Supplemental Materials for

Evolution of local mutation rate and its determinants

Nadezhda V. Terekhanova,[*,1,2,5] Vladimir B. Seplyarskiy,[*,2,3,5] Ruslan A. Soldatov,[1,2] Georgii A. Bazykin[1,2,3,4]

Contents

Supplemental Text 4

Text S1. LMRs estimated from 100Kb genomic windows 4

Text S2. Statistical power of mLMR 4

Text S3. Exclusion of GC-biased gene conversion-prone mutations improves estimates of LMR 5

Text S4. The variance in LMR explained by the human genomic features obtained from different tissues 5

Text S5. Exceptionally high correlation between human and gibbon LMRs 6

Text S6. The association between changes in recombination and changes in LMR is not due to GC-biased gene conversion 6

Figure S1. pLMR variation in the human lineage, and dLMR variation in primate species, explained by the mLMR in the human lineage. 8

Figure S2. dLMR variation in primate species and mouse explained by genomic features and dLMR (A and C) or pLMR (B and D) in the human lineage. 9

Figure S3. Local mutation rate (LMR) variation in primate species and mouse explained by genomic features and dLMR (A and C) or pLMR (B and D) in the human lineage, for alignment columns without gaps or ambiguous nucleotides in any of the species. 10

Figure S4. The distribution of the correlation coefficients between the phylogenetic distance from the human and the fraction of the variance in dLMR explained by the mLMR, obtained by bootstrapping. 11

Figure S5. Scatterplots for raw correlations of LMRs in primate species and human genomic features. 16
Figure S6. Heatmap for the absolute values of correlation coefficients between all possible pairs of human LMR and genomic features.. 17
Figuew S7. Distribution of the correlation coefficients between the phylogenetic distance from the human and the fraction of the variance in dLMR explained by each genomic feature (ANOVA Type III analysis). 18
Figure S8. Densities of human recombination scores (A-B) in HDRs (A) and CDRs (B) and densities of chimpanzee recombination scores (C-D) in HDRs (C) and CDRs (D). 19
Figure S9. Variance of the LMR explained by the genomic features obtained from different tissues. 21
Figure S10. LMR variation in primate species explained by genomic features and dLMR in the human lineage. 22
Figure S11. Densities of human recombination scores (A-B) in HARs (A) and CARs (B) and densities of chimpanzee recombination scores (C-D) in HARs (C) and CARs (D) on the data without excluding W↔S substitutions. 23
Figure S12. Densities of human recombination scores (A-B) in HDRs (A) and CDRs (B) and densities of chimpanzee recombination scores (C-D) in HDRs (C) and CDRs on the data without excluding W↔S substitutions.. 24
Figure S13. Densities of human gBGC scores in HARs (A), HDRs (B), CARs (C), and CDRs (D) on the data without excluding W↔S substitutions. 25
Figure S14. Correlation of the observed and bootstrapped LMRs in 1Mb genomic windows. 26
Figure S15. Local mutation rate variation in primate species explained by genomic features and LMR in the human lineage. 27
Figure S16. LMR explained by individual genomic features. 28
Figure S17. Local mutation rate (LMR) variation in primate species explained by genomic features and LMR in the human lineage in 100Kb windows (A), and with the addition of mouse (B). 29

Figure S18. Variation in human LMR explained by the estimate of the same variable based on a subsample of mutations. 30

Figure S19. dLMR variation in primate species explained by genomic features and pLMR in the human lineage. 31

Figure S20. Correlation of the human pLMR and gibbon dLMR with (black line) and without (blue line) exclusion of the 8 chromosome 1 Mb genomic windows. 32

Figure S21. Local mutation rate variation in primate species explained by genomic features and LMR in the human lineage. 33

Figure S22. Densities of human BGC scores in HARs (A), HDRs (B), CARs (C), and CDRs (D). 34

Supplemental Tables 35

Table S1. The fraction of the variance in human and non-human dLMR explained by the features obtained from different tissue types. 35

Table S2. Average human and chimpanzee dLMRs in the HDRs, CDRs, HARs and CARs. 36

Supplemental Text

Text S1. LMRs estimated from 100Kb genomic windows

To test the robustness of our results to the choice of the window size, we repeated the analyses of the correlation between the LMRs for the 23,551 and 16,449 100Kb genomic windows from primate and primate-mouse alignments respectively.

The correlation between the LMRs of human and chimpanzee in 100 Kb windows (R^2=0.36, $P < 2.2×10^{-16}$; Supplemental Fig. S17) was lower than that observed in 1Mb windows (see main text), probably due to a higher random component of the variance in smaller windows. The decay of this correlation with phylogenetic distance was similar to that observed for 1Mb windows: the minimal value among primates was observed in the marmoset (R^2=0.11, $P < 2.2×10^{-16}$, Supplemental Fig. S17A), and was even lower in mouse (R^2=0.06, $P < 2.2×10^{-16}$ for 100Kb windows; Supplemental Fig. S17B).

Text S2. Statistical power of mLMR

While the mLMR is the most direct proxy for the LMR, its estimates are less robust than those of pLMR and dLMR because there are fewer events. For example, in humans, our analysis included a total of 636,954 mutations for dLMR and 1,093,223 mutations for pLMR, but only 11,429 mutations for mLMR. The correlation between the mLMR and the dLMR, and between the mLMR and the pLMR, even in the same species is therefore expected to be low simply due to the low sample size.

To show this, we subsampled the nucleotides divergent (i.e., contributing to dLMR) or polymorphic (i.e., contributing to pLMR) in our sample, and asked how well the subsamples explained the variance observed in the entire sample. With lower sample sizes, the R^2 values also were lower (Supplemental Fig. S18). When a subsample size of 11,429 mutations was used, the R^2 for dLMR or pLMR subsample vs. the entire sample was ~0.24. Therefore, a low correlation between the mLMR and dLMR, and between mLMR and pLMR, even in the same species is partially explainable by the reduced sample size.

Text S3. Exclusion of GC-biased gene conversion-prone mutations improves estimates of LMR

dLMR and even pLMR estimates can be confounded by the differential contribution of gene conversion in different genomic regions. This effect can be singled out using the fact that gene conversion is GC-biased (gBGC), i.e., favors fixation of strong (S: G or C) over weak (W: A or T) alleles (Duret and Arndt 2008; Duret and Galtier 2009; Glemin, et al. 2015). As expected, exclusion of W↔S substitutions slightly increases the correlation between dLMR and mLMR, and between pLMR and mLMR (Supplemental Fig. S1). Overall, the correlation between the pLMR and dLMR is rather low (Supplemental Fig. S19, A and C), but is substantially increased when the W↔S substitutions are excluded (Fig. 1). Interestingly, to increase the correlation between the pLMR and dLMR, it is sufficient to exclude the W↔S substitutions from the divergence data (Supplemental Fig. S19, B and D), and subsequent exclusion of such substitutions from the human polymorphism data adds little to the correlation, suggesting that gBGC impacts dLMR more than pLMR. In the main text, we exclude W↔S substitutions in all analyses.

Text S4. The variance in LMR explained by the human genomic features obtained from different tissues

The feature annotation in the main text is based on the embryonic stem cells. To study how our results depend on the properties of the specific tissue, we also explained the variance in human and non-human LMR by genomic features obtained from five other tissues. The obtained results (Supplemental Table S1) were very similar to those from the embryonic stem cells. A previous analysis (Schuster-Bockler and Lehner 2012) based on the GM12878 lymphoblastoid cell line explained ~28% of the variance in human LMR. We were able to explain 36% of the variance with a smaller set of features for this tissue (Supplemental Table S1). The difference is likely due to differences in genome assemblies (hg18 in ref. (Schuster-Bockler and Lehner 2012), hg19 in our analysis), details of alignment (divergence data was extracted from EPO whole-genome alignments in ref. (Schuster-Bockler and Lehner 2012) and from the UCSC genome browser in our study), masking and/or inference of LMR.

Additionally, we performed ANOVA type III analyses to infer the individual impacts of the genomic features on the explained variance in LMRs (Supplemental Fig. S9). We obtained nearly identical results to those in Fig. 2A. In particular, in all tissues, the contribution of recombination decreased with the phylogenetic distance.

Text S5. Exceptionally high correlation between human and gibbon LMRs

To understand the cause of the unexpectedly high correlation between the LMRs of human and gibbon, we plotted their LMRs for all 1MB genomic windows (Supplemental Fig. S20). We found that the high correlation is largely driven by the few strongly mutated windows in the ~15-megabase segment of the human chromosome 8 (region 8p), which is known to have mutated particularly rapidly in humans (Nusbaum, et al. 2006). For unknown reasons, this region has also mutated particularly rapidly in the gibbon (Supplemental Fig. S20). Exclusion of chromosome 8 reduced the human-gibbon LMRs correlation from ~0.65 to ~0.55 (see Supplemental Fig. S21).

Text S6. The association between changes in recombination and changes in LMR is not due to GC-biased gene conversion

To validate the association between the changes in LMR and changes in recombination, we additionally used the human gBGC rate (Capra, et al. 2013; Glemin, et al. 2015) as a proxy for the recombination rate. The human gBGC scores were significantly elevated in the HARs and CARs, compared with the genome average (one-sided Wilcoxon rank sum test, $P=7.34\times10^{-9}$ and P=0.04, respectively; Supplemental Fig. S22, A and C). Conversely, gBGC scores were decreased in the HDRs and CDRs (P=0.007 and P=0.006, respectively; Supplemental Fig. S22, B and D). Therefore, the gBGC data supports the role of species-specific recombination hotspots in acceleration of the LMR.

Conceivably, the association between the recombination and mutation rates could arise from the direct effect of gBGC (Duret and Arndt 2008) which confounds the substitution spectrum by favoring the fixation of W→S mutations. However, the higher recombination in the regions with the elevated species-specific LMR is not explained by excessive gBGC, as it is also observed when W→S and S→W nucleotide substitutions are excluded, both when recombination scores (Fig. 3 and Supplemental Figs. S8, S11-S12) and gBGC scores (Supplemental Figs. S13, 22) are used as proxy for recombination rate. This implies that the effect of recombination on LMR is not confined to the effect of biased nucleotide changes through gBGC. Therefore, the role of recombination dependent mutagenesis in the species-specific changes in LMR is supported by two different approaches.

Supplemental Figures

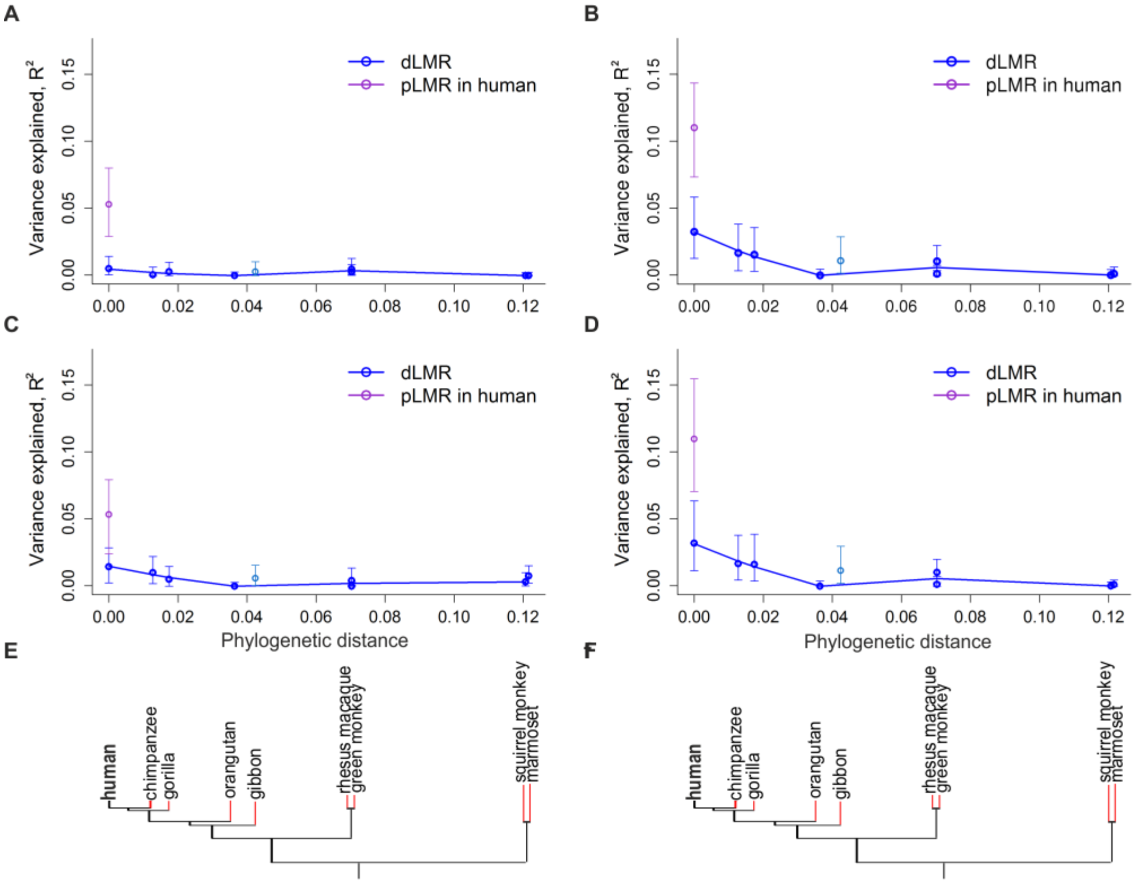


Figure S1. pLMR variation in the human lineage, and dLMR variation in primate species, explained by the mLMR in the human lineage.

(A-B) W↔S substitutions included; (C-D) W↔S substitutions excluded. (A and C) repeats excluded (same dataset as in the Main text); (B and D) repeats not excluded. Including the regions with repeats yielded 63% more de novo mutations, and stronger correlations. (E-F) Phylogenetic tree of the considered species. Notations are as in Figure 1.

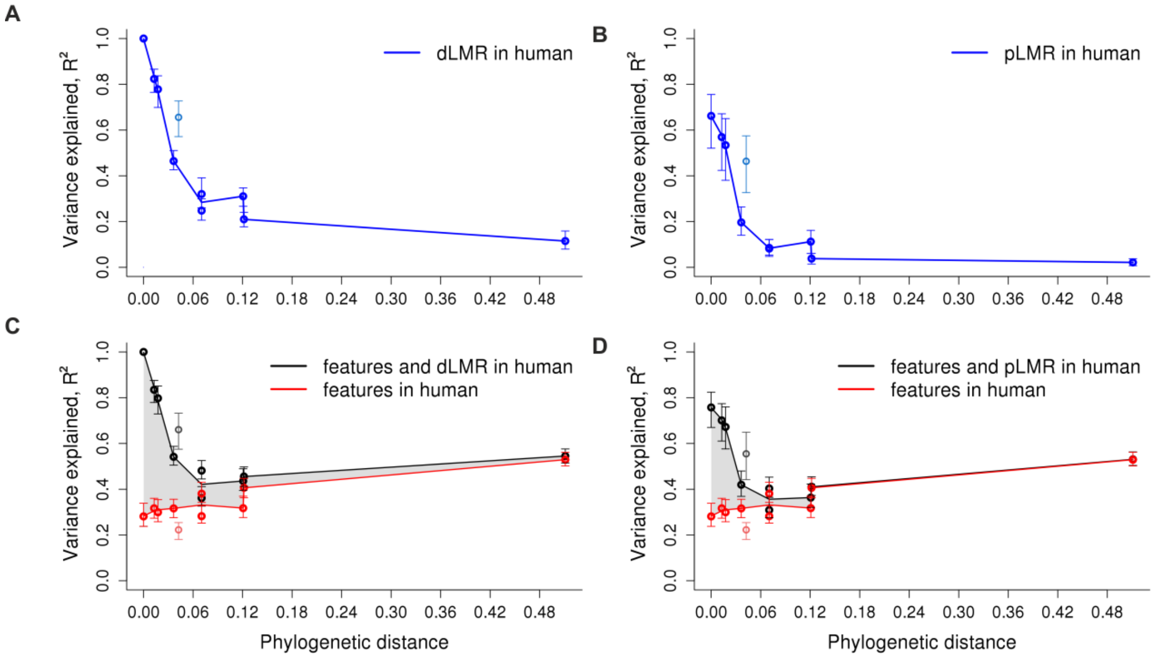


Figure S2. dLMR variation in primate species and mouse explained by genomic features and dLMR (A and C) or pLMR (B and D) in the human lineage.

Notations are as in Figure 1. Species are the same as in Figure 1, with the addition of mouse (the rightmost data point). The 9 genomic features included in the model are: recombination, MAF, DHSs, replication, GC-content, exonic density, H3K9me3, H3K27ac, H3K27me3.

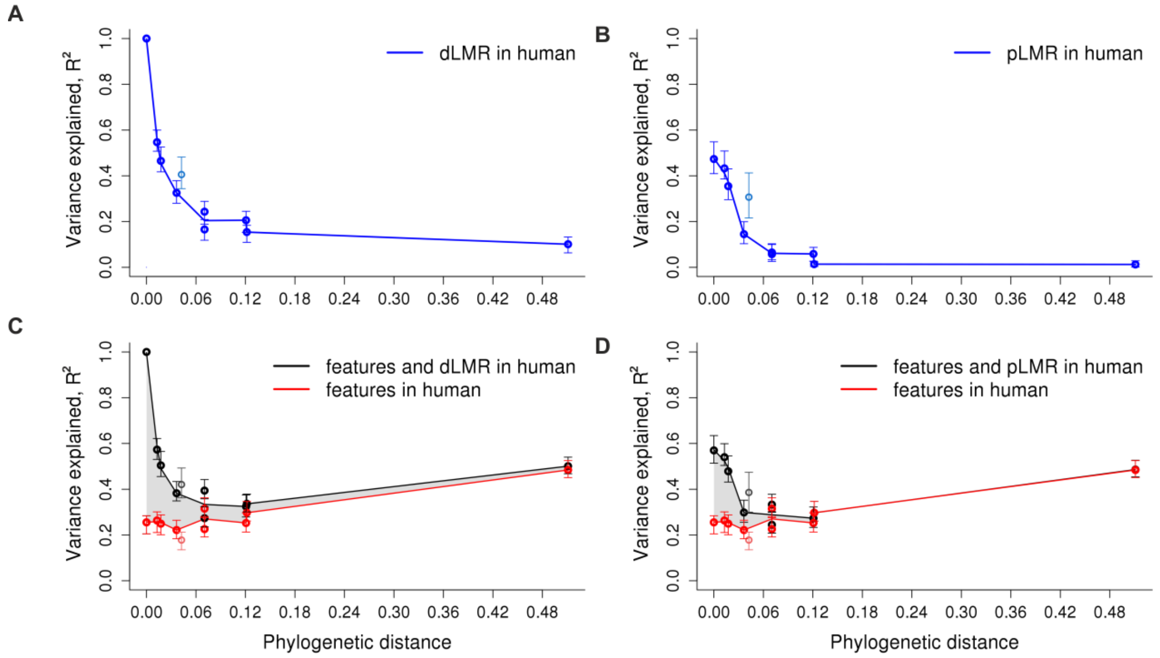


Figure S3. Local mutation rate (LMR) variation in primate species and mouse explained by genomic features and dLMR (A and C) or pLMR (B and D) in the human lineage, for alignment columns without gaps or ambiguous nucleotides in any of the species.

Notations are as in Figure 1. Species are the same as in Figure 1, with the addition of mouse (the rightmost data point). Genomic features included in the model are the same as in Supplemental Figure S2.

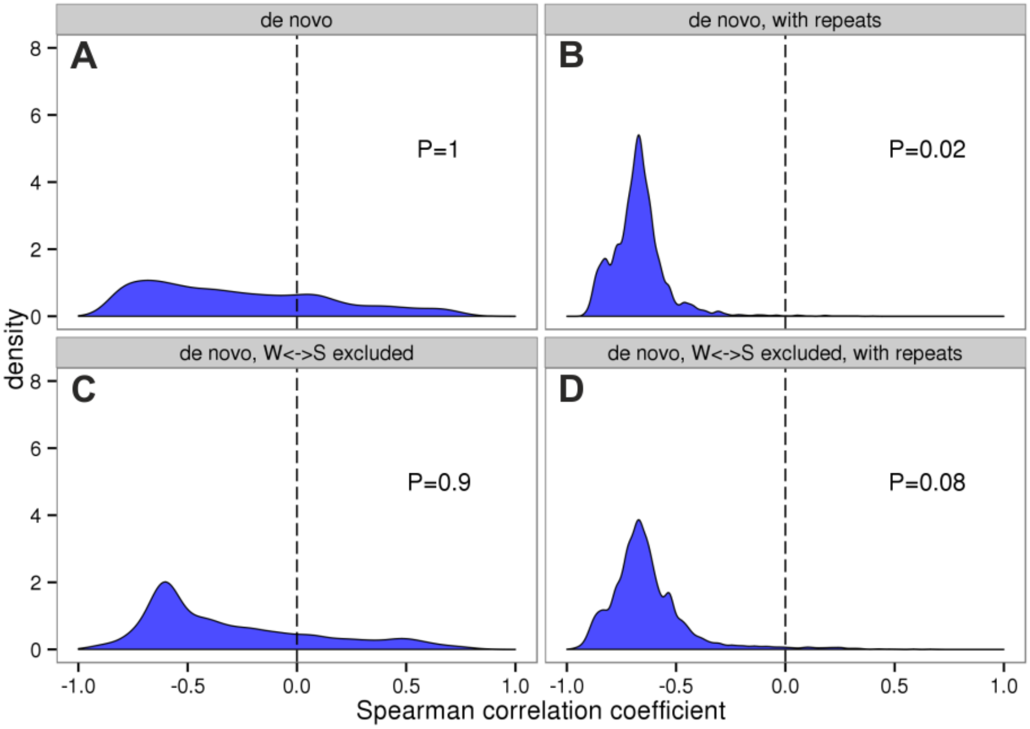
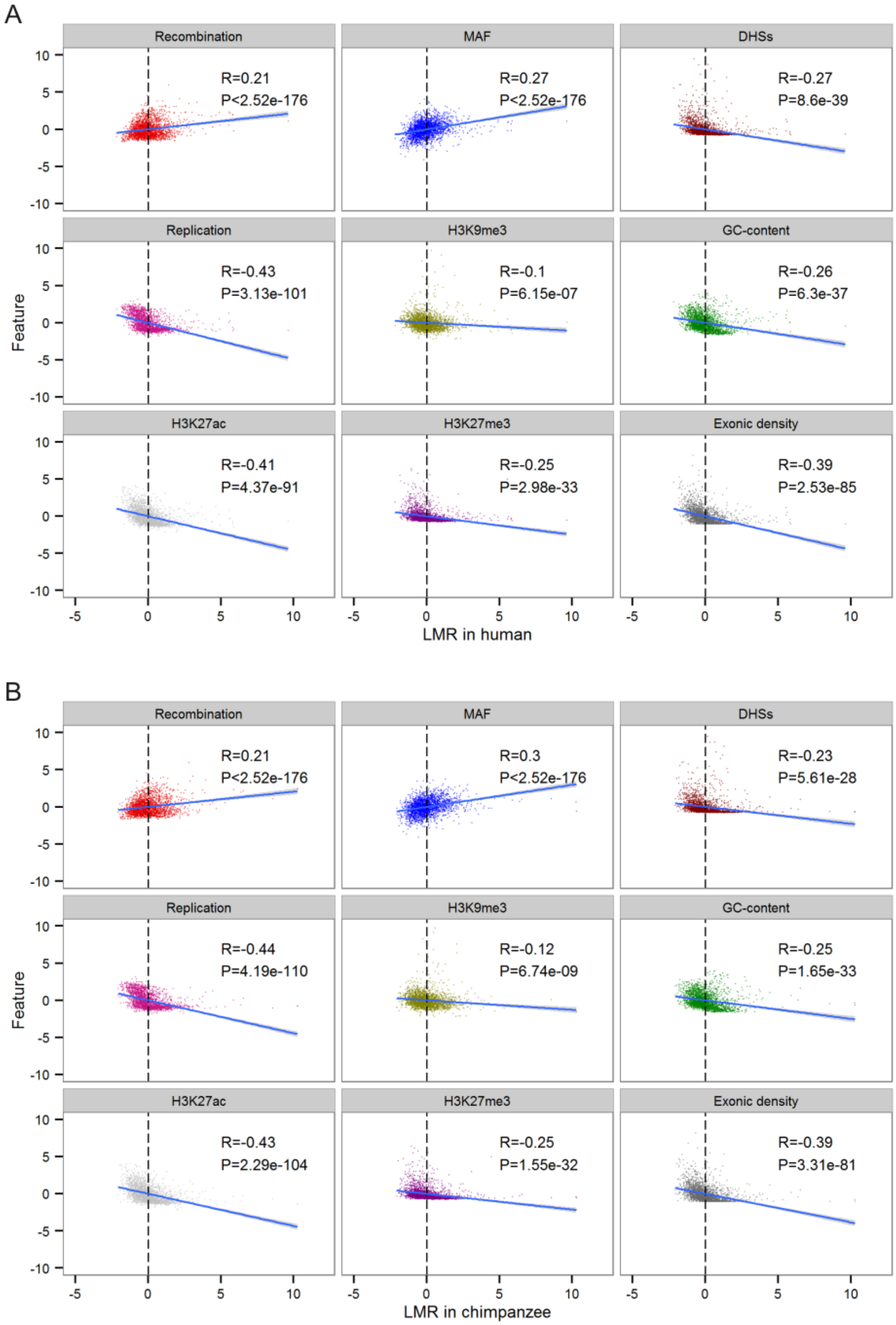
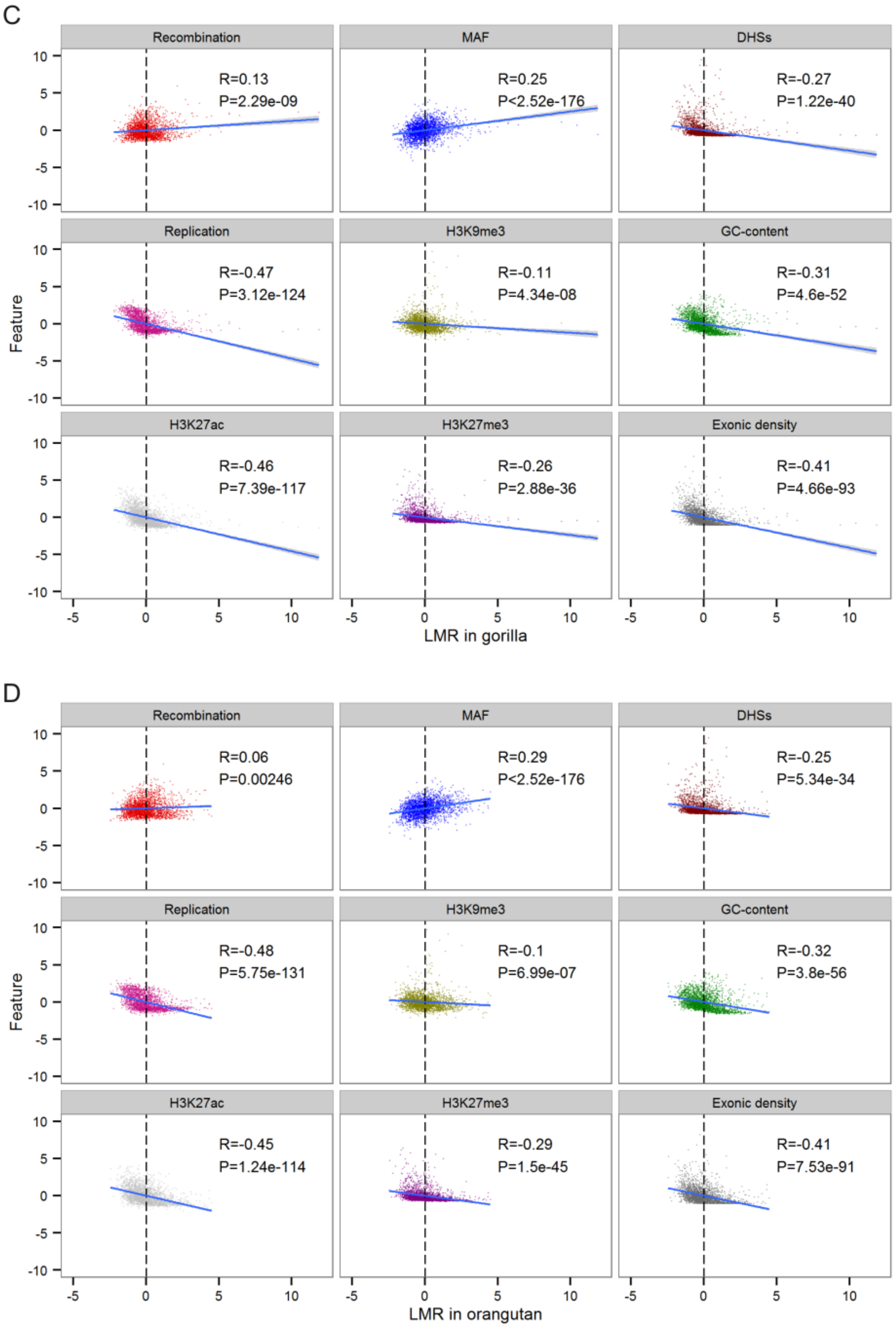


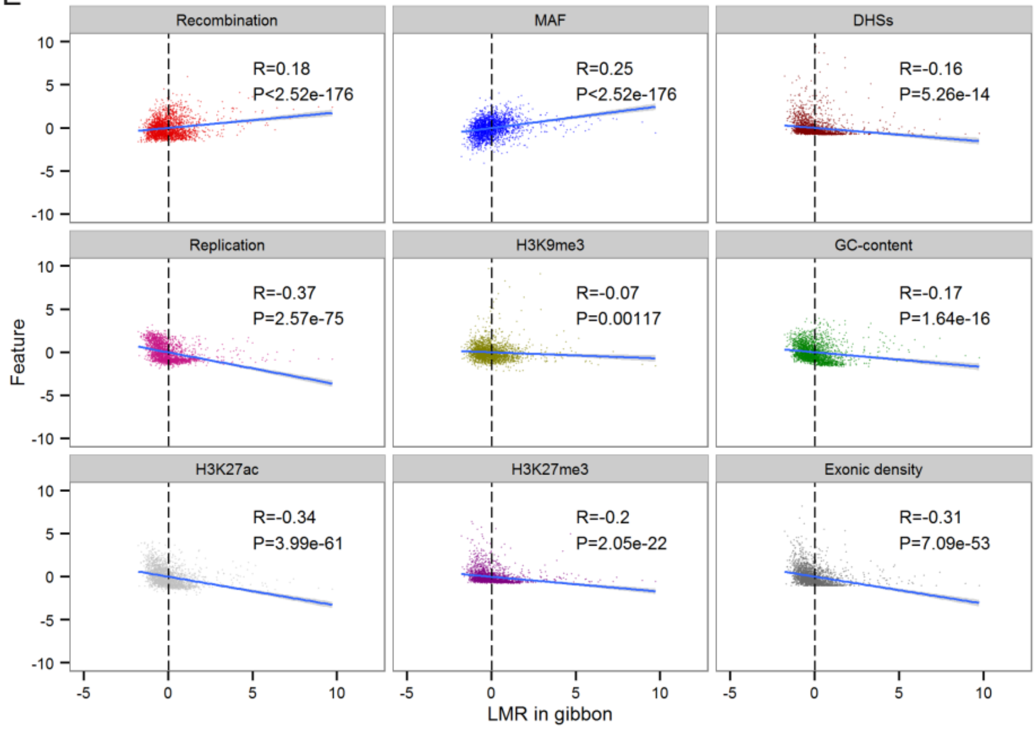
Figure S4. The distribution of the correlation coefficients between the phylogenetic distance from the human and the fraction of the variance in dLMR explained by the mLMR, obtained by bootstrapping.

(A-B) W↔S substitutions included; (C-D) W↔S substitutions excluded. (A and C) repeats excluded (same dataset as in the Main text); (B and D) repeats not excluded. Distributions of Spearman's correlation coefficients were obtained from 10,000 window-wise bootstrapping trials. For each panel, the P-value is calculated as the smaller of the left and right quantiles at *x*=0 with the Bonferroni correction applied.

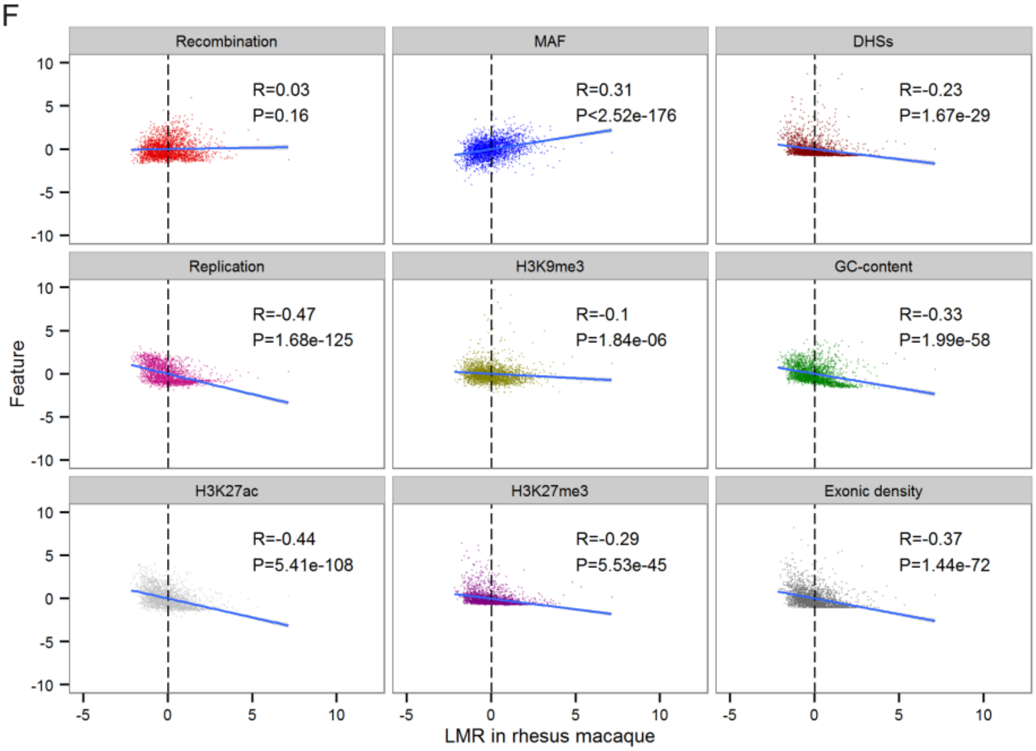

A
Recombination
R=0.21
P<2.52e-176
MAF
R=0.27
P<2.52e-176
DHSs
R=-0.27
P=8.6e-39
Replication
R=-0.43
P=3.13e-101
H3K9me3
R=-0.1
P=6.15e-07
GC-content
R=-0.26
P=6.3e-37
H3K27ac
R=-0.41
P=4.37e-91
H3K27me3
R=-0.25
P=2.98e-33
Exonic density
R=-0.39
P=2.53e-85
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in human
B
Recombination
R=0.21
P<2.52e-176
MAF
R=0.3
P<2.52e-176
DHSs
R=-0.23
P=5.61e-28
Replication
R=-0.44
P=4.19e-110
H3K9me3
R=-0.12
P=6.74e-09
GC-content
R=-0.25
P=1.65e-33
H3K27ac
R=-0.43
P=2.29e-104
H3K27me3
R=-0.25
P=1.55e-32
Exonic density
R=-0.39
P=3.31e-81
Feature
LMR in chimpanzee

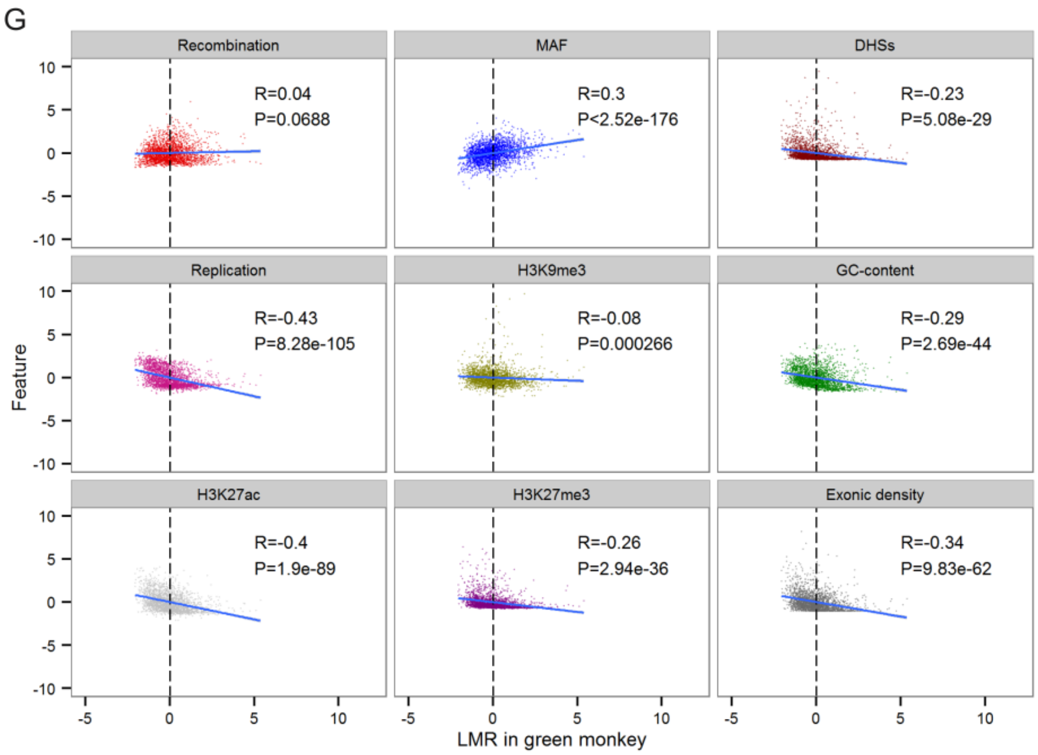

C
Recombination
R=0.13
P=2.29e-09
MAF
R=0.25
P<2.52e-176
DHSs
R=-0.27
P=1.22e-40
Replication
R=-0.47
P=3.12e-124
H3K9me3
R=-0.11
P=4.34e-08
GC-content
R=-0.31
P=4.6e-52
H3K27ac
R=-0.46
P=7.39e-117
H3K27me3
R=-0.26
P=2.88e-36
Exonic density
R=-0.41
P=4.66e-93
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in gorilla
D
Recombination
R=0.06
P=0.00246
MAF
R=0.29
P<2.52e-176
DHSs
R=-0.25
P=5.34e-34
Replication
R=-0.48
P=5.75e-131
H3K9me3
R=-0.1
P=6.99e-07
GC-content
R=-0.32
P=3.8e-56
H3K27ac
R=-0.45
P=1.24e-114
H3K27me3
R=-0.29
P=1.5e-45
Exonic density
R=-0.41
P=7.53e-91
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in orangutan

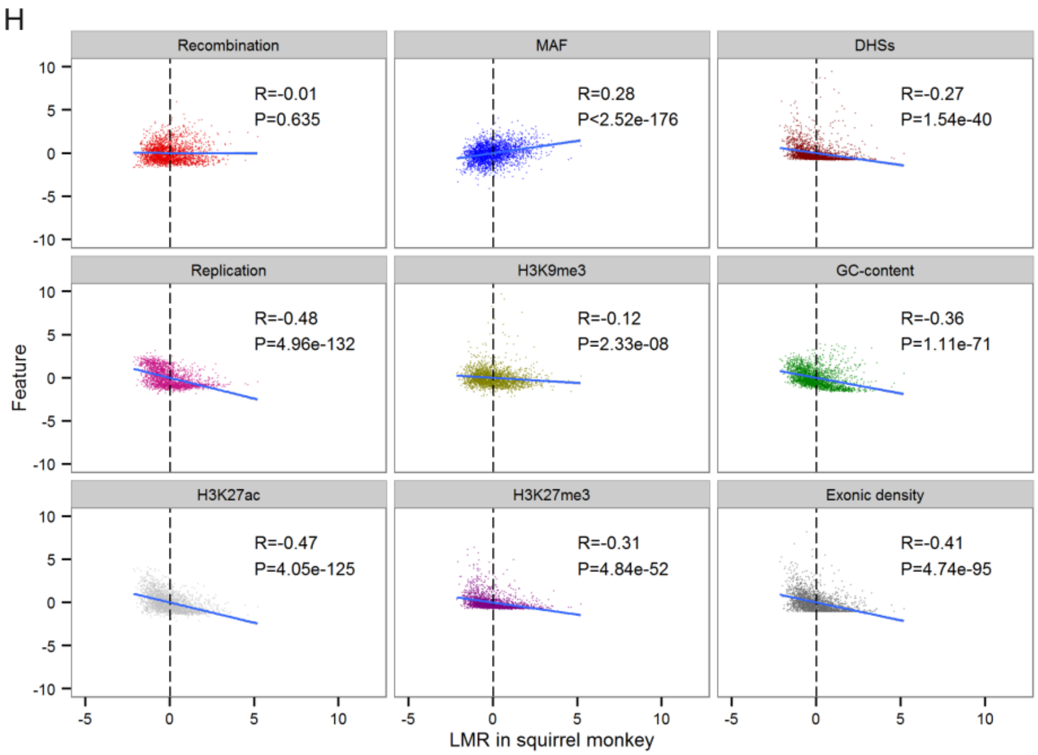
E


Recombination
R=0.18
P<2.52e-176
MAF
R=0.25
P<2.52e-176
DHSs
R=-0.16
P=5.26e-14
Replication
R=-0.37
P=2.57e-75
H3K9me3
R=-0.07
P=0.00117
GC-content
R=-0.17
P=1.64e-16
H3K27ac
R=-0.34
P=3.99e-61
H3K27me3
R=-0.2
P=2.05e-22
Exonic density
R=-0.31
P=7.09e-53
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in gibbon

F


Recombination
R=0.03
P=0.16
MAF
R=0.31
P<2.52e-176
DHSs
R=-0.23
P=1.67e-29
Replication
R=-0.47
P=1.68e-125
H3K9me3
R=-0.1
P=1.84e-06
GC-content
R=-0.33
P=1.99e-58
H3K27ac
R=-0.44
P=5.41e-108
H3K27me3
R=-0.29
P=5.53e-45
Exonic density
R=-0.37
P=1.44e-72
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in rhesus macaque


G
Recombination
R=0.04
P=0.0688
MAF
R=0.3
P<2.52e-176
DHSs
R=-0.23
P=5.08e-29
Replication
R=-0.43
P=8.28e-105
H3K9me3
R=-0.08
P=0.000266
GC-content
R=-0.29
P=2.69e-44
H3K27ac
R=-0.4
P=1.9e-89
H3K27me3
R=-0.26
P=2.94e-36
Exonic density
R=-0.34
P=9.83e-62
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in green monkey


H
Recombination
R=-0.01
P=0.635
MAF
R=0.28
P<2.52e-176
DHSs
R=-0.27
P=1.54e-40
Replication
R=-0.48
P=4.96e-132
H3K9me3
R=-0.12
P=2.33e-08
GC-content
R=-0.36
P=1.11e-71
H3K27ac
R=-0.47
P=4.05e-125
H3K27me3
R=-0.31
P=4.84e-52
Exonic density
R=-0.41
P=4.74e-95
Feature
10
5
0
-5
-10
-5
0
5
10
LMR in squirrel monkey

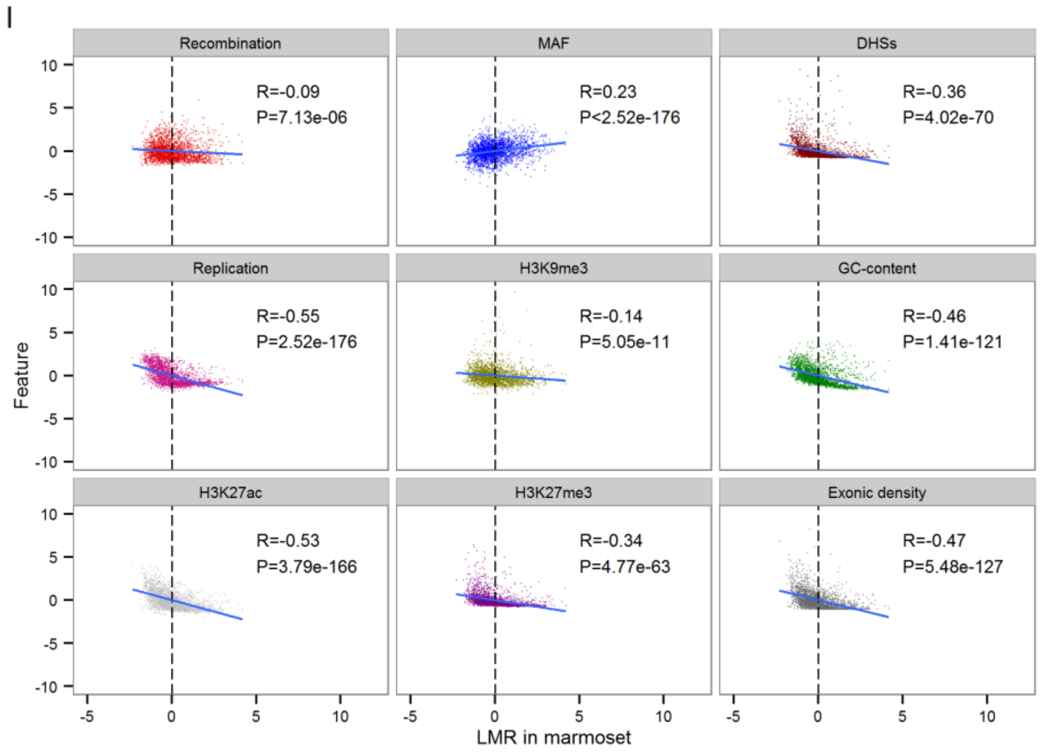


Figure S5. Scatterplots for raw correlations of LMRs in primate species and human genomic features.

(A) Human; (B) chimpanzee; (C) gorilla; (D) orangutan; (E) gibbon; (F) rhesus macaque; (G) green monkey; (H) squirrel monkey; (I) marmoset.

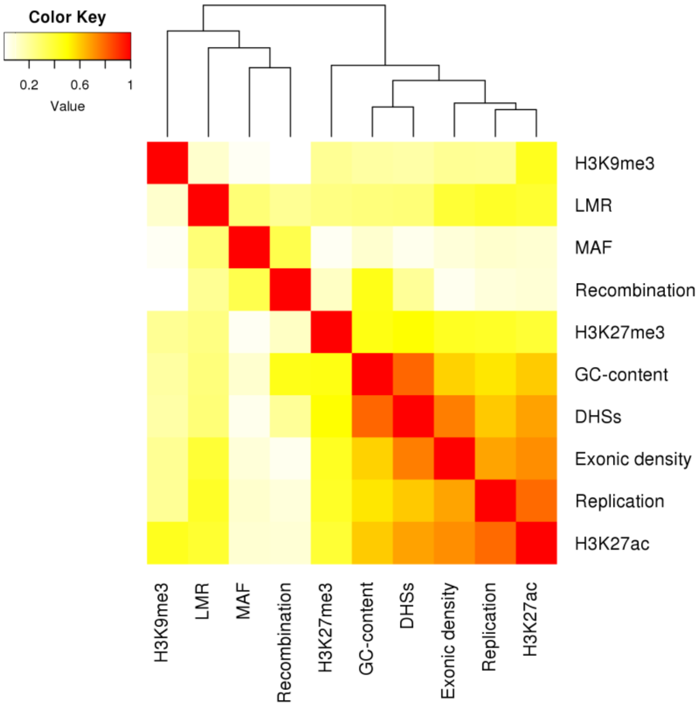
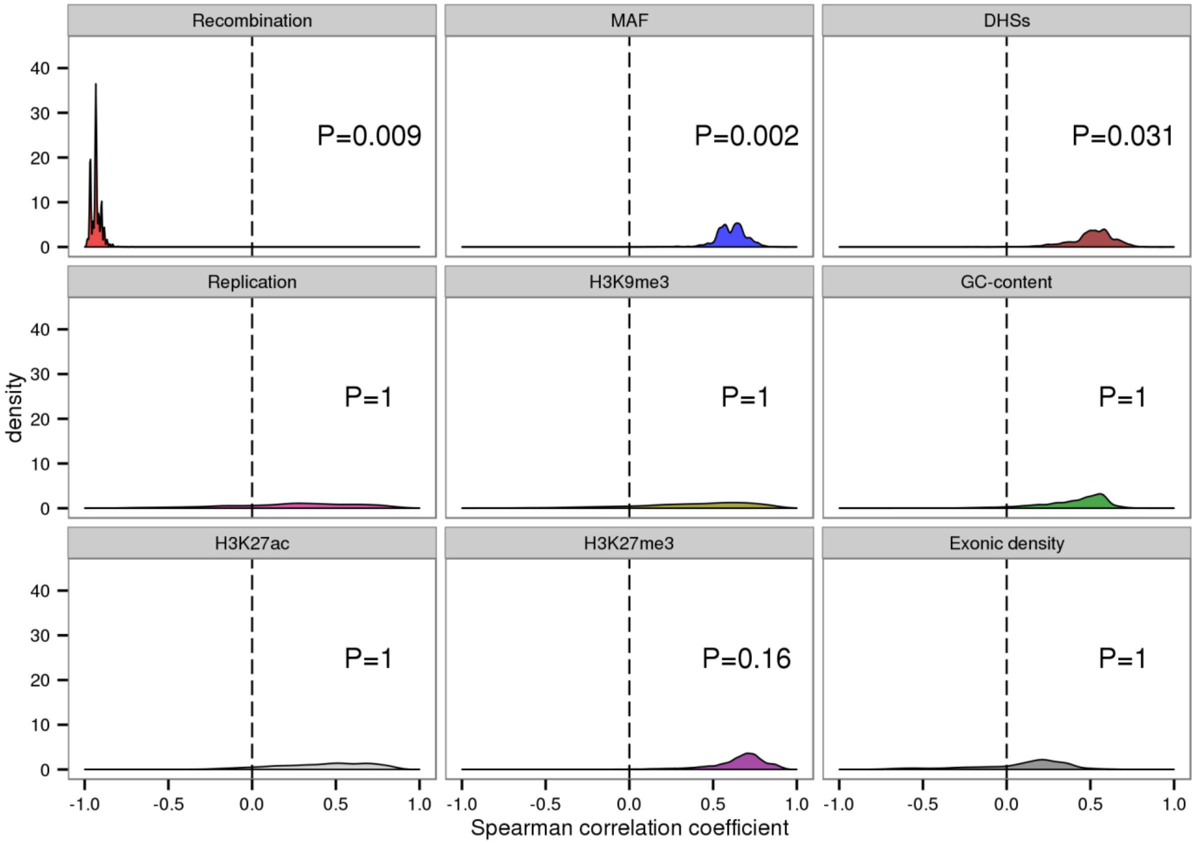


Figure S6. Heatmap for the absolute values of correlation coefficients between all possible pairs of human LMR and genomic features.

Warmer colors correspond to higher absolute values of the Pearson correlation coefficients.



Figuew S7. Distribution of the correlation coefficients between the phylogenetic distance from the human and the fraction of the variance in dLMR explained by each genomic feature (ANOVA Type III analysis).

Distributions of Spearman’s correlation coefficients were obtained from 10,000 window-wise bootstrapping trials. For each feature, the P-value reported is the smaller of the left and right quantiles at x=0 with the Bonferroni correction applied. Genomic features included in the model are the same as in Supplemental Figure S2.

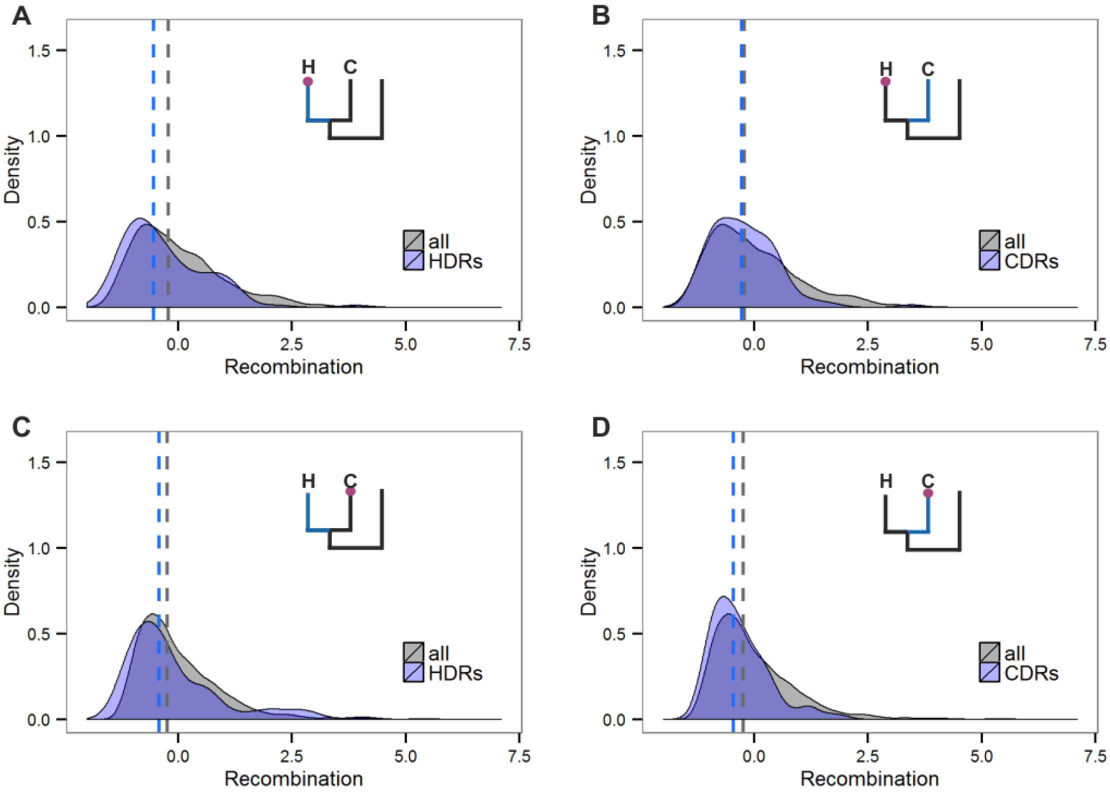
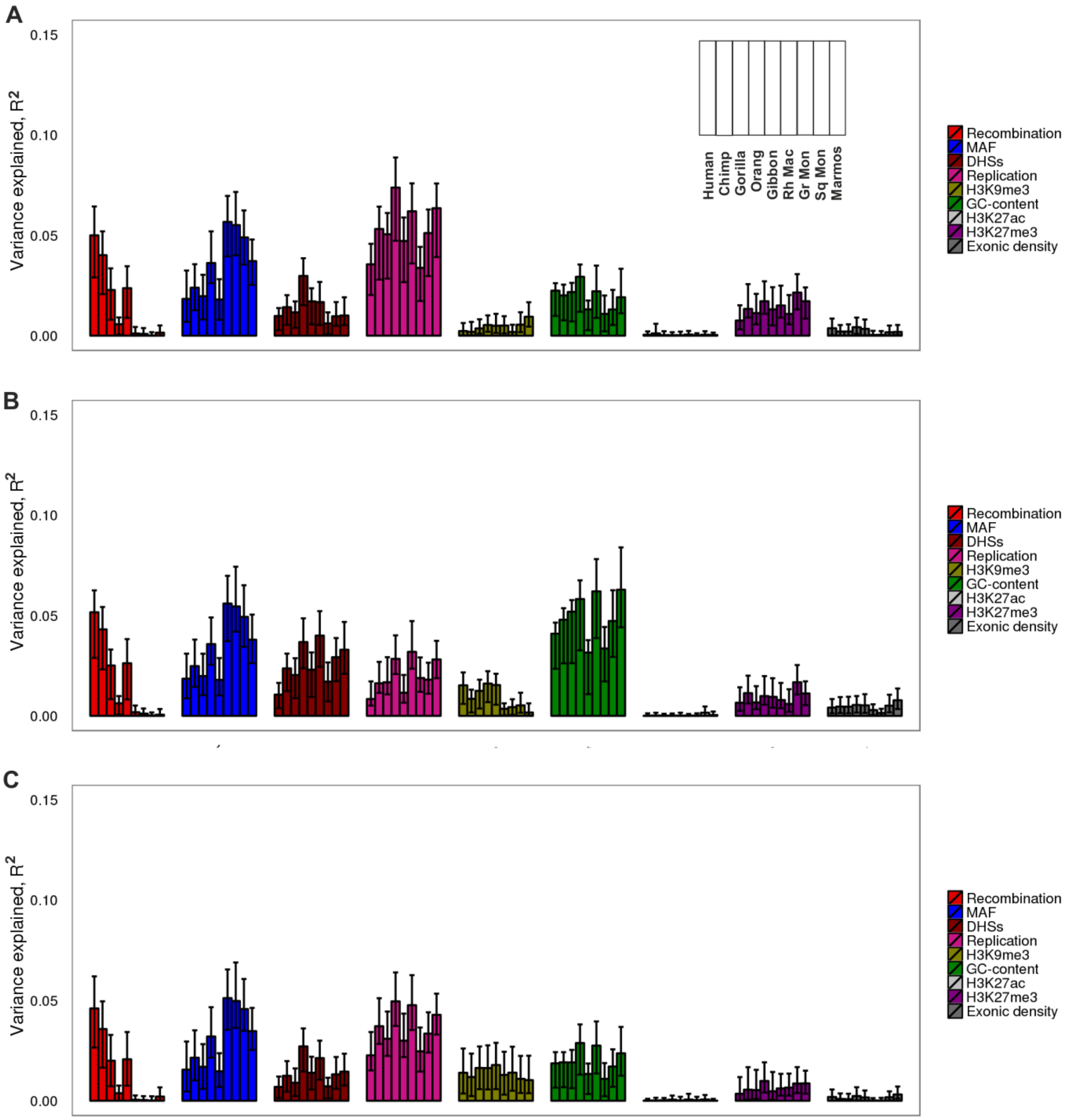


Figure S8. Densities of human recombination scores (A-B) in HDRs (A) and CDRs (B) and densities of chimpanzee recombination scores (C-D) in HDRs (C) and CDRs (D).

The schematic phylogenies show the lineage in which the LMR was decreased, in blue; and the species in which recombination was measured, as a circle. The dashed line corresponds to the median recombination rate in the DRs and all genomic regions.


A
B
C
0.15
0.10
0.05
0.00
Variance explained, R^2
Human
Chimp
Gorilla
Orang
Gibbon
Rh Mac
Gr Mon
Sq Mon
Marmos
Recombination
MAF
DHSs
Replication
H3K9me3
GC-content
H3K27ac
H3K27me3
Exonic density

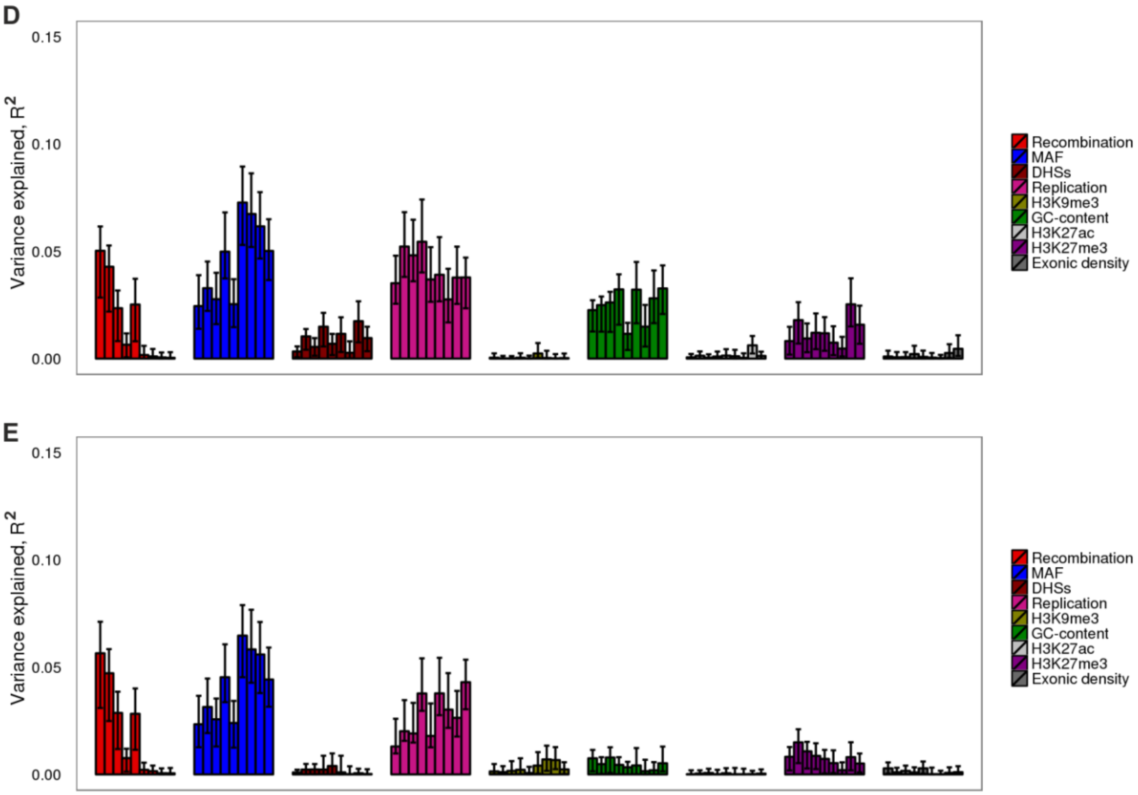


Figure S9. Variance of the LMR explained by the genomic features obtained from different tissues.

(A) GM12878; (B) HUVEC; (C) NHEK; (D) Hela-S3; (E) K562. Notations are as in Fig. 2A.). Genomic features included in the model are the same as in Supplemental Figure S2.

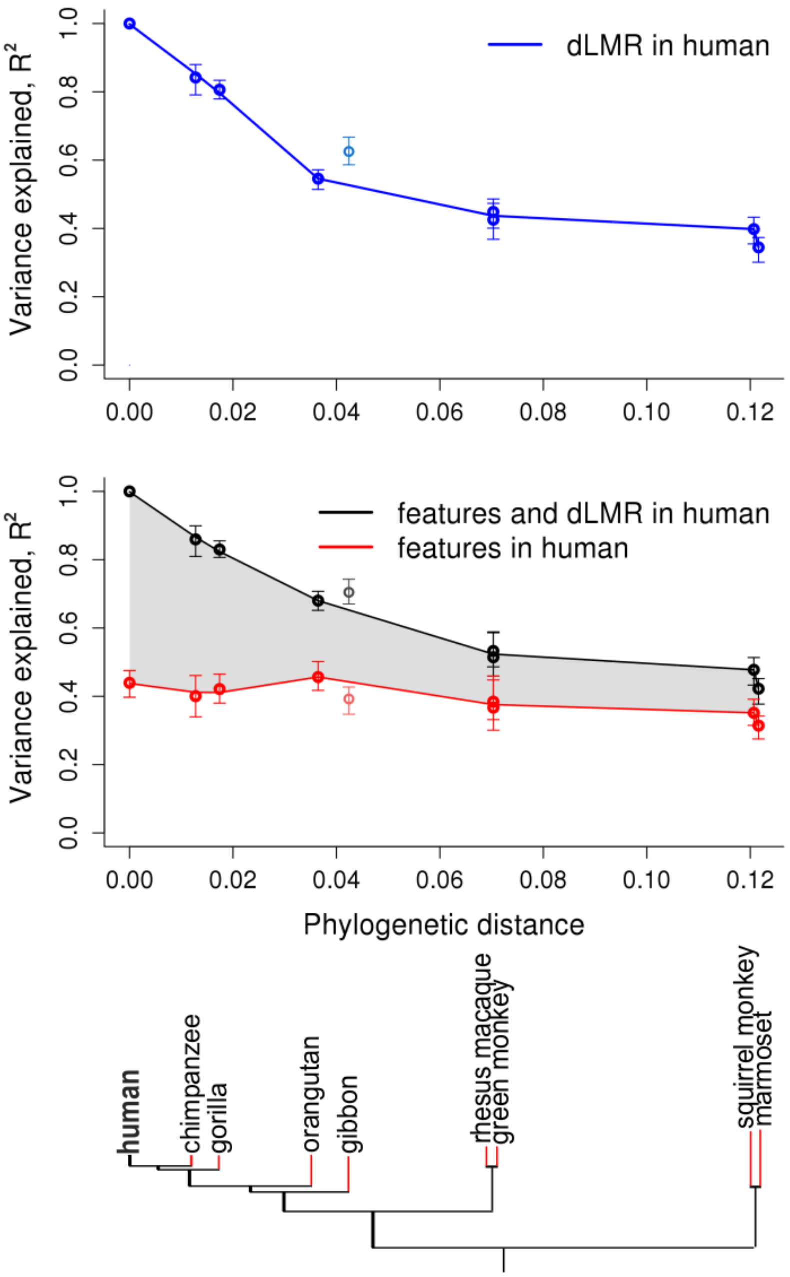


Figure S10. LMR variation in primate species explained by genomic features and dLMR in the human lineage.

dLMRs were inferred using maximum likelihood approach without excluding W↔S substitutions. Notations as in the Figure 1. Genomic features included in the model are the same as in Supplemental Figure S2.

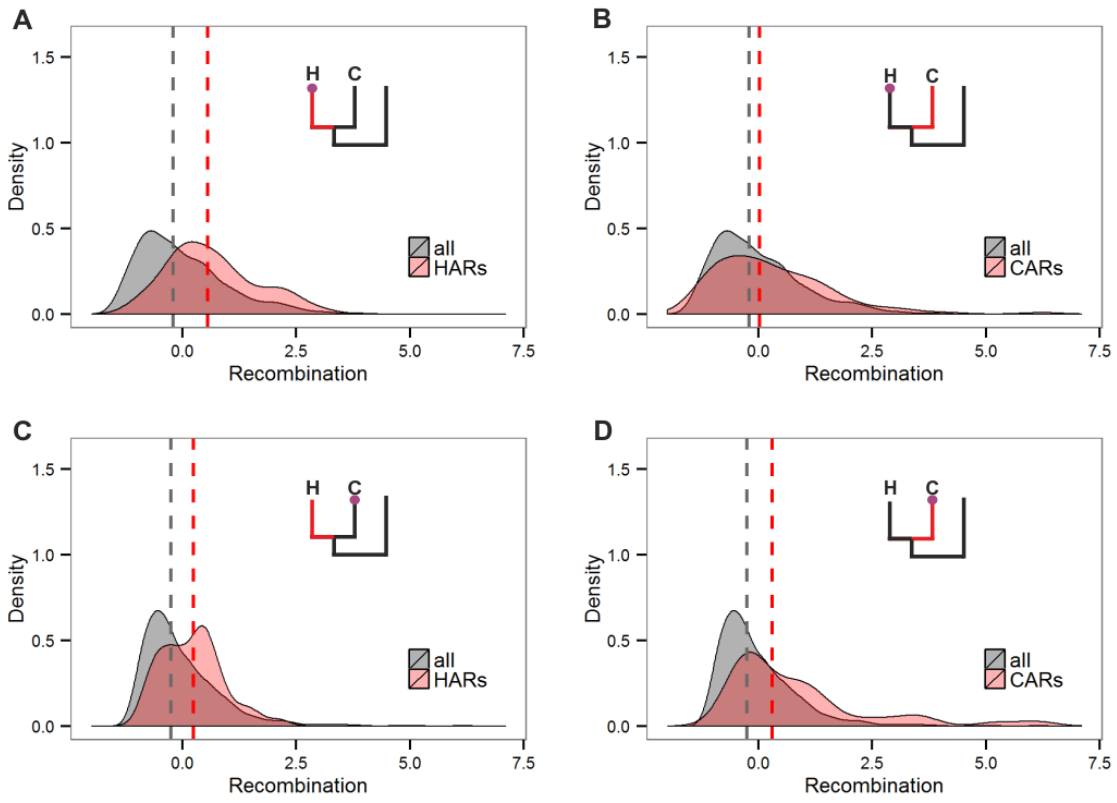


Figure S11. Densities of human recombination scores (A-B) in HARs (A) and CARs (B) and densities of chimpanzee recombination scores (C-D) in HARs (C) and CARs (D) on the data without excluding W↔S substitutions.

Notations are the same as in the Figure 3.

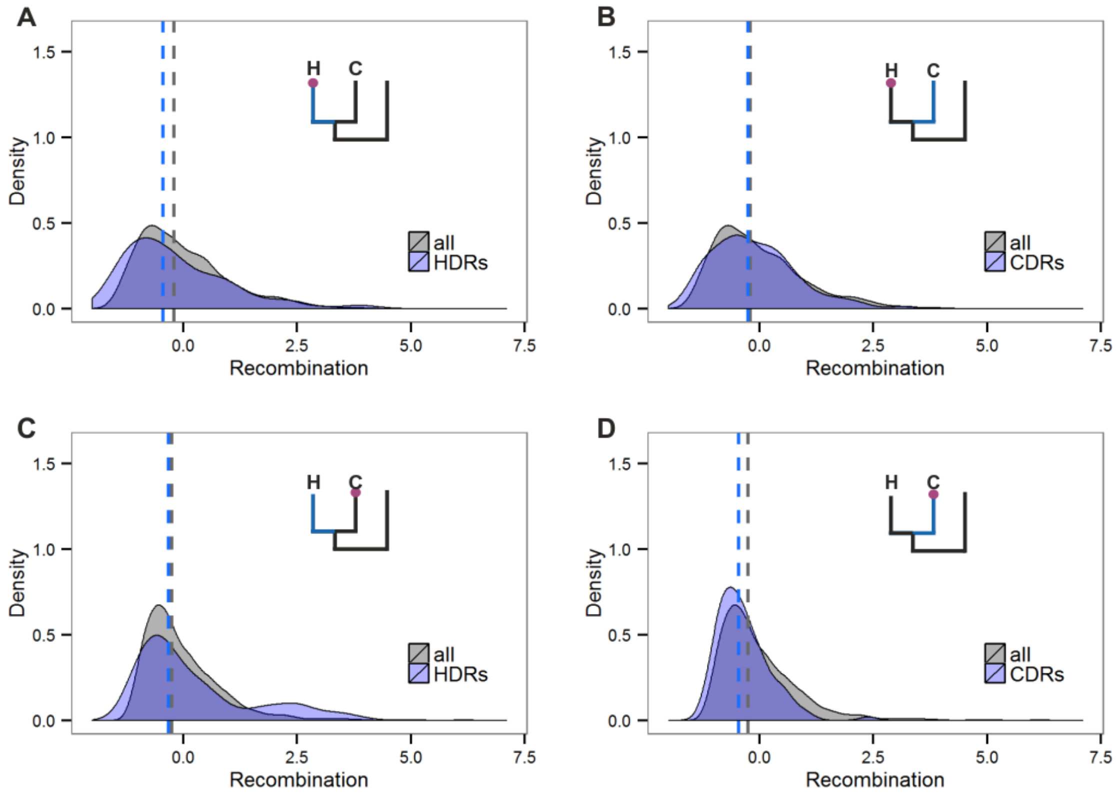


Figure S12. Densities of human recombination scores (A-B) in HDRs (A) and CDRs (B) and densities of chimpanzee recombination scores (C-D) in HDRs (C) and CDRs on the data without excluding W↔S substitutions.

Notations are the same as in the Supplemental Figure S8.

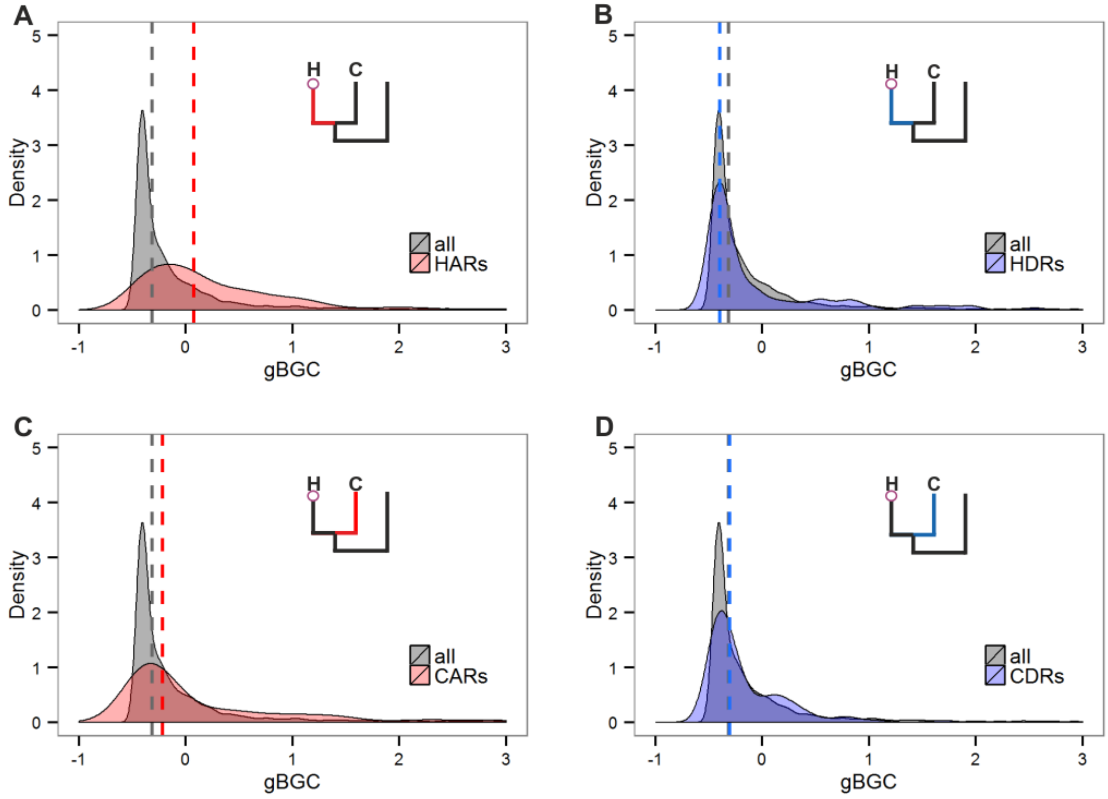


Figure S13. Densities of human gBGC scores in HARs (A), HDRs (B), CARs (C), and CDRs (D) on the data without excluding W↔S substitutions.

The schematic phylogenies show the lineage in which the LMR was increased, in red or decreased, in blue; and the species in which gBGC was measured (human), as an open circle. The dashed line corresponds to the median of the gBGC rates in ARs, DRs and all genomic regions.

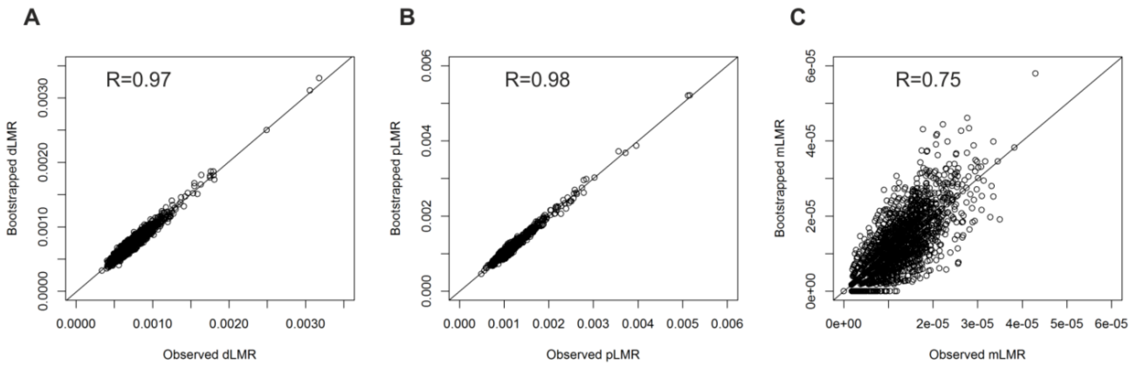


Figure S14. Correlation of the observed and bootstrapped LMRs in 1Mb genomic windows.

(A) dLMR; (B) pLMR; (C) mLMR.

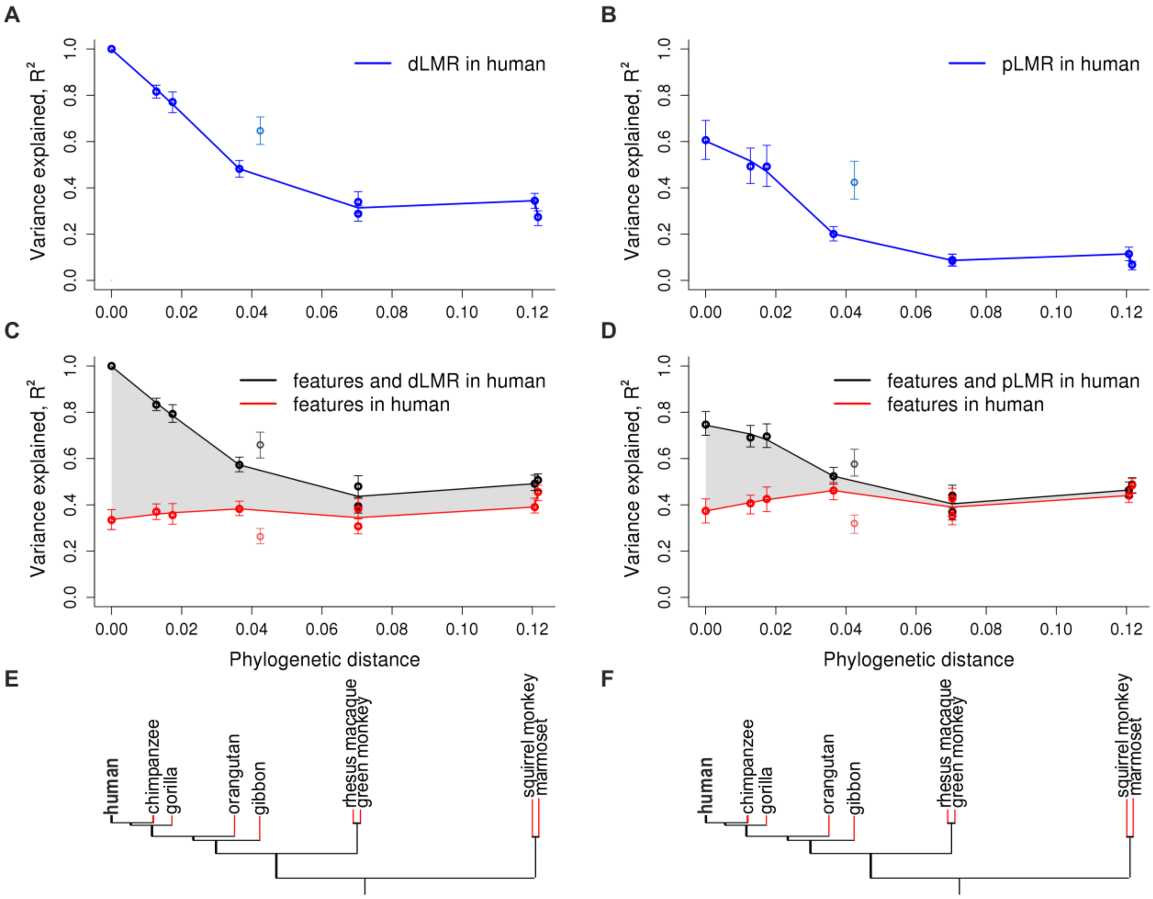


Figure S15. Local mutation rate variation in primate species explained by genomic features and LMR in the human lineage.

pLMRs and dLMRs are calculated accounting for the changes in the individual substitution rates between species. Notations are the same as in Figure 1. Genomic features included in the model are the same as in Supplemental Figure S2.

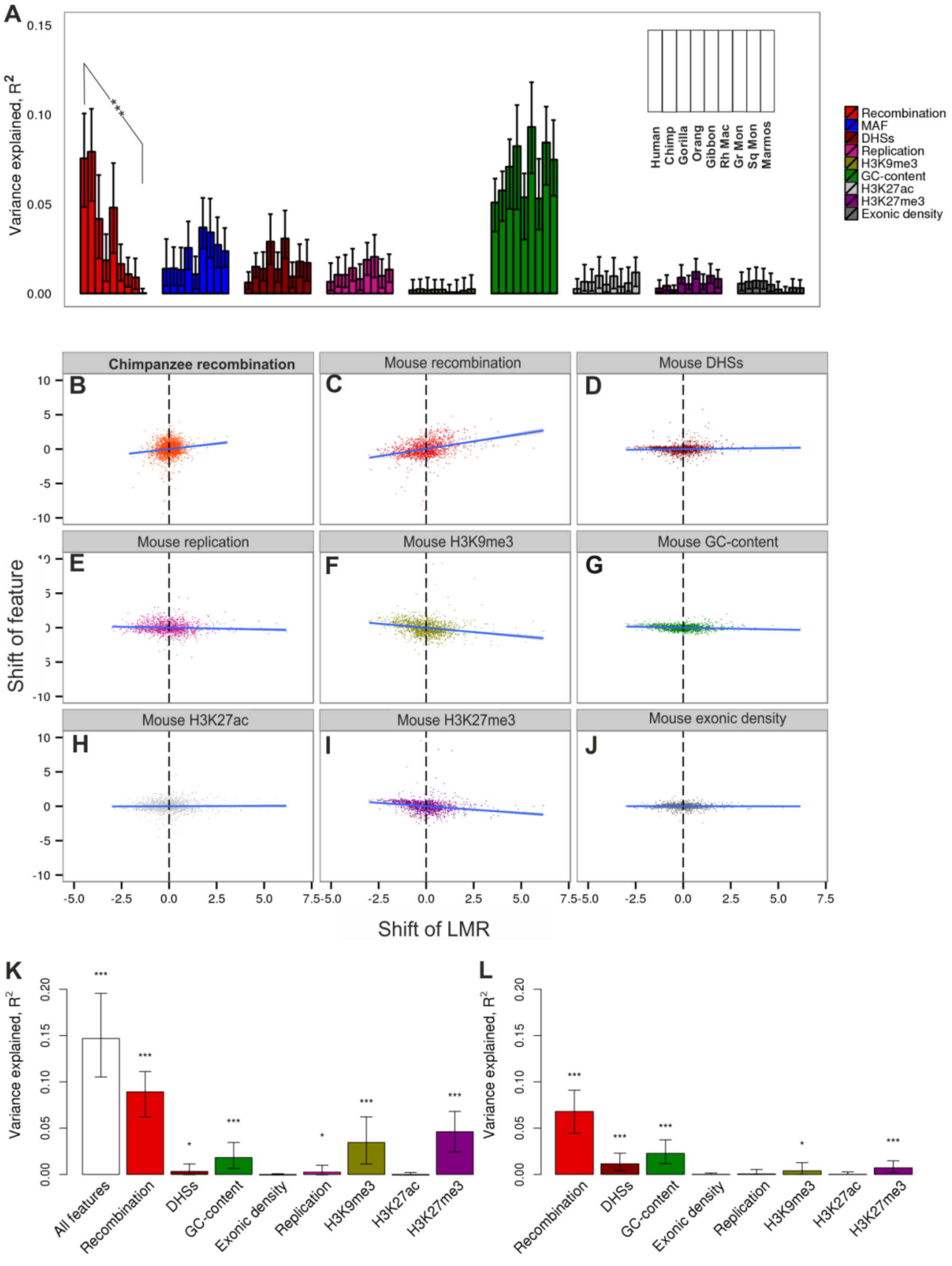


Figure S16. LMR explained by individual genomic features.

dLMRs are calculated accounting for the changes in the individual substitution rates between species. Notations are the same as in Figure 2.

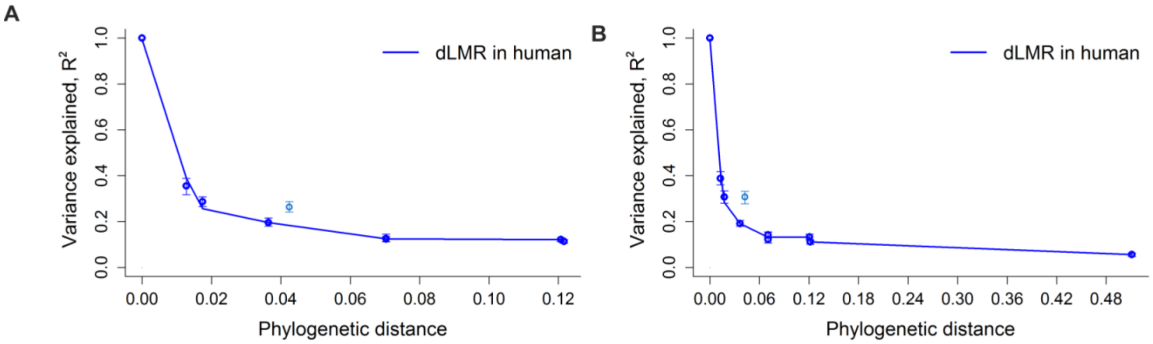


Figure S17. Local mutation rate (LMR) variation in primate species explained by genomic features and LMR in the human lineage in 100Kb windows (A), and with the addition of mouse (B).

Notations are as in Figure 1A.

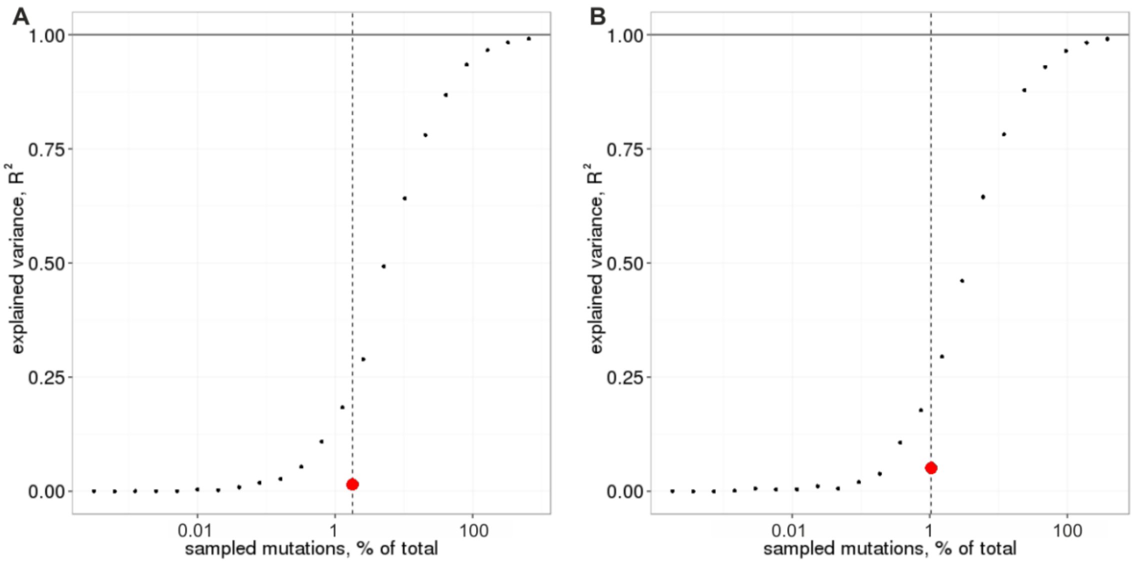


Figure S18. Variation in human LMR explained by the estimate of the same variable based on a subsample of mutations.

(A) dLMR; (B) pLMR. Horizontal axis, fraction of sampled mutations; vertical axis, fraction of the explained variance (R^2). The dashed line corresponds to 11,429 mutations. The circles show the correlations between the mLMR and dLMR (in A) or pLMR (in B).

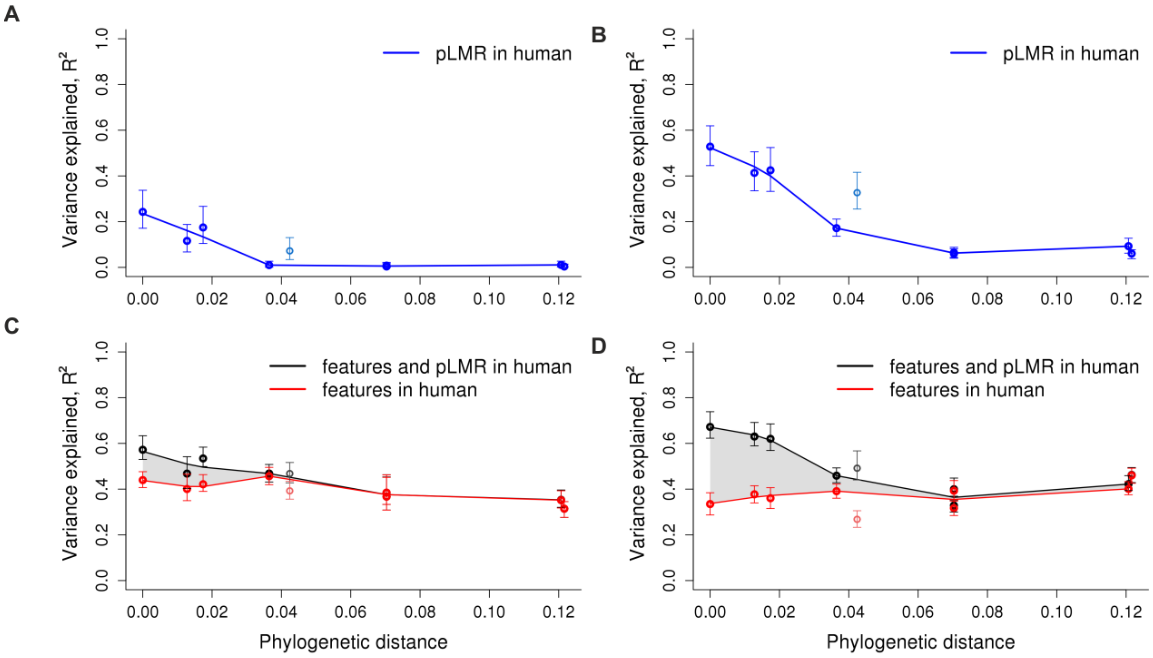


Figure S19. dLMR variation in primate species explained by genomic features and pLMR in the human lineage.

Notations are as in Fig. 1. (A and C) W↔S substitutions included in dLMR and pLMR estimates; (B and D) W↔S substitutions excluded in dLMR estimates, and included in pLMR estimates. Genomic features included in the model are the same as in Supplemental Figure S2.

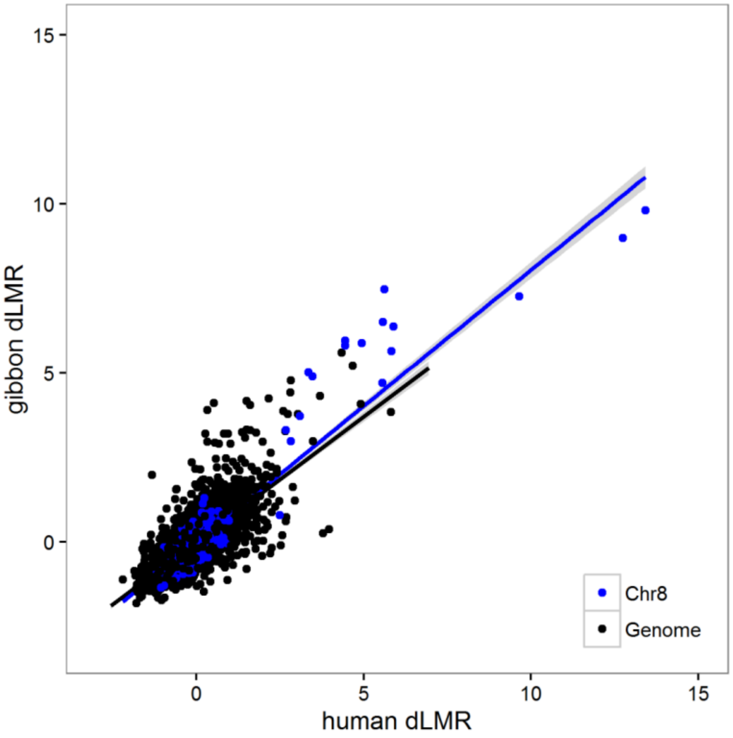


Figure S20. Correlation of the human and gibbon LMRs with (black line) and without (blue line) exclusion of the 8 chromosome 1 Mb genomic windows.

Blue dots correspond to the windows from the 8th chromosome; black dots correspond to the all other genomic windows.

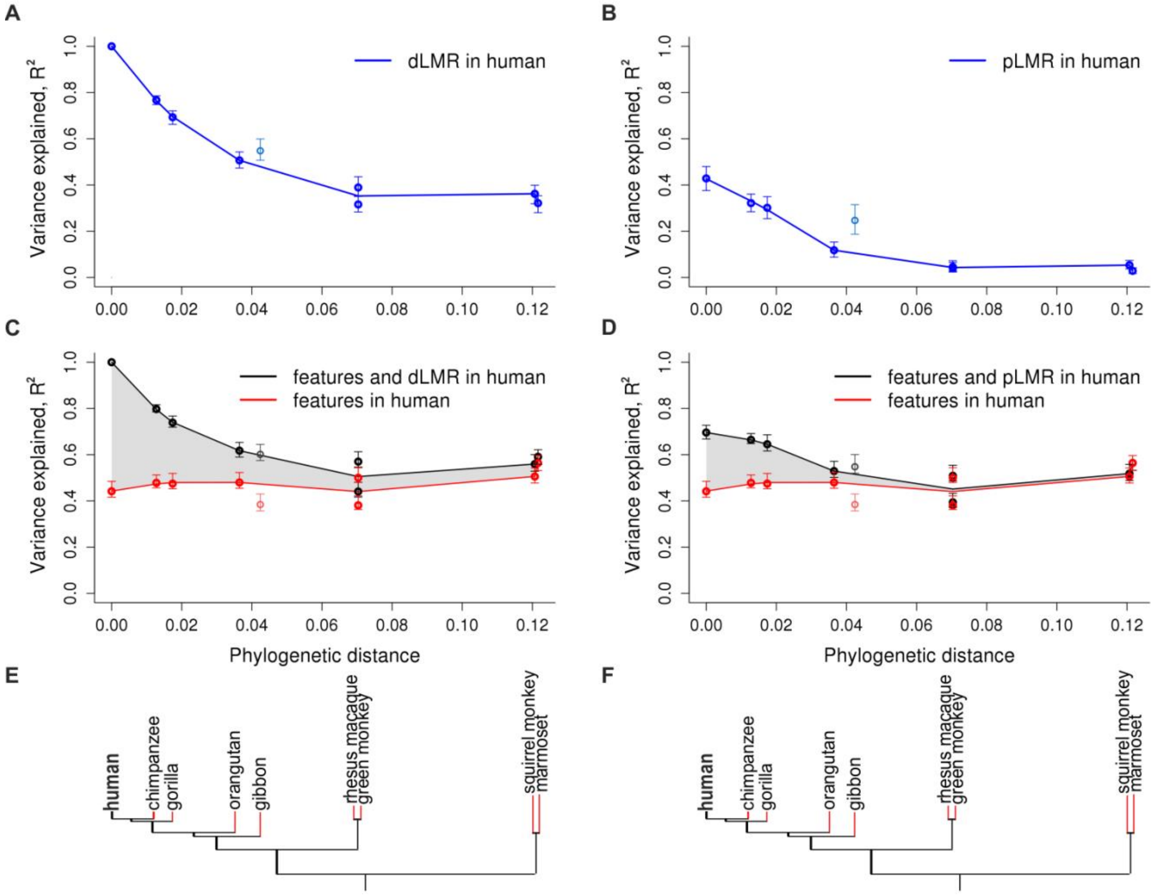


Figure S21. Local mutation rate variation in primate species explained by genomic features and LMR in the human lineage.

pLMRs and dLMRs are calculated on 1 Mb genomic windows excluding those from the 8 chromosome. Notations are the same as in Figure 1.

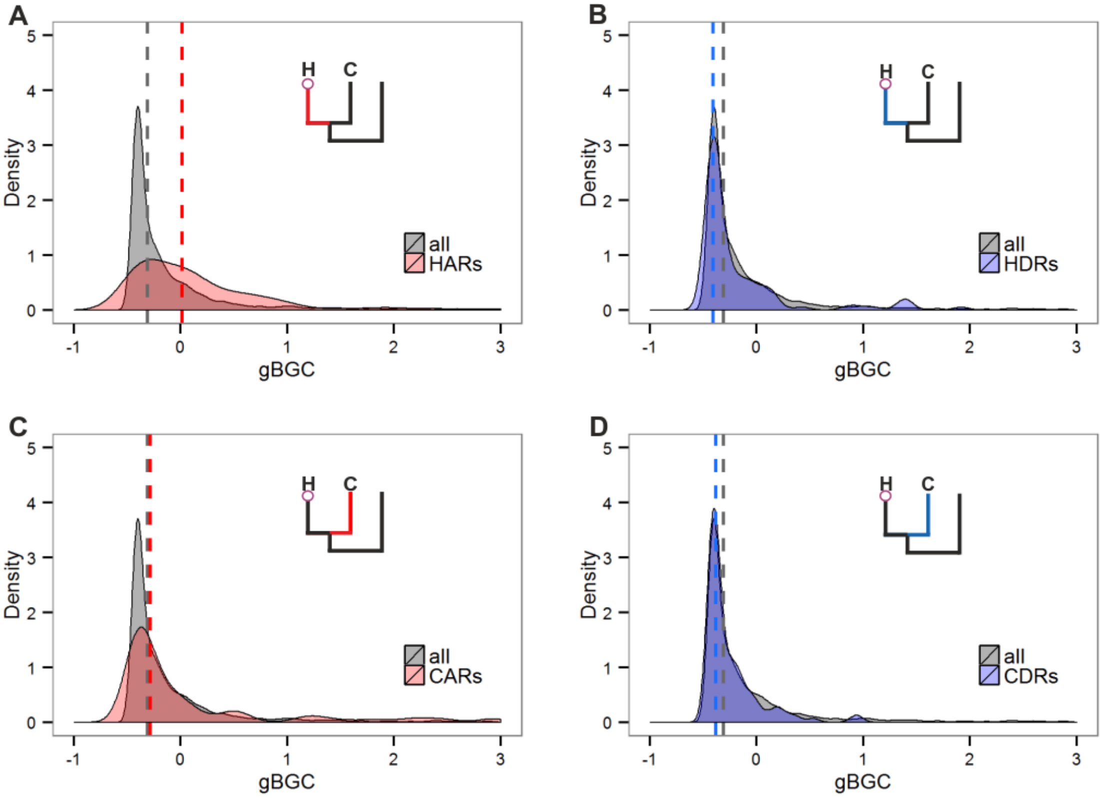


Figure S22. Densities of human BGC scores in HARs (A), HDRs (B), CARs (C), and CDRs (D).

The schematic phylogenies show the lineage in which the LMR was increased, in red or decreased, in blue; and the species in which gBGC was measured (human), as an open circle. The dashed line corresponds to the median of the gBGC rates in ARs, DRs and all genomic regions.

Supplemental Tables

Table S1. The fraction of the variance in human and non-human dLMR explained by the features obtained from different tissue types.

The set of 9 genomic features is the same as in Fig. 1. Five of these features varied between tissues: replication timing, number of DHSs, and densities of the three histone marks.

						Rhesus	Green	Squirrel	
	Human	Chimpanzee	Gorilla	Orangutan	Gibbon	Macaque	Monkey	Monkey	Marmoset
H1Esc	0.33	0.37	0.36	0.38	0.26	0.38	0.31	0.39	0.46
K562	0.32	0.35	0.33	0.35	0.23	0.34	0.28	0.35	0.42
NHEK	0.37	0.40	0.39	0.43	0.30	0.41	0.32	0.40	0.47
GM12878	0.36	0.39	0.38	0.42	0.28	0.39	0.30	0.40	0.47
HelaS3	0.33	0.35	0.34	0.34	0.24	0.31	0.26	0.35	0.40
Huvec	0.37	0.39	0.38	0.43	0.29	0.39	0.31	0.41	0.47

Table S2. Average human and chimpanzee dLMRs in the HDRs, CDRs, HARs and CARs.

The values in brackets are the percentage relative to the genome average.

	HDRs	CDRs	Genome Average	HARs	CARs
Human	0.00058 (79%)	0.00070 (95%)	0.00074	0.00104 (141%)	0.00076 (103%)
Chimpanzee	0.00080 (97%)	0.00066 (79%)	0.00083	0.00098 (118%)	0.0010 (122%)

References

Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A 2013. A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. PLoS Genet 9: e1003684.

Duret L, Arndt PF 2008. The Impact of Recombination on Nucleotide Substitutions in the Human Genome. PLoS Genet 4: e1000071.

Duret L, Galtier N 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. Annual Review of Genomics and Human Genetics 10: 285-311.

Glemin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L 2015. Quantification of GC-biased gene conversion in the human genome. Genome Res 25: 1215-1228.

Nusbaum C, Mikkelsen TS, Zody MC, Asakawa S, Taudien S, Garber M, Kodira CD, Schueler MG, Shimizu A, Whittaker CA, Chang JL, Cuomo CA, Dewar K, FitzGerald MG, Yang X, Allen NR, Anderson S, Asakawa T, Blechschmidt K, Bloom T, Borowsky ML, Butler J, Cook A, Corum B, DeArellano K, DeCaprio D, Dooley KT, Dorris L, Engels R, Glöckner G, Hafez N, Hagopian DS, Hall JL, Ishikawa SK, Jaffe DB, Kamat A, Kudoh J, Lehmann R, Lokitsang T, Macdonald P, Major JE, Matthews CD, Mauceli E, Menzel U, Mihalev AH, Minoshima S, Murayama Y, Naylor JW, Nicol R, Nguyen C, O'Leary SB, O'Neill K, Parker SCJ, Polley A, Raymond CK, Reichwald K, Rodriguez J, Sasaki T, Schilhabel M, Siddiqui R, Smith CL, Sneddon TP, Talamas JA, Tenzin P, Topham K, Venkataraman V, Wen G, Yamazaki S, Young SK, Zeng Q, Zimmer AR, Rosenthal A, Birren BW, Platzer M, Shimizu N, Lander ES 2006. DNA sequence and analysis of human chromosome 8. Nature 439: 331-335.

Schuster-Bockler B, Lehner B 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature 488: 504-507.