Supporting Online Material

Supplemental Figures S1 to S14

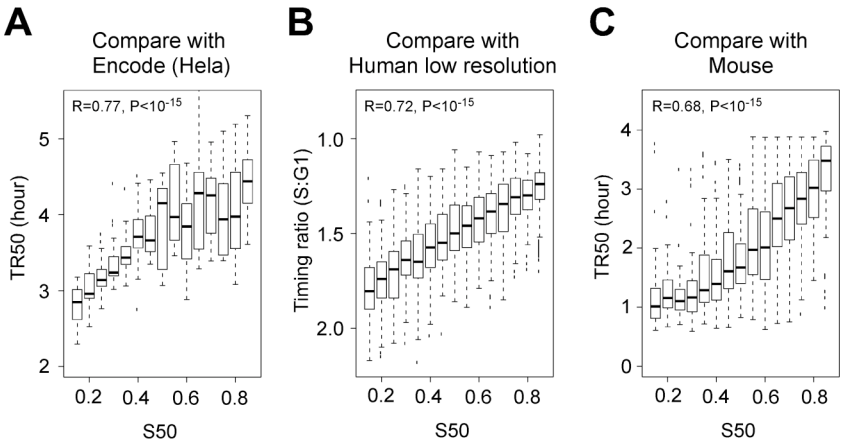

Figure S1. Correlations between replication timing data from this study (HeLa cells) (box plot representation) and (A) mean timing values computed in 100-kbp windows within ENCODE regions (HeLa cells, 1% of the human genome) (Karnani et al. 2007), (B) timing values from the whole human genome low resolution data from lymphocytes (computed in spotted DNA fragments) (Woodfine et al. 2004) and (C) timing values from mouse lymphocytes (Farkash-Amar et al. 2008) computed in human/mouse syntenic blocks retrieved from (Ma et al. 2006).

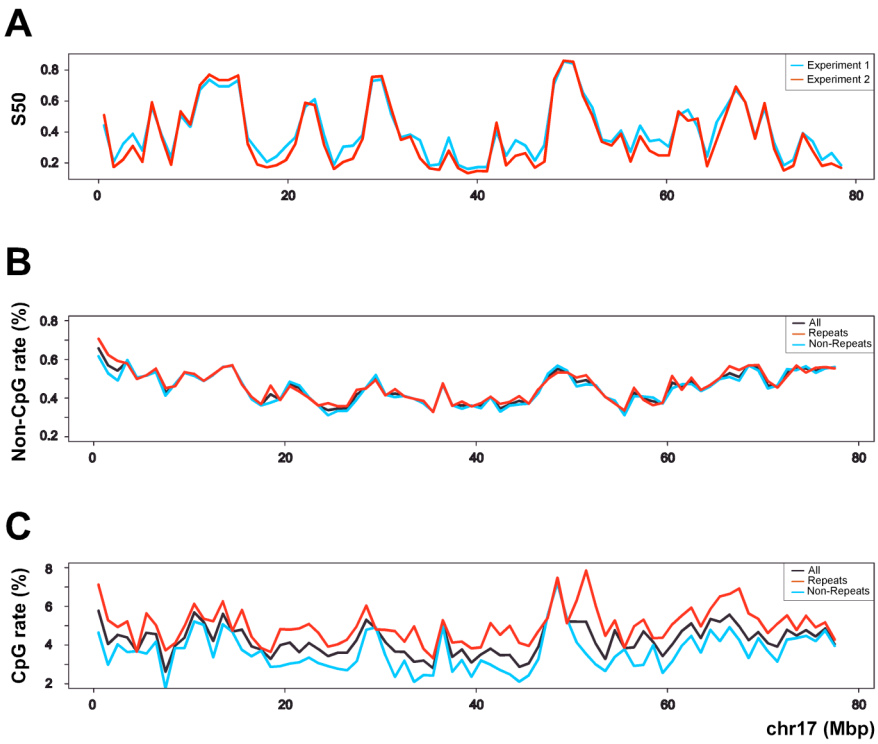

Figure S2. (A) Replication timing profiles (S50) along the human chromosome 17 obtained in experiments 1 and 2 (see main text, Methods). (B) Profiles of the non-CpG global substitution rate in all non-coding regions of chromosome 17 (black), in repeated (red) and non-repeated (blue) sequences. (C) Profiles of CpG transition rates in all regions of chromosome 17 (black), in repeated (red) and non-repeated sequences (blue). Rates and replication timing were computed in 1 Mbp windows.

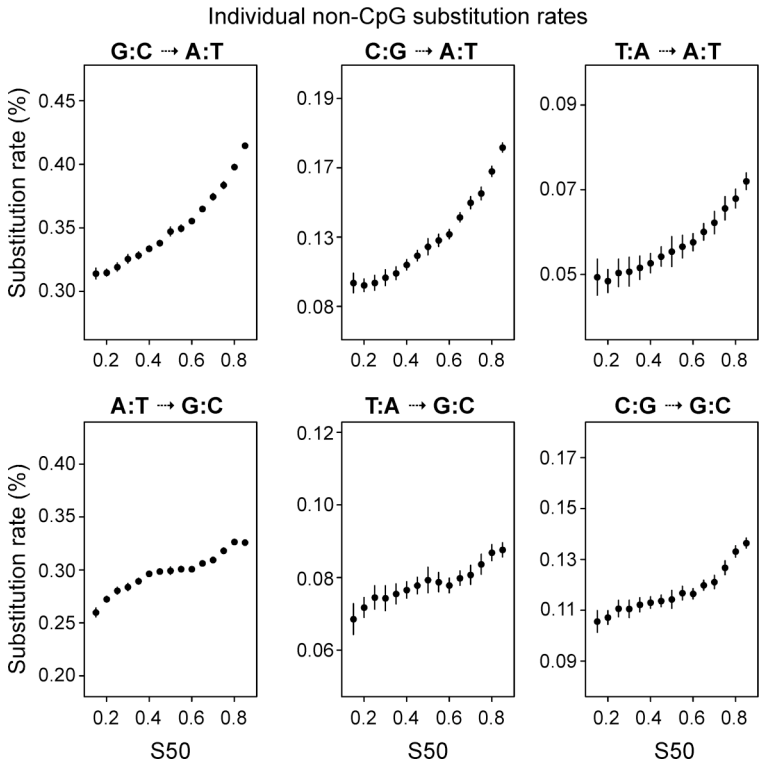

Figure S3. Variation of human individual substitution rates with replication timing. Non-CpG substitutions are computed in 100-kbp windows and timing values are as in main text Fig. 2. The relative increase of substitution rate during the S phase is G:C→A:T (35%), C:G→A:T (83%), T:A→A:T (55%), A:T→G:C (20%), T:A→G:C (23%), C:G→G:C (30%).

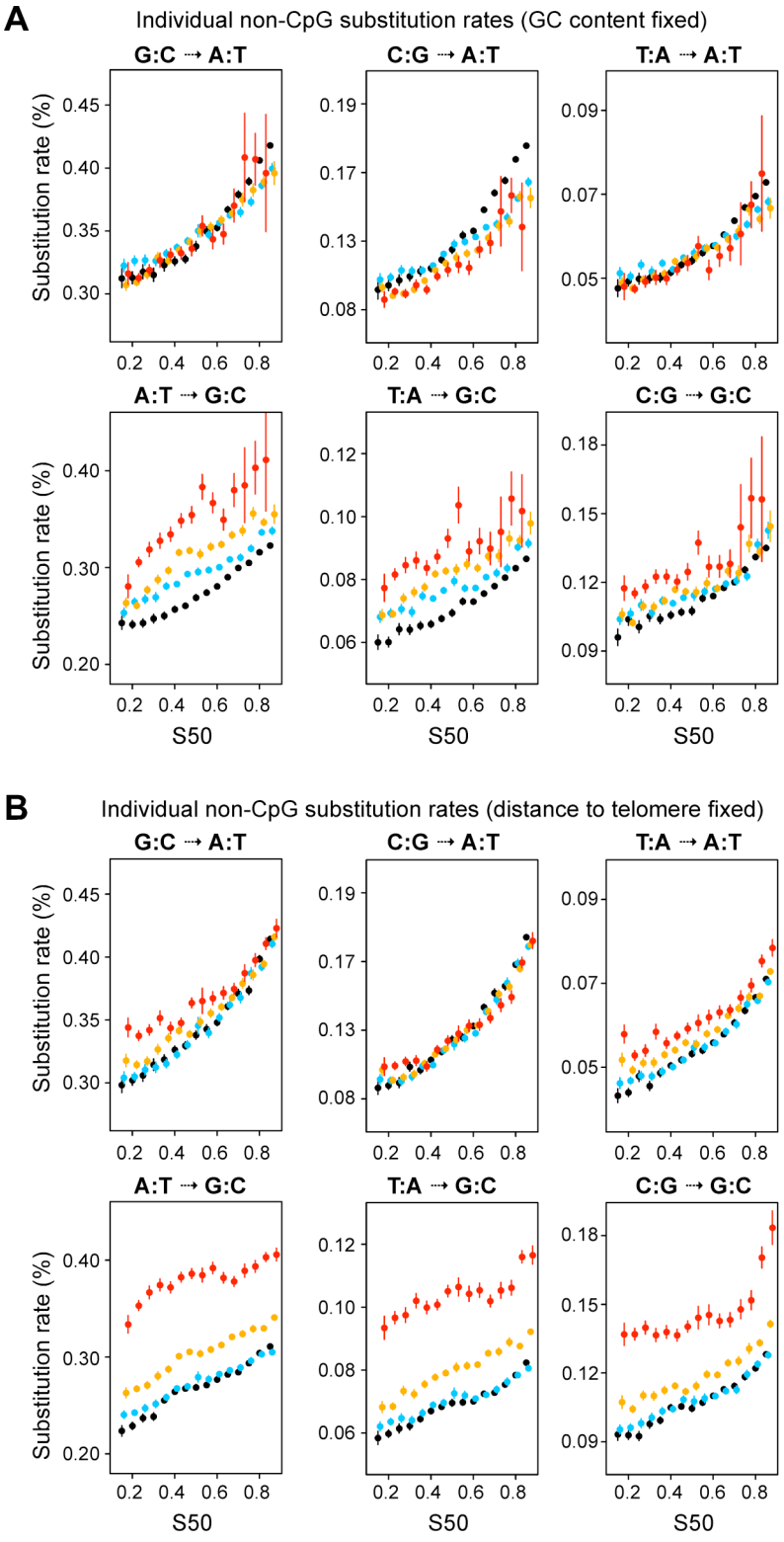
A
Individual non-CpG substitution rates (GC content fixed)
G:C → A:T
C:G → A:T
T:A → A:T
A:T → G:C
T:A → G:C
C:G → G:C
Substitution rate (%)
S50
B
Individual non-CpG substitution rates (distance to telomere fixed)
G:C → A:T
C:G → A:T
T:A → A:T
A:T → G:C
T:A → G:C
C:G → G:C
Substitution rate (%)
S50

Figure S4. Variation of individual human non-CpG substitution rates with replication timing when (A) controlling for GC content: black, GC≤37%; blue, 37<GC≤42%; orange, 42<GC≤52%; red, GC>52%. (B) Control for distance to telomere (DT): black, DT>50 Mbp; blue, 30<DT≤50 Mbp; orange, 10<DT≤30 Mbp; red, DT≤10 Mbp. Replication timing values and non-CpG substitution rates are determined in 100 kbp windows. N→W display close parallel lines while N→S show widely-separated lines, as expected from the biased gene conversion mechanism (Berglund et al. 2009; Duret and Arndt 2008; Galtier et al. 2009).

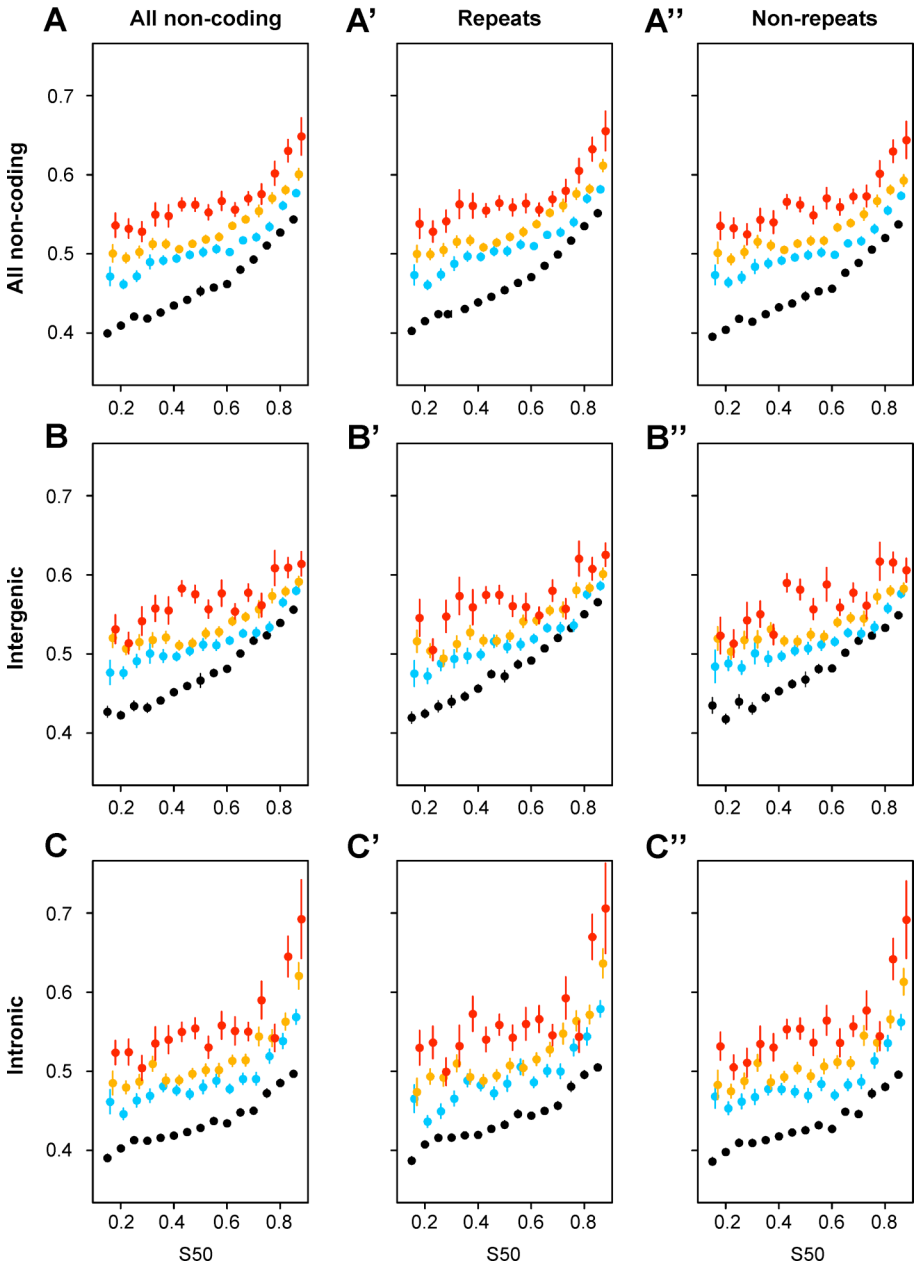

Figure S5. Variations of non-CpG global substitution rate with replication timing when controlling for crossover rate, in intergenic and intronic regions, in repeated elements and regions that do not contain repeated elements. Timing values and rates are determined in non-coding regions in 100 kbp windows. (A-C) All non-coding sequences; (A'-C') repeated elements; (A''-C'') non-repeated sequences. (A, A', A") all non-coding sequences; (B, B', B") intergenic sequences; (C, C', C") intronic sequences: black, CO≤1 cM/Mbp; blue, 1<CO≤2 cM/Mbp; orange, 2<CO≤4 cM/Mbp; red, CO>4 cM/Mbp. The similarity of the substitution rates variation was also observed when controlling for GC and DT (data not shown).

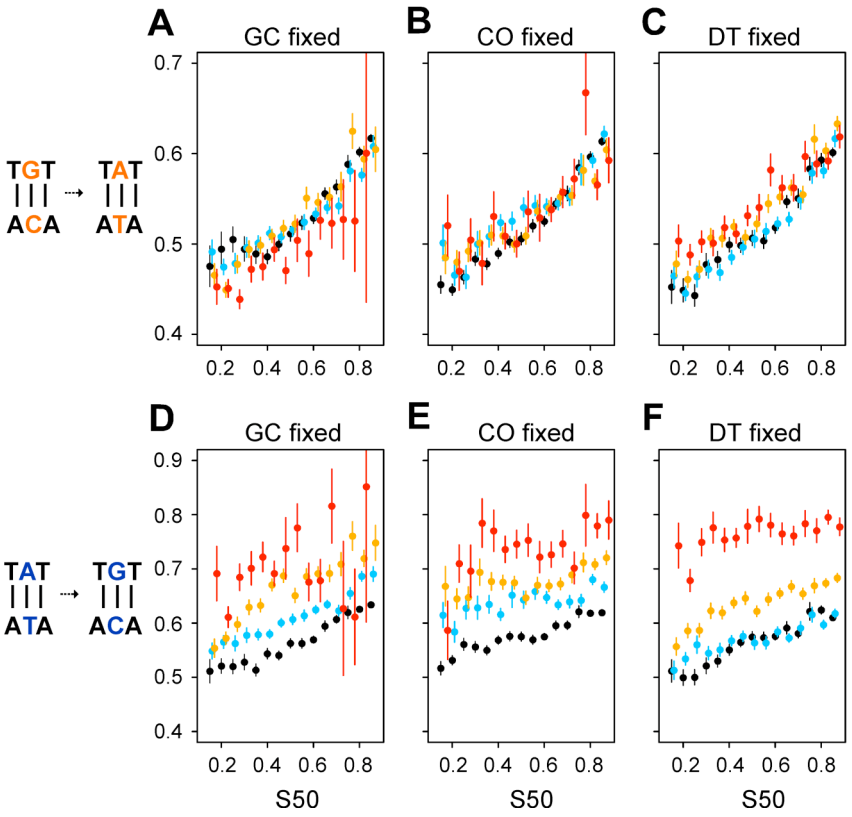

Figure S6. Variation of individual substitution rates with replication timing when controlling for the flanking nucleotide context. The TGT:ACA→TAT:ATA and TAT:ATA→TGT:ACA transition rates are examined after controlling for GC, CO and DT. Substitution rates can depend on the identity of flanking nucleotides (Hwang and Green 2004). Timing values and substitution rates are determined in 100 kbp windows. (A, D) Control of GC content: GC≤37%; blue, 37<GC≤42%; orange, 42<GC≤52%; red, GC>52%; (B, E) control of CO: black, CO≤1 cM/Mbp; blue, 1<CO≤2 cM/Mbp; orange, 2<CO≤4 cM/Mbp; red, CO>4 cM/Mbp; (C, F) control of DT: black, DT>50 Mbp; blue, 30<DT≤50 Mbp; orange, 10<DT≤30 Mbp; red, DT≤10 Mbp.

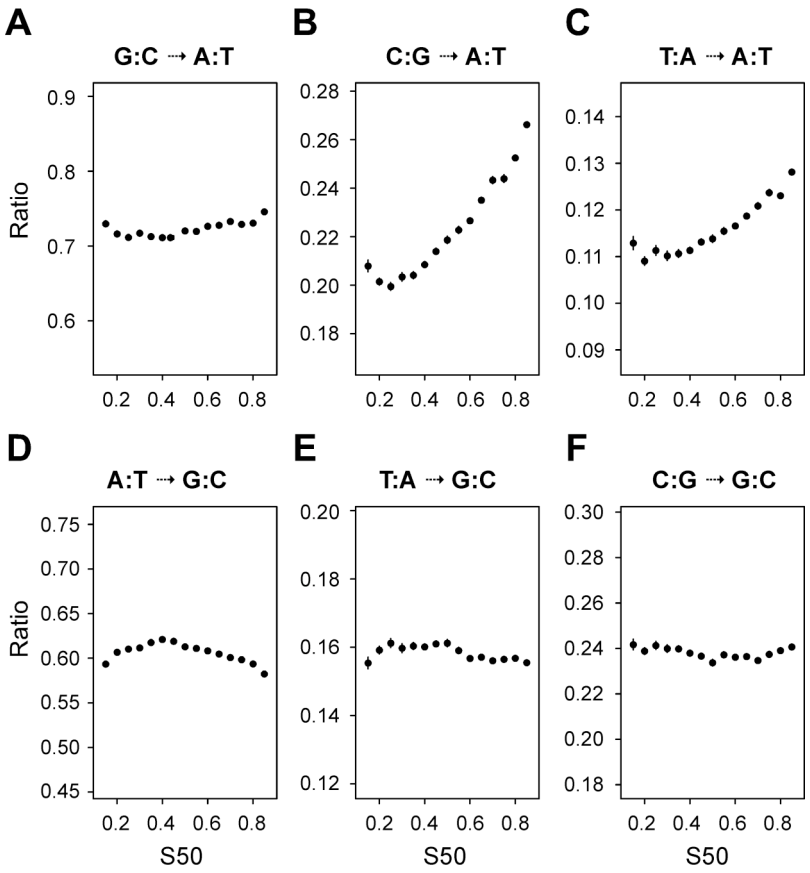

Figure S7. Variation of the relative non-CpG substitution rates with replication timing. A relative substitution rate is the ratio of an individual substitution rate to the global non-CpG substitution rate. Substitution rates were computed in non-coding regions, in 100-kbp windows (the ratio of extreme values on the ordinate axis is the same in all graphs).

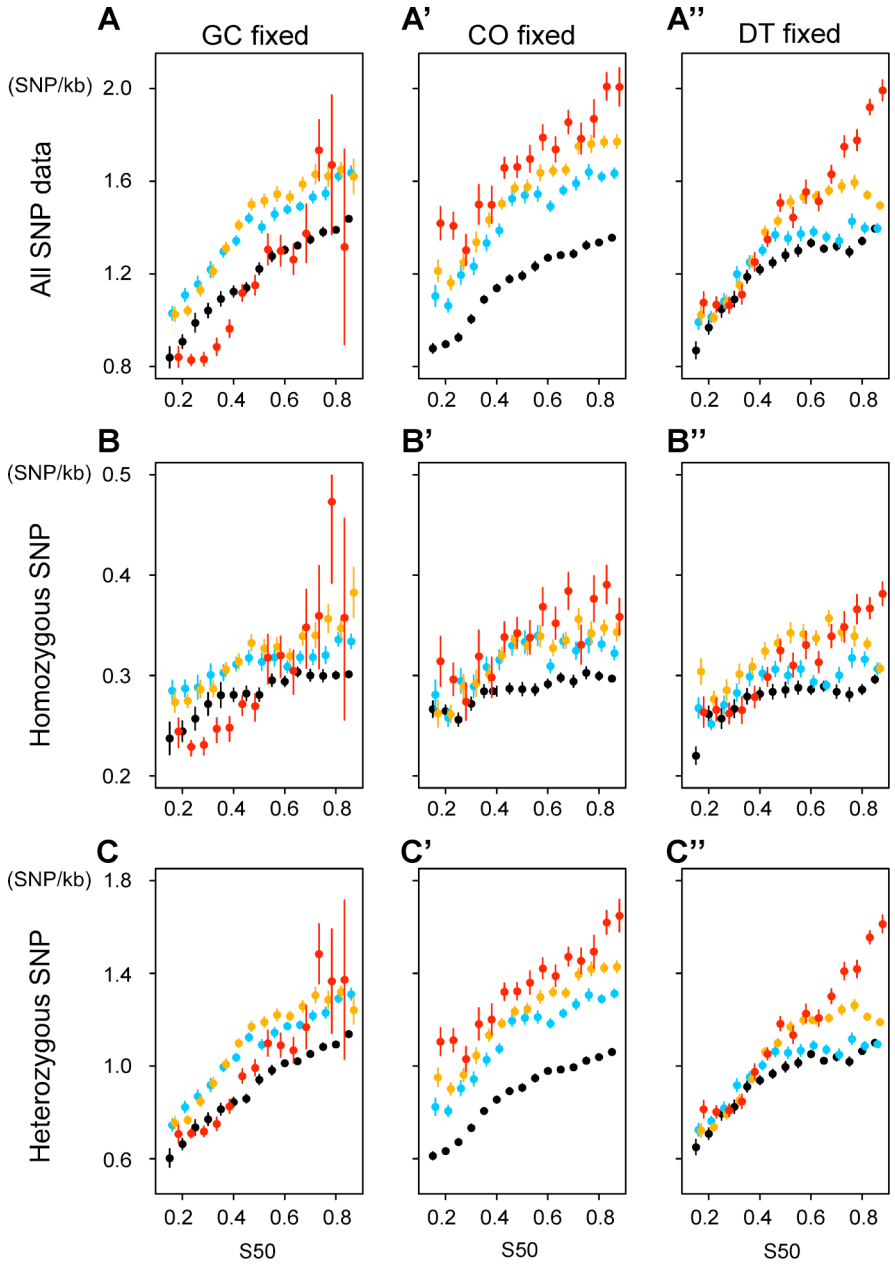

Figure S8. Variation of human non-CpG diversity with replication timing. Density of total SNP (A-A"), homozygous SNP (B-B"), and heterozygous SNP (C-C") of the YRI population; data retrieved from the International HAMAP project (Frazer et al. 2007a). (A-C) Variation of diversity when controlling for GC content: black, GC≤37%; blue, 37<GC≤42%; orange, 42<GC≤52%; red, GC>52%. (A'-C') Variation of diversity when controlling for crossover rate: black, CO≤1 cM/Mbp; blue, 1<CO≤2 cM/Mbp; orange, 2<CO≤4 cM/Mbp; red, CO>4 cM/Mbp. (A"-C") Variation of diversity when controlling for distance to telomeres: black, DT>50 Mbp; blue, 30<DT≤50 Mbp; orange, 10<DT≤30 Mbp; red, DT≤10 Mbp. Replication timing values and SNP densities are determined in 100-kbp windows. Similar results were obtained for the SNP data of JPT, CHB and CEU populations from the HAMAP project (data not shown).

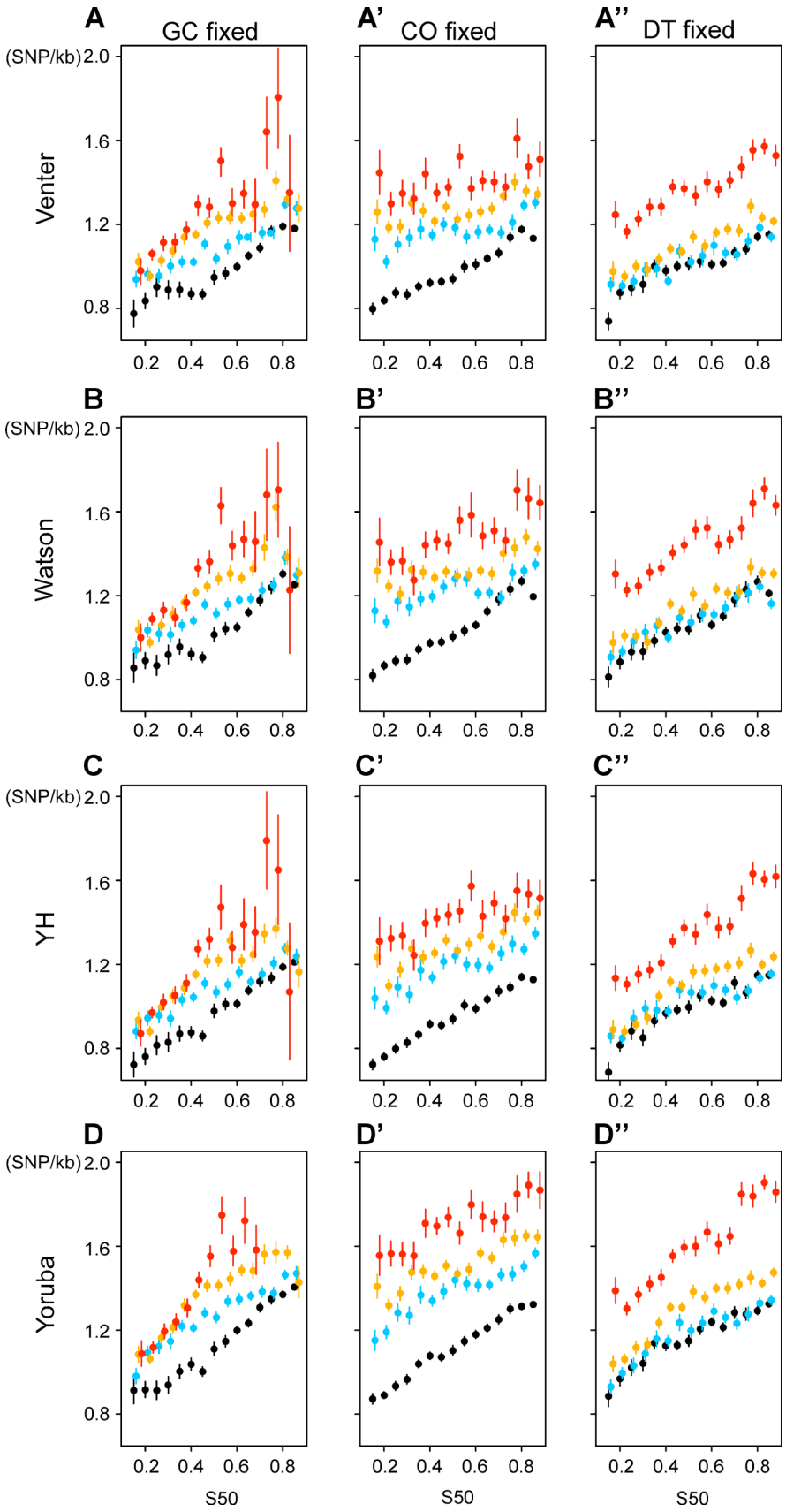
A
GC fixed
A'
CO fixed
A''
DT fixed
B
B'
B''
C
C'
C''
D
D'
D''
(SNP/kb)
Venter
Watson
YH
Yoruba
2.0
1.6
1.2
0.8
0.2
0.4
0.6
0.8
S50

Figure S9. Variation of the human non-CpG diversity in individual genomes with replication timing. Density of total SNP from Venter's genome (Levy et al. 2007) (A-A"), Watson's genome (Wheeler et al. 2008) (B-B"), YH genome (Wang et al. 2008) (C-C") and Yoruba's genome (Bentley et al. 2008) (D-D") as function of S50. (A-D) Variation of diversity when controlling for GC content: black, GC≤37%; blue, 37<GC≤42%; orange, 42<GC≤52%; red, GC>52%; (A'-D') variation diversity when controlling for CO: black, CO≤1 cM/Mbp; blue, 1<CO≤2 cM/Mbp; orange, 2<CO≤4 cM/Mbp; red, CO>4 cM/Mbp; (A"-D") variation of diversity when controlling for DT: black, DT>50 Mbp; blue, 30<DT≤50 Mbp; orange, 10<DT≤30 Mbp; red, DT≤10 Mbp. Replication timing values and SNP densities are determined in 100-kbp windows.

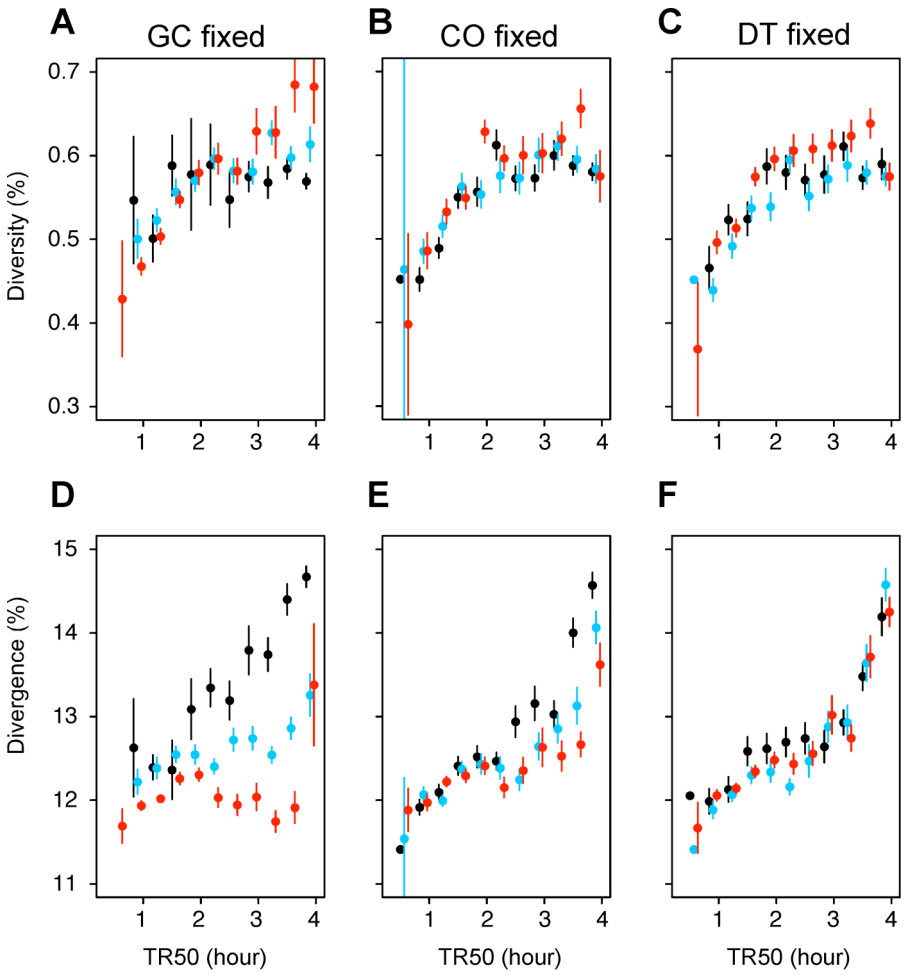

Figure S10. Variation of mouse diversity (A-C) and mouse-rat divergence (D-F) with replication timing. On the abscissa, replication timing (TR50) values determined in 1-Mbp windows (Farkash-Amar et al. 2008); in ordinate, the mean value of mouse diversity or mouse-rat divergence ± SEM in percent computed in non-coding regions. (A) Mouse diversity (Frazer et al. 2007b) as function of replication timing, when controlling for GC content: black, GC≤38%; blue, 38<GC≤43%; red, GC>43%. (B) Same as in (A) when controlling for crossover rate: black, CO≤0.5 cM/Mbp; blue, 0.5<CO≤1 cM/Mbp; red, CO>1 cM/Mbp. (C) Same as in (A) when controlling for distance to telomeres: black, DT>100 Mbp; blue, 50<DT≤100 Mbp; red, DT≤50 Mbp. (D-F) Mouse-rat divergence when controlling for GC (D), CO (E) and DT (F) as in (A-C). The relative increase of divergence (18%) is smaller than that of the diversity (35%). This small increase likely results from substitution saturation due to the long evolutionary time since mouse/rat diverged from their common ancestor.

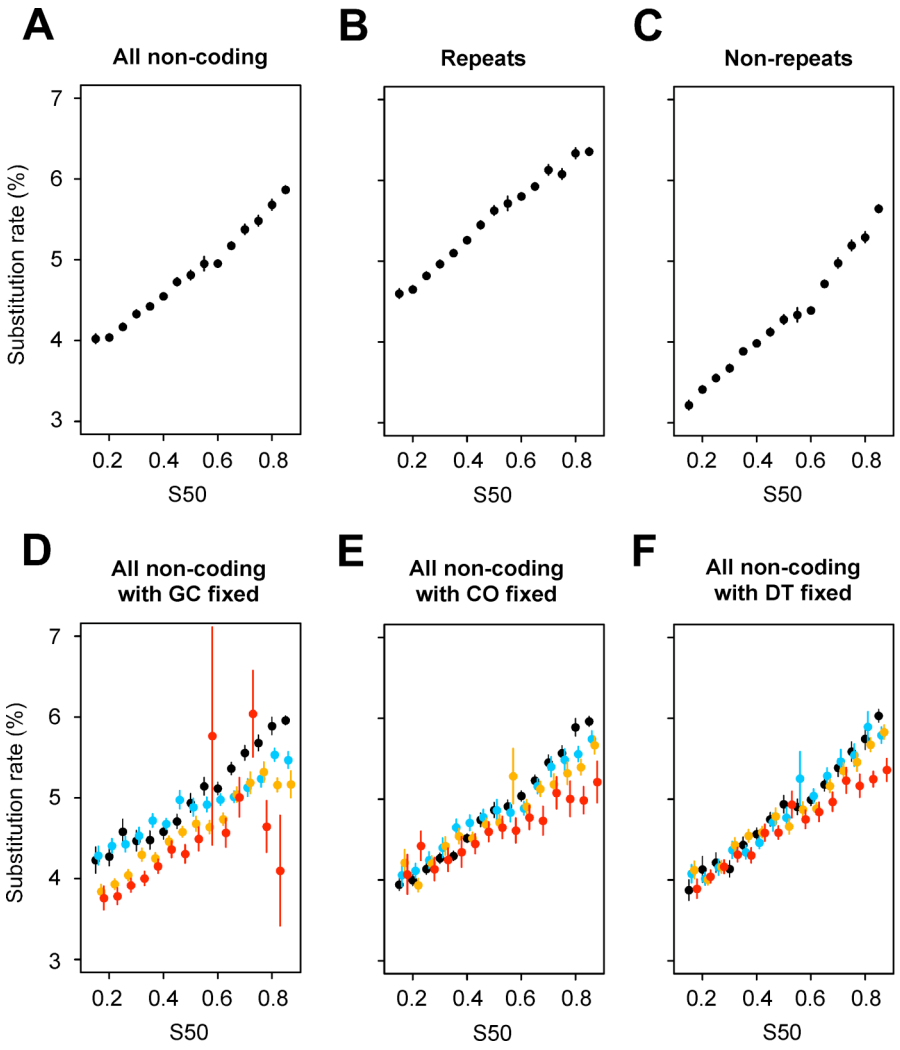

Figure S11. Variation of human CpG substitution rates with replication timing for all non-coding (A), repeated (B) and non-repeated (C) sequences. Variation of CpG substitution rates for non-coding sequences when controlling for GC (D), CO (E) and DT (F). Black, GC≤37%, CO≤1 cM/Mbp, DT>50 Mbp; blue, 37<GC≤42%, 1<CO≤2 cM/Mbp, 30<DT≤50 Mbp; orange, 42<GC≤52%, 2<CO≤4 cM/Mbp, 10<DT≤30 Mbp; red, GC>52%, CO>4 cM/Mbp, DT≤10 Mbp. Replication timing values and CpG substitution rates are determined in 100 kbp windows. It is noteworthy that the rate is higher in repeated than in non-repeated sequences which likely results from higher mean methylation level in repeated than in non-repeated elements (Goll and Bestor 2005).

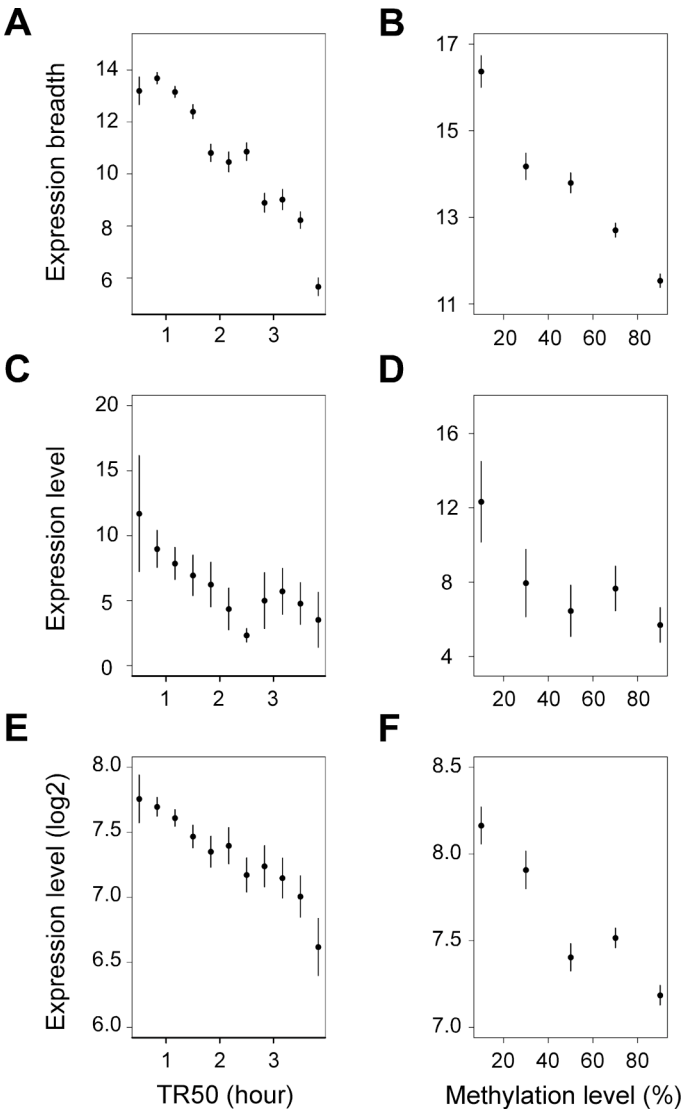

Figure S12. Dependence of mouse gene expression breadth (Semon et al. 2005) on replication timing (A) and on methylation level (B). Dependence of mouse expression level in embryonic stem cells (Semon et al. 2005) on replication timing (C) and on methylation level (D). Dependence of mouse expression level in spermatogonia (Chalmel et al. 2007) with replication timing (E) and with methylation level (F). Methylation levels were retrieved from (Meissner et al. 2008). Replication timing values were retrieved from (Farkash-Amar et al. 2008). Expression breadth, expression level, methylation level and replication timing were computed for each gene.

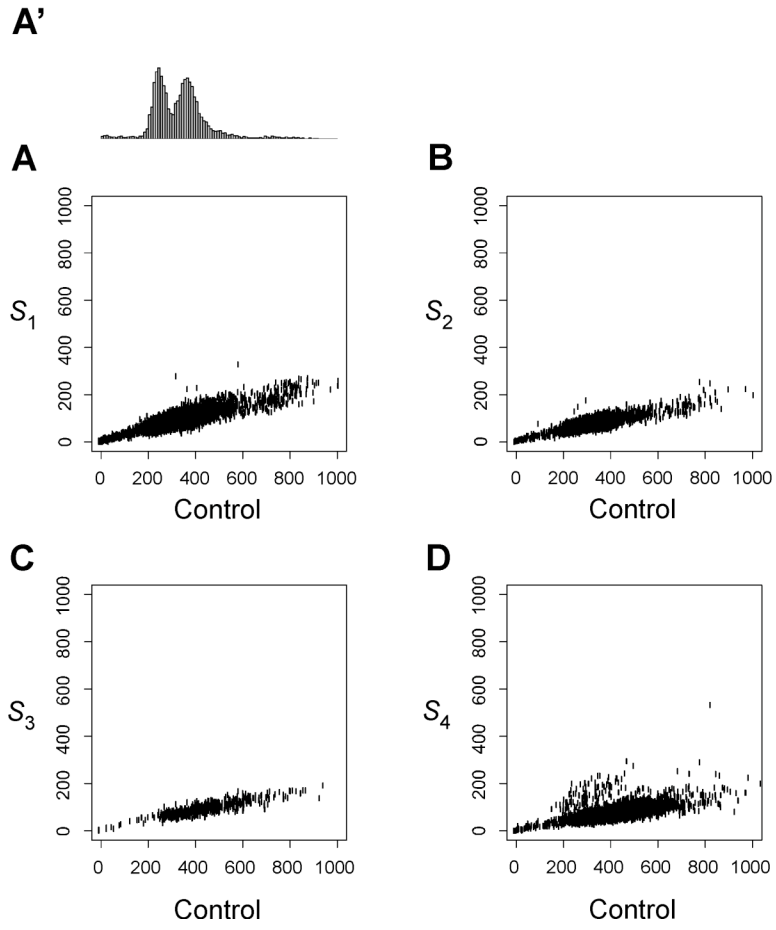

Figure S13. Correlations between the density of sequence reads (reads/100kbp) for each sample of the S_i periods of the S phase and for the control sample, within the corresponding background regions: (A) S_1; (B) S_2; (C) S_3 and (D) S_4 (see Methods). (A') shows the bimodal distribution of the density of reads for the control sample, within the background regions of S_1. Similar results were observed for the S_2, S_3 and S_4 periods.

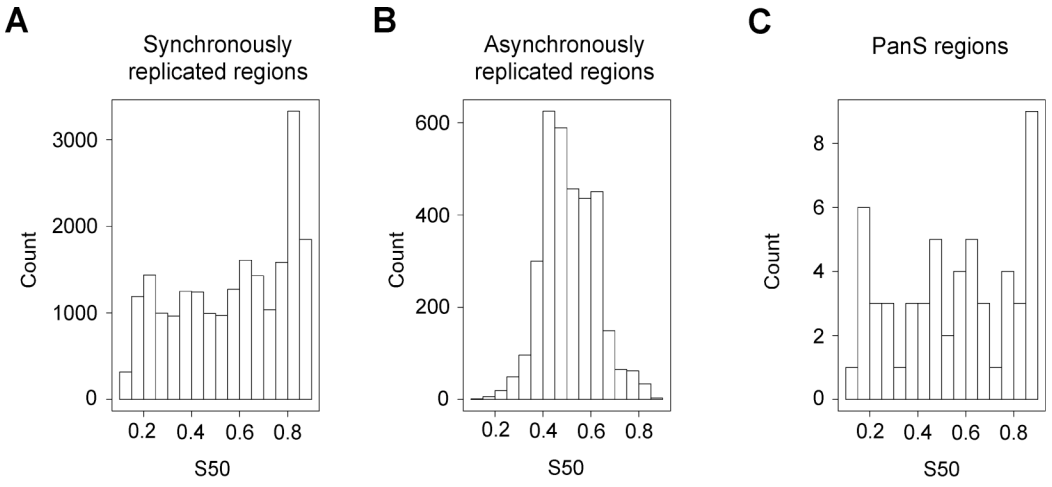

Figure S14. Replication timing profile (S50): (A) for regions that are synchronously replicated; (B) for regions that are asynchronously replicated; (C) for pan-S regions defined in (Karnani et al. 2007) (see main text, Methods).

Supplemental Tables S1 to S7

Table S1. Partial correlation analysis of human substitution rates and diversity. Analyzing variables are GC, CO, LDT and S50 defined in main text, Table 1; analysis was performed in 100kbp-windows.

	GC|(CO,LDT,S50)		CO|(GC,LDT,S50)		LDT|(GC,CO,S50)		S50|(GC,CO,LDT)	
	R	**P**	**R**	**P**	**R**	**P**	**R**	**P**
Global Rate	-0.07	2.0×10^{-29}	0.25	$<1 \times 10^{-300}$	-0.31	$<1 \times 10^{-300}$	0.29	$<1 \times 10^{-300}$
N→S								
A:T→G:C	-0.03	4.7×10^{-8}	0.33	$<1 \times 10^{-300}$	-0.39	$<1 \times 10^{-300}$	0.25	$<1 \times 10^{-300}$
T:A→G:C	-0.06	2.7×10^{-21}	0.26	$<1 \times 10^{-300}$	-0.35	$<1 \times 10^{-300}$	0.18	1.7×10^{-178}
C:G→G:C	-0.08	9.5×10^{-33}	0.13	1.8×10^{-90}	-0.27	$<1 \times 10^{-300}$	0.16	1.5×10^{-151}
N→W								
G:C→A:T (non CpG)	-0.10	1.4×10^{-59}	0.08	1.8×10^{-39}	-0.11	1.1×10^{-62}	0.26	$<1 \times 10^{-300}$
C:G→A:T	-0.17	9.9×10^{-164}	0.05	6.5×10^{-17}	-0.07	1.4×10^{-29}	0.29	$<1 \times 10^{-300}$
T:A→A:T	-0.15	4.5×10^{-135}	0.14	5.3×10^{-112}	-0.14	3.3×10^{-106}	0.21	2.2×10^{-255}
CpG								
G:C→A:T	-0.07	1.4×10^{-31}	<-0.01	0.48	-0.01	0.07	0.13	1.6×10^{-97}
Diversity (YRI)	-0.12	6.0×10^{-79}	0.34	$<1 \times 10^{-300}$	-0.08	4.0×10^{-35}	0.15	3.3×10^{-135}
" " (Venter)	0.01	0.02	0.17	1.0×10^{-168}	-0.12	1.6×10^{-84}	0.13	5.2×10^{-101}
" " (Watson)	0.01	0,13	0,16	6.6×10^{-151}	-0,12	1.4×10^{-88}	0,14	4.0×10^{-114}
" " (YH)	-0,02	7.4×10^{-5}	0,21	1.9×10^{-254}	-0,12	7.1×10^{-84}	0,13	1.5×10^{-92}
" " (Yoruba)	-0.02	2.0×10^{-3}	0.26	$<1 \times 10^{-300}$	-0.15	1.1×10^{-127}	0.16	9.0×10^{-148}

Table S2. Partial correlation analysis of human substitution rates and diversity. Analyzing variables are as in Table S1 with 1-Mbp windows.

	GC|(CO,LDT,S50)		CO|(GC,LDT,S50)		LDT|(GC,CO,S50)		S50|(GC,CO,LDT)	
	R	P	R	P	R	P	R	P
Global Rate	-0.17	5.9×10^{-19}	0.36	7.8×10^{-89}	-0.37	3.1×10^{-92}	0.34	1.3×10^{-78}
N→S								
A:T→G:C	-0.17	4.0×10^{-19}	0.50	1.4×10^{-194}	-0.47	2.1×10^{-163}	0.28	2.1×10^{-51}
T:A→G:C	-0.19	4.0×10^{-23}	0.44	1.4×10^{-137}	-0.43	5.3×10^{-133}	0.21	8.5×10^{-28}
C:G→G:C	-0.13	2.2×10^{-11}	0.21	6.0×10^{-29}	-0.31	3.2×10^{-61}	0.19	4.3×10^{-23}
N→W								
G:C→A:T (non CpG)	-0.18	4.7×10^{-20}	0.11	9.1×10^{-9}	-0.18	1.2×10^{-21}	0.39	5.3×10^{-102}
C:G→A:T	-0.28	7.7×10^{-50}	0.03	0.09	-0.15	3.8×10^{-15}	0.41	3.8×10^{-115}
T:A→A:T	-0.27	1.7×10^{-57}	0.24	2.5×10^{-36}	-0.20	3.3×10^{-25}	0.32	8.5×10^{-67}
CpG								
G:C→A:T	-0.13	3.1×10^{-11}	-0.03	0.07	-0.06	0.4×10^{-3}	0.30	2.6×10^{-59}
Diversity (YRI)	-0.28	2.5×10^{-53}	0.50	1.4×10^{-197}	-0.04	0.05	0.06	9.7×10^{-4}
" " (Venter)	-0.07	9.8×10^{-5}	0.30	1.1×10^{-61}	-0.17	5.1×10^{-19}	0.14	5.7×10^{-13}
" " (Watson)	-0.07	4.9×10^{-4}	0.27	1.6×10^{-49}	-0.18	3.6×10^{-21}	0.16	4.4×10^{-18}
" " (YH)	-0.16	9.5×10^{-18}	0.36	2.6×10^{-90}	-0.16	5.2×10^{-17}	0.10	2.1×10^{-7}
" " (Yoruba)	-0.16	4.0×10^{-18}	0.43	4.1×10^{-133}	-0.18	1.4×10^{-21}	0.13	4.6×10^{-12}

Table S3. Multivariate regression analysis of substitution rates and diversity. Same as in Table 1 of the main text with 1-Mbp windows.

	GC			CO			LDT			S50			Full model	
	slope	%	P	slope	%	P	slope	%	P	slope	%	P	R^2	P
Global Rate	-0.21	6.8	1×10^{-18}	0.35	29	8×10^{-83}	-0.38	31	9×10^{-86}	0.39	33	7×10^{-74}	0.49	$<1 \times 10^{-300}$
N→S														
A:T→G:C	-0.19	-0.6	7×10^{-19}	0.47	46	5×10^{-168}	-0.45	43	3×10^{-144}	0.28	12	2×10^{-49}	0.59	$<1 \times 10^{-300}$
T:A→G:C	-0.23	-0.1	1×10^{-22}	0.43	44	1×10^{-123}	-0.45	45	6×10^{-120}	0.22	10	3×10^{-27}	0.51	$<1 \times 10^{-300}$
C:G→G:C	-0.19	5.9	3×10^{-11}	0.23	26	3×10^{-28}	-0.36	45	2×10^{-58}	0.24	24	1×10^{-22}	0.28	8×10^{-191}
N→W														
G:C→A:T (non CpG)	-0.23	25	9×10^{-20}	0.11	1.2	1×10^{-8}	-0.19	2.7	3×10^{-21}	0.48	71	4×10^{-94}	0.40	5×10^{-295}
C:G→A:T	-0.34	40	6×10^{-48}	0.03	-0.6	0.09	-0.14	-2.5	5×10^{-15}	0.46	63	3×10^{-105}	0.51	$<1 \times 10^{-300}$
T:A→A:T	-0.36	38	1×10^{-46}	0.23	7.5	3×10^{-35}	-0.20	4.4	1×10^{-24}	0.38	50	2×10^{-63}	0.43	$<1 \times 10^{-300}$
CpG														
G:C→A:T	-0.18	29	4×10^{-11}	-0.04	1.8	0.07	-0.06	-2.5	4×10^{-3}	0.40	72	1×10^{-56}	0.29	$<1 \times 10^{-300}$
Diversity (YRI)	-0.40	22	4×10^{-51}	0.58	70	1×10^{-170}	-0.04	1.9	0.05	0.08	5.6	1×10^{-3}	0.35	2×10^{-249}
" " (Venter)	-0.11	-0.4	1×10^{-4}	0.35	59	8×10^{-59}	-0.20	28	9×10^{-19}	0.18	13	7×10^{-13}	0.23	3×10^{-148}
" " (Watson)	-0.10	0.8	5×10^{-4}	0.31	50	1×10^{-47}	-0.21	30	7×10^{-21}	0.22	19	7×10^{-13}	0.22	4×10^{-144}
" " (YH)	-0.24	5.8	2×10^{-17}	0.41	64	3×10^{-84}	-0.19	21	8×10^{-17}	0.13	9.4	2×10^{-7}	0.26	7×10^{-173}
" " (Yoruba)	-0.23	2.9	7×10^{-18}	0.47	67	3×10^{-120}	-0.20	20	3×10^{-21}	0.16	10	6×10^{-12}	0.33	2×10^{-236}

Table S4. Partial correlation analysis of individual relative substitution rates. A relative substitution rate is the ratio of an individual substitution rate to the global non-CpG substitution rate, measured in 100-kbp windows as indicated in Fig. S7. Only C:G→A:T and T:A→A:T substitution rates display significant positive correlation with replication timing.

	S50|(GC,CO,LDT)	P
G:C→A:T	-0.02	3.3×10^{-4}
C:G→A:T	0.16	6.8×10^{-141}
T:A→A:T	0.05	6.0×10^{-13}
A:T→G:C	-0.03	3.5×10^{-6}
C:G→G:C	-0.01	0.03
T:A→G:C	-0.02	2.0×10^{-4}
CpG	0.01	0.15

Table S5. Multivariate regression analysis of human CpG transition rate and non-CpG global substitution rate using GC content, methylation level and replication timing as predictors (Methods). Same analysis with the mouse CpG diversity.

		GC		Methylation		S50		Full model	
		%	P	%	P	%	P	R^2	P
CpG	Human	21.0	0.03	77.5	1.7×10^{-6}	1.5	0.82	0.04	4.1×10^{-7}
	Mouse	14.5	7.1×10^{-4}	73.5	2.3×10^{-62}	12.0	1.4×10^{-3}	0.20	5.3×10^{-98}
Non-CpG	Human	10.6	2.6×10^{-4}	28.5	1.3×10^{-3}	60.9	7.0×10^{-8}	0.05	1.8×10^{-9}
	Mouse	4.3	0.48	30.4	1.8×10^{-13}	65.3	1.1×10^{-13}	0.12	2.7×10^{-56}

Table S6. Sequencing data.

		Control	S_1	S_2	S_3	S_4
Experiment 1	Total	15,272,351	6,929,165	7,122,751	7,126,848	6,763,692
	Mapped	13,387,734	5,813,777	6,035,088	5,974,135	5,476,494
	Uniquely mapped	10,455,231	4,219,675	4,575,618	4,540,586	3,857,883
Experiment 2	Total		8,844,035	8,613,551	15,022,063	15,464,546
	Mapped		7,542,379	7,450,728	12,756,054	13,105,749
	Uniquely mapped		5,691,001	5,791,303	9,863,398	9,670,309

Table S7. Relative amounts (%) of genomic regions that replicate at different periods S_i of the S phase. Synchronously and asynchronously replicated regions are defined in the main text (Methods). Values correspond to mean relative amounts of experiments 1 and 2.

	Non-significantly enriched in all periods	Synchronously replicated						Asynchronously replicated					
		Significantly enriched in *n* adjacent period(s)				Significantly enriched in *n* non-adjacent periods		Significantly enriched in *n* adjacent period(s)				Significantly enriched in *n* non- adjacent periods	
		n=1	*n*=2	*n*=3	*n*=4	*n*=2	*n*=3 [a]	*n*=1 [b]	*n*=2	*n*=3	*n*=4	*n*=2	*n*=3 [a]
		21.45	26.36	20.82	13.98	1.31	3.66	0.24	0.07	2.42	4.19	0.24	0.25
Sum (%)		82.61				4.97		6.92				0.49	
Sum (%)	5.01	87.58						7.41					

[a] For example, regions significantly enriched in the S_1, S_2 and S_4, or S_1, S_3 and S_4 periods.
[b] Although only one window is significantly enriched, the corresponding enrichment is too small for this window to be synchronously replicated.

References

Bentley, D.R. Balasubramanian, S. Swerdlow, H.P. Smith, G.P. Milton, J. Brown, C.G. Hall, K.P. Evers, D.J. Barnes, C.L. Bignell, H.R. et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53-59.

Berglund, J., Pollard, K.S., and Webster, M.T. 2009. Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol.* **7:** e26.

Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B. et al. 2007. The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. USA* **104:** 8346-8351.

Duret, L. and Arndt, P.F. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* **4:** e1000071.

Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. 2008. Global organization of replication time zones of the mouse genome. *Genome Res.* **18:** 1562-1570.

Frazer, K.A. Ballinger, D.G. Cox, D.R. Hinds, D.A. Stuve, L.L. Gibbs, R.A. Belmont, J.W. Boudreau, A. Hardenbol, P. Leal, S.M. et al. 2007a. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449:** 851-861.

Frazer, K.A., Eskin, E., Kang, H.M., Bogue, M.A., Hinds, D.A., Beilharz, E.J., Gupta, R.V., Montgomery, J., Morenzoni, M.M., Nilsen, G.B. et al. 2007b. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448:** 1050-1053.

Galtier, N., Duret, L., Glemin, S., and Ranwez, V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25:** 1-5.

Goll, M.G. and Bestor, T.H. 2005. Eukaryotic cytosine methyltransferases. *Ann. Rev. Biochem.* **74:** 481-514.

Hwang, D.G. and Green, P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl. Acad. Sci. USA* **101:** 13994-14001.

Karnani, N., Taylor, C., Malhotra, A., and Dutta, A. 2007. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome Res.* **17:** 865-876.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G. et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* **5:** e254.

Ma, J., Zhang, L., Suh, B.B., Raney, B.J., Burhans, R.C., Kent, W.J., Blanchette, M., Haussler, D., and Miller, W. 2006. Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16:** 1557-1565.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B. et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454:** 766-770.

Semon, M., Mouchiroud, D., and Duret, L. 2005. Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance. *Hum. Mol. Genet.* **14:** 421-427.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y. et al. 2008. The diploid genome sequence of an Asian individual. *Nature* **456:** 60-65.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452:** 872-876.

Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I., and Carter, N.P. 2004. Replication timing of the human genome. *Hum. Mol. Genet.* **13:** 191-202.