidates, the predicted DNA-contacting amino acids within the zinc fingers are identical between human and chimpanzee. Chimpanzee PRDM9, however, has a dramatically different predicted binding sequence (fig. S5). Although PRDM9 has multiple zinc fingers in both species (12 and 13 respectively), the DNA-contacting residues -1, 2, 3, and 6 are only shared between species in the first finger (Fig. 2C). Such rapid evolution is exceptional. Comparing these residues among all 544 C2H2-containing zinc-finger protein human-chimpanzee ortholog pairs, PRDM9 is the most diverged (*P* = 0.0018). The PRDM9 sequences in five additional mammals (elephant, mouse, rat, macaque and orangutan) exhibit rapid evolution, variation in zinc-finger number (between 8 and

12), and patterns of substitution suggestive of complex repeat shuffling (Fig. 2C) (20).

Multiple lines of evidence point to a role for the orthologous mouse gene, *Prdm9*, in recombination. *Prdm9* lies within a 5.1-Mb region that contains a locus that influences genome-wide hotspot locations (10, 11) and is exclusively expressed during meiotic prophase, with mice in which *Prdm9* has been knocked out showing infertility and failure to properly repair DSBs (21). Mouse PRDM9 trimethylates lysine 4 of histone H3 (H3K4me3) (21), an epigenetic mark specifically enriched on mouse chromatids carrying recombination initiation sites within the mouse hotspot *Psmb9* (22). In yeast, mutation of the sole gene, *Set1*, encoding H3K4me3 reduces crossover activity at 84% of hotspots (23). The lack of well-defined target-sequence specificity of Set1 (which is not a zinc-finger protein) may indicate why no dominant hotspot motif has been identified in yeast. Intriguingly, *Prdm9* is also the only species-incompatibility gene yet identified in mouse (24), with differences among nine PRDM9 zinc fingers between mouse strains potentially playing a causal role in male sterility.

Baudat *et al.* find that variation in *PRDM9* among humans correlates with variability in genome-wide hotspot usage, and PRDM9 binds the 13-bp motif in a sequence-specific manner in vitro (25). The findings of both studies imply that PRDM9 determines human hotspot locations, with PRDM9 evolution explaining lack of hotspot conservation in other species. Exactly how PRDM9 functions, for example through altering transcription of DSB repair genes or directly recruiting DSB repair proteins, remains unknown. These findings also raise the question of why such an important gene is evolving so rapidly. The

DNA sequence of the zinc-finger array of PRDM9 constitutes a coding minisatellite, suggesting a high intrinsic mutation rate resulting from repeat instability. However, patterns of evolution within the zinc-finger array, notably the clustering and coordination of changes at sites that interact with DNA bases, strongly suggest positive selection on binding specificity (20). Selection could possibly arise from the gradual degradation of hotspots through BGC, leading to a loss in fitness either through the promotion of deleterious alleles within hotspots (15) or through having insufficient crossover events to support proper disjunction (14, 15). Alternatively, the rapid evolution of *PRDM9* could be indicative of genetic conflict, such as meiotic drive or conflict involving mobile elements (26, 27). Although there is no direct evidence for this, mouse *Prdm9* lies within one of the inversions characterizing the meiotic-drive *t*-complex (28).

References and Notes

1. J. L. Gerton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11383 (2000).
2. G. A. McVean *et al.*, *Science* **304**, 581 (2004).
3. K. Paigen *et al.*, *PLoS Genet.* **4**, e1000119 (2008).
4. S. E. Ptak *et al.*, *Nat. Genet.* **37**, 429 (2005).
5. W. Winckler *et al.*, *Science* **308**, 107 (2005).
6. S. Myers, C. Freeman, A. Auton, P. Donnelly, G. McVean, *Nat. Genet.* **40**, 1124 (2008).
7. A. J. Webb, I. L. Berg, A. Jeffreys, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10471 (2008).
8. Materials and methods are available as supporting material on Science Online.
9. D. G. Mets, B. J. Meyer, *Cell* **139**, 73 (2009).
10. E. D. Parvanov, S. H. Ng, P. M. Petkov, K. Paigen, M. Lichten, *PLoS Biol.* **7**, e36 (2009).
11. C. Grey, F. Baudat, B. de Massy, M. Lichten, *PLoS Biol.* **7**, e35 (2009).
12. A. Nicolas, D. Treco, N. P. Schultes, J. W. Szostak, *Nature* **338**, 35 (1989).
13. T. Nagylaki, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6278 (1983).

14. G. Coop, S. R. Myers, *PLoS Genet.* **3**, e35 (2007).
15. A. Boulton, R. S. Myers, R. J. Redfield, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8058 (1997).
16. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267 (2002).
17. R. A. Gibbs *et al.*, Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
18. <http://genome.ucsc.edu/cgi-bin/hgGateway?clade=mammal&org=Orangutan&db=0> (2007).
19. A. V. Persikov, R. Osada, M. Singh, *Bioinformatics* **25**, 22 (2009).
20. P. L. Oliver *et al.*, *PLoS Genet.* **5**, e1000753 (2009).
21. K. Hayashi, K. Yoshida, Y. Matsui, *Nature* **438**, 374 (2005).
22. J. Buard, P. Barthès, C. Grey, B. de Massy, *EMBO J.* **28**, 2616 (2009).
23. V. Borde *et al.*, *EMBO J.* **28**, 99 (2009).
24. O. Mihola, Z. Trachtulec, C. Vlcek, J. C. Schimenti, J. Forejt, *Science* **323**, 373 (2009).
25. F. Baudat *et al.*, *Science* **327**, 836 (2010); published online 31 December 2009 (10.1126/science.1183439).
26. D. C. Presgraves, *Bioessays* **29**, 386 (2007).
27. G. J. Wyckoff, W. Wang, C. I. Wu, *Nature* **403**, 304 (2000).
28. J. Schimenti, *Trends Genet.* **16**, 240 (2000).
29. We thank N. Mundy for advice and provision of chimpanzee samples and C. Mitchell and E. Nerrienet for assisting in chimpanzee sample collection. We thank D. Falush and G. Coop for helpful conversations. Part of the work was completed while S.M. was a fellow at the Broad Institute. We would like to acknowledge funding from the Leverhulme Trust (to G.M.), the Royal Society (to P.D.), and the Wellcome Trust (to S.M., C.F., G.M., and P.D.). The chimpanzee genotype data for 22 hotspot candidate regions generated is available online with the SOM as Database S1.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1182363/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2
References
Database S1

23 September 2009; accepted 17 December 2009
Published online 31 December 2009;
10.1126/science.1182363
Include this information when citing this paper.

The *Lmo2* Oncogene Initiates Leukemia in Mice by Inducing Thymocyte Self-Renewal

Matthew P. McCormack,^{1,2*} Lauren F. Young,¹ Sumitha Vasudevan,¹ Carolyn A. de Graaf,³ Rosalind Codrington,⁴ Terence H. Rabbitts,⁴ Stephen M. Jane,^{1,2} David J. Curtis^{1,2}

The *LMO2* oncogene causes a subset of human T cell acute lymphoblastic leukemias (T-ALL), including four cases that arose as adverse events in gene therapy trials. To investigate the cellular origin of *LMO2*-induced leukemia, we used cell fate mapping to study mice in which the *Lmo2* gene was constitutively expressed in the thymus. *Lmo2* induced self-renewal of committed T cells in the mice more than 8 months before the development of overt T-ALL. These self-renewing cells retained the capacity for T cell differentiation but expressed several genes typical of hematopoietic stem cells (HSCs), suggesting that *Lmo2* might reactivate an HSC-specific transcriptional program. Forced expression of one such gene, *Hhex*, was sufficient to initiate self-renewal of thymocytes in vivo. Thus, *Lmo2* promotes the self-renewal of preleukemic thymocytes, providing a mechanism by which committed T cells can then accumulate additional genetic mutations required for leukemic transformation.

Despite significant improvements in the treatment of T cell acute lymphoblastic leukemia (T-ALL), about 20% of chil-

dren and the majority of adults succumb to resistant or relapsed disease (1). A common cause of T-ALL is the overexpression of onco-

genic transcription factors during T cell development in the thymus, resulting from chromosomal translocations or deletions (2). One such oncogene is *LMO2*, whose expression is activated in about 9% of pediatric T-ALL cases (3). *LMO2* is a member of the LMO (LIM-only) class of transcription factors that contain zinc-binding finger-like motifs termed LIM domains, which function in protein-protein interactions (4). *Lmo2* does not bind DNA directly but mediates transcriptional activation and repression by binding to other transcription factors to mediate the formation of multimeric complexes (5).

¹Rotary Bone Marrow Research Laboratories, Royal Melbourne Hospital, Grattan Street, Parkville, Victoria 3050, Australia.

²Department of Medicine, University of Melbourne, Parkville, Victoria 3010, Australia. ³The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3050, Australia, and Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia. ⁴Leeds Institute of Molecular Medicine, Wellcome Trust Brenner Building, St. James's University Hospital, Leeds LS9 7TF, UK.

*To whom correspondence should be addressed. E-mail mccormack@wehi.edu.au