Drive Against Hotspot Motifs in Primates Implicates the \textit{PRDM9} Gene in Meiotic Recombination
Simon Myers, et al.
\textit{Science} \textbf{327}, 876 (2010);
DOI: 10.1126/science.1182363

\textit{This copy is for your personal, non-commercial use only.}

\textbf{If you wish to distribute this article to others}, you can order high-quality copies for your colleagues, clients, or customers by \textbf{clicking here.}

\textbf{Permission to republish or repurpose articles or portions of articles} can be obtained by following the guidelines \textbf{here.}

\textbf{The following resources related to this article are available online at www.sciencemag.org (this information is current as of October 19, 2011):}

\textbf{Updated information and services}, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/327/5967/876.full.html

\textbf{Supporting Online Material} can be found at:
http://www.sciencemag.org/content/suppl/2009/12/30/science.1182363.DC1.html

A list of selected additional articles on the Science Web sites \textbf{related to this article} can be found at:
http://www.sciencemag.org/content/327/5967/876.full.html#related

This article \textbf{cites 25 articles}, 9 of which can be accessed free:
http://www.sciencemag.org/content/327/5967/876.full.html#ref-list-1

This article has been \textbf{cited by 4 article(s)} on the ISI Web of Science

This article has been \textbf{cited by 14 articles} hosted by HighWire Press; see:
http://www.sciencemag.org/content/327/5967/876.full.html#related-urls
poorly understood, and the elucidation of such mechanisms could lead to the discovery of novel therapeutics.

References and Notes
11. Materials and methods are available as supporting material on Science online.
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,† Rory Bowden, Afidalina Tumian, Ronald E. Bontrop, Colin Freeman, Tammie S. MacFie,† Gil McVean, Peter Donnelly,*§

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, CCNCCNNTNCCNC, 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 Velvorous 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) (6) 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 Caenohabditis 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).
Fig. 1. Recombination rates and patterns of motif gain and loss in human and chimpanzee. For additional details, see (8). (A) Estimated HapMap Phase II recombination rate across the 40-kb surrounding 16 human THE1 elements (red line) and six L2 elements (blue line) containing the 13-bp core motif. Rates are smoothed using a 2-kb sliding window slid in 50-bp increments, averaged across elements. Horizontal dashed line indicates the center of the repeat. (B) Average estimated recombination rate for the western chimpanzee data around the 16 THE1 elements (red line) and six L2 elements (blue line) containing the 13-bp core motif. Other details are the same as (A). (C) Numbers of core motif gains (left bars) versus losses (right bars), inferred using macaque and orangutan outgroup information (8), in humans (orange bars) and chimpanzees (light blue bars) on three backgrounds: THE1, L2, and non-repeat (NR). For each background, gains are shown as a fraction of motifs currently present in each species and losses as a fraction of motifs inferred in the human-chimpanzee ancestor. The intervals flanking the plot on each side show exact 1-sided 95% confidence intervals and associated P values for testing equality of gain/loss rate between the species (8).

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
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

<table>
<thead>
<tr>
<th>Sequence background</th>
<th>Core motif CCTCCCTNNCCAC</th>
<th>Degenerate motif CCNCCNNTNCNCC</th>
<th>Human only</th>
<th>Chim only</th>
<th>Ratio (( P ) value)</th>
<th>Empirical ( P ) value</th>
<th>Human only</th>
<th>Chim only</th>
<th>Ratio (( P ) value)</th>
<th>Empirical ( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>425</td>
<td>515</td>
<td>1.21 (0.0018**)</td>
<td>0.0033**</td>
<td>1.04 (3.2e-05**)</td>
<td>0.0020**</td>
<td>19448</td>
<td>20245</td>
<td>1.04 (3.2e-05**)</td>
<td>0.0020**</td>
</tr>
<tr>
<td>THE1</td>
<td>20</td>
<td>39</td>
<td>1.95 (0.0092**)</td>
<td>0.0505**</td>
<td>1.52 (0.0128*)</td>
<td>0.0093**</td>
<td>50</td>
<td>76</td>
<td>1.52 (0.0128*)</td>
<td>0.0093**</td>
</tr>
<tr>
<td>L2</td>
<td>40</td>
<td>47</td>
<td>1.57 (0.0338*)</td>
<td>0.0307*</td>
<td>1.15 (0.0193*)</td>
<td>0.0219*</td>
<td>432</td>
<td>496</td>
<td>1.15 (0.0193*)</td>
<td>0.0219*</td>
</tr>
<tr>
<td>AluY,Sc,Sg</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.08 (0.0006**)</td>
<td>0.1119</td>
<td>3642</td>
<td>3924</td>
<td>1.08 (0.0006**)</td>
<td>0.1119</td>
</tr>
<tr>
<td>Other repeats</td>
<td>99</td>
<td>131</td>
<td>1.32 (0.0204*)</td>
<td>0.0346*</td>
<td>1.01 (0.1868)</td>
<td>0.4373</td>
<td>10126</td>
<td>10254</td>
<td>1.01 (0.1868)</td>
<td>0.4373</td>
</tr>
<tr>
<td>Non-repeats</td>
<td>276</td>
<td>298</td>
<td>1.08 (0.2135)</td>
<td>0.2206</td>
<td>1.06 (0.0021**)</td>
<td>0.0215*</td>
<td>5198</td>
<td>5495</td>
<td>1.06 (0.0021**)</td>
<td>0.0215*</td>
</tr>
</tbody>
</table>

\* \( P \) < 0.05, \** \( P \) < 0.01.

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

\[ P = \frac{\text{number of observed differences}}{\text{expected number of differences}} \]

\[ \text{Degeneracy} = \frac{\text{number of observed differences}}{\text{number of possible differences}} \]

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.

We next investigated the function of the 13-nucleotide oligomer motif. Previously, we suggested that the human hotspot motif was probably bound by a zinc finger protein with at least 12 zinc fingers, on the basis of an extended 30- to 40-bp region of weaker sequence specificity containing the motif and a 3-bp periodicity of influential bases (6). We therefore set out to identify candidates for such a protein using a computational algorithm that predicts DNA binding specificity for C2H2 zinc-finger proteins (19). Among the 691 identified human C2H2 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 51-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 chromatin 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-incompatible 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 l-complex (28).

References and Notes
8. Materials and methods are available as supporting material on Science Online.
25. F. Baudat et al., Science 327, 836 (2010); published online 31 December 2009 (10.1126/science.1183439).
29. We thank N. Mundy for advice and provision of chimpanzee samples and C. Mitchell and E. Nerrieter for assisting in chimpanzee sample collection. We thank D. Faush 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. Curtis1,2

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

www.sciencemag.org SCIENCE VOL 327 12 FEBRUARY 2010 879