

**Supplementary Materials for**  
**“Fast Evolution of Core Promoters in Primate Genomes”**

Han Liang, Yeong-Shin Lin and Wen-Hsiung Li\*

This file includes the following figure and tables:

1. Figure S1

The proportion of transitions or AT↔GC substitutions in core promoters

2. Table S1

Inferred positively selected human core promoters

3. Table S2

Rate analysis on inferred positively selected human core promoters

4. Table S3

GO term analysis of inferred positively selected genes

5. Table S4

The SNPs with an extreme *iHS* value

Fig. S1. The proportion of transitions or AT↔GC substitutions in core promoters

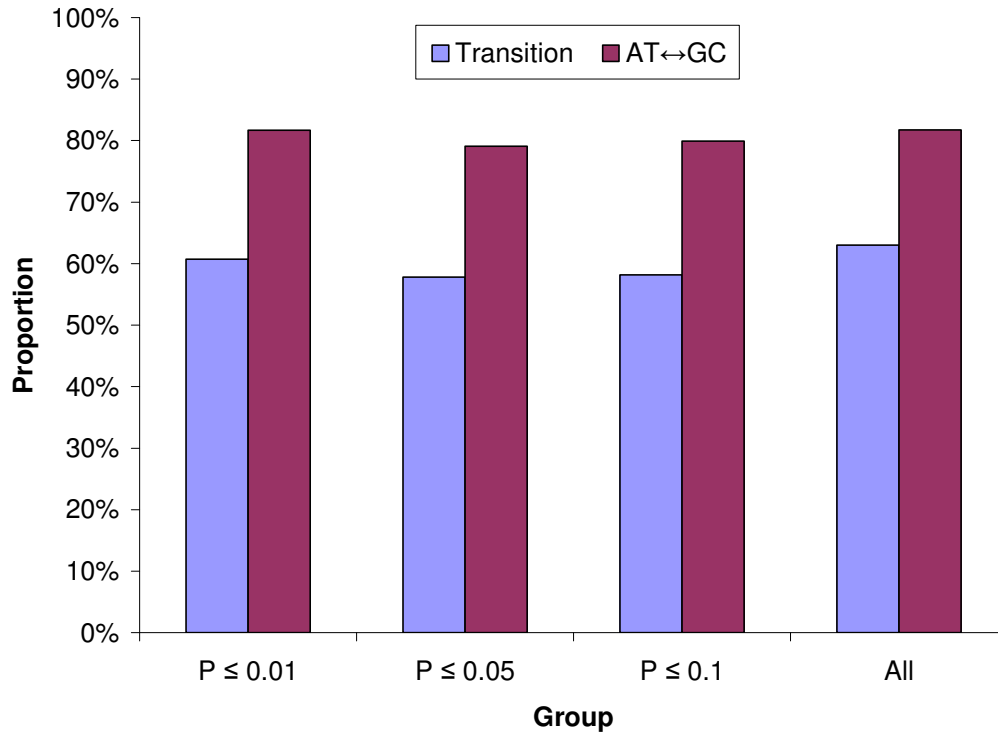


Table S1. Inferred positively selected human core promoters

Promote location (Hg18)	Gene_ID	Gene_Symbol	RefSeq	K <sub>core</sub>	K <sub>pseudo</sub>	K <sub>4</sub>
chr1:23367840-23367959	7798	LUZP1	NM_033631	0.1502	0.0187	0.0702
chr1:151805302-151805421	6273	S100A2	NM_005978	0.1433	0.0175	0.1067
chr2:120233635-120233754	5775	PTPN4	NM_002830	0.1673	0.0085	0.0282
chr2:179051624-179051743	51661	FKBP7	NM_181342	0.0895	0.0000	0.0328
chr3:51397224-51397343	7873	ARMET	NM_006010	0.1774	0.0000	0.0222
chr6:74287320-74287439	1915	EEF1A1	NM_001402	0.0831	0.0000	0.0159
chr6:133177451-133177570	6206	RPS12	NM_001016	0.3223	0.0421	0.0972
chr7:44496899-44497018	23386	NUDCD3	NM_015332	0.1868	0.0288	0.0458
chr8:11697171-11697290	2222	FDFT1	NM_004462	0.2301	0.0309	0.0613
chr8:25372694-25372813	157313	CDCA2	NM_152562	0.2002	0.0093	0.0464
chr9:38383103-38383222	219	ALDH1B1	NM_000692	0.1400	0.0175	0.0569
chr9:94095703-94095822	3376	IARS	NM_013417	0.1480	0.0088	0.0468
chr11:120668866-120668985	6309	SC5DL	NM_001024956	0.4687	0.0182	0.0648
chr11:122437507-122437626	3312	HSPA8	NM_006597	0.2047	0.0110	0.0595
chr12:52131830-52131949	5094	PCBP2	NM_005016	0.0751	0.0000	0.0155
chr12:55129071-55129190	8914	TIMELESS	NM_003920	0.1048	0.0000	0.0313
chr13:46268830-46268949	2098	ESD	NM_001984	0.1148	0.0088	0.0821
chr16:86357152-86357271	54758	KLHDC4	NM_017566	0.2132	0.0278	0.0601
chr19:8484945-8485064	84330	ZNF414	NM_032370	0.1209	0.0097	0.0370
chr20:1047121-1047240	9491	PSMF1	NM_006814	0.1296	0.0000	0.0000
chr20:17498496-17498615	11034	DSTN	NM_001011546	0.2128	0.0093	0.0000
chr20:21231755-21231874	22803	XRN2	NM_012255	0.1418	0.0177	0.0220
chr22:39676601-39676720	9978	RBX1	NM_014248	0.1157	0.0096	0.0000
chrX:150315926-150316045	203547	LOC203547	NM_001017980	0.0862	0.0000	0.0278

Gene_Symbol	GO terms
LUZP1	N.A.
S100A2	Binding
PTPN4	catalytic activity; protein binding; hydrolase activity; structural molecule activity;
FKBP7	catalytic activity; protein binding; isomerase activity
ARMET	protein binding
EEF1A1	catalytic activity; protein binding; nucleic acid binding; hydrolase activity; translation regulator activity
RPS12	nucleic acid binding; structural molecule activity
NUDCD3	N.A.
FDFT1	catalytic activity; oxidoreductase activity; transferase activity
CDCA2	protein binding
ALDH1B1	catalytic activity; oxidoreductase activity;
IARS	catalytic activity; ligase activity
SC5DL	catalytic activity; oxidoreductase activity
HSPA8	catalytic activity; protein binding; hydrolase activity
PCBP2	protein binding; nucleic acid binding
TIMELESS	protein binding;
ESD	catalytic activity; hydrolase activity
KLHDC4	N.A.
ZNF414	nucleic acid binding
PSMF1	protein binding; enzyme regulator activity
DSTN	protein binding;
XRN2	catalytic activity; protein binding; nucleic acid binding; hydrolase activity
RBX1	protein binding;
LOC203547	N.A.

Table S2. Rate analysis on inferred positively selected human core promoters

Promoter Location	$K_{\text{pseudo}}^*$	$K_{\text{between}}$	$K_{\text{core}}$	$K_{\text{intron}}$	site_4f	n1	n2
chr1:23367840-23367959	0.039	0.0444	0.1502	0.0529	146	6	4
chr1:151805302-151805421	0.0308	0.0458	0.1433	0.0588	34	8	5
chr2:120233635-120233754	0.0625	0.0496	0.1673	0.0398	337	8	4
chr2:179051624-179051743	0.0516	0.0546	0.0895	0.0645	65	4	3
chr3:51397224-51397343	0.0622	0.041	0.1774	0.0557	47	6	8
chr6:74287320-74287439	0.0293	0.0406	0.0831	NA	195	2	5
chr6:133177451-133177570	0.0398	0.0598	0.3223	0.02	49	12	8
chr7:44496899-44497018	0.1045	0.0612	0.1868	0.0367	119	7	7
chr8:11697171-11697290	0.0593	0.0739	0.2301	0.0842	128	10	5
chr8:25372694-25372813	0.0334	0.0634	0.2002	0.0517	306	5	8
chr9:38383103-38383222	0.0521	0.0535	0.14	0.0385	176	8	3
chr9:94095703-94095822	0.0552	0.0642	0.148	0.0712	444	4	6
chr11:120668866-120668985	0.0473	0.0615	0.4687	0.0444	87	4	0
chr11:122437507-122437626	0.0886	0.0912	0.2047	0.1254	263	10	6
chr12:52131830-52131949	0.059	0.0612	0.0751	0.0228	133	5	2
chr12:55129071-55129190	0.0297	0.0388	0.1048	0.0561	407	4	4
chr13:46268830-46268949	0.0453	0.0724	0.1148	0.0636	99	7	2
chr16:86357152-86357271	0.0885	0.0898	0.2132	0.0758	149	8	5
chr19:8484945-8485064	0.0419	0.0881	0.1209	0.0793	116	7	2
chr20:1047121-1047240	0.0409	0.0524	0.1296	0.0668	102	5	4
chr20:17498496-17498615	0.0368	0.0476	0.2128	0.0623	49	9	7
chr20:21231755-21231874	0.0504	0.0637	0.1418	0.0572	332	6	4
chr22:39676601-39676720	0.075	0.0758	0.1157	0.0899	27	2	8
chrX:150315926-150316045	0.0334	0.0266	0.0862	0.0365	38	5	2

$K_{\text{pseudo}}^*$ , the rate of extended pseudo promoter, which is calculated from the 1 kb upstream non-coding region that is 1 kb away from a core promoter

$K_{\text{between}}$  is calculated from the region between a core promoter and the corresponding pseudo promoter

$K_{\text{intron}}$ , the rate of intron region, which is calculated from a 1 kb randomly chosen intronic region

site\_4f, the number of four-fold degenerate sites

n1, the number of substitutions at the 35-bp flanking regions within a core promoter

n2, the number of substitutions at the middle 50 bp of a core promoter

Table S3. GO term analysis of inferred positively selected genes

GO Term	Category description	<i>n</i>	<i>N</i>	<i>PI</i>	<i>P2</i>
Biological process					
GO:0008152	metabolic process	16	1113	0.05	/
GO:0009987	cellular process	16	1378	0.39	/
GO:0043170	macromolecule metabolic process	14	957	0.076	/
GO:0006139	nucleobase\, nucleoside\, nucleotide and nucleic acid metabolic process	6	590	/	0.54
GO:0009058	biosynthetic process	5	176	0.035	/
GO:0050789	regulation of biological process	4	660	/	0.11
GO:0050896	response to stimulus	4	218	0.20	/
GO:0009056	catabolic process	3	96	0.082	/
GO:0006928	cell motility	2	50	0.10	/
GO:0007275	multicellular organismal development	1	258	/	0.21
GO:0006519	amino acid and derivative metabolic process	1	40	0.354	/
GO:0006118	electron transport	0	39	/	0.65
GO:0006810	transport	0	253	/	0.05
GO:0030154	cell differentiation	0	221	/	0.079
GO:0007154	cell communication	0	359	/	0.014
GO:0008219	cell death	0	130	/	0.24
GO:0046903	secretion	0	45	/	0.61
Molecular function					
GO:0005488	binding	16	1400	0.43	/
GO:0003824	catalytic activity	10	628	0.1	/
GO:0005515	protein binding	10	846	0.43	/
GO:0003676	nucleic acid binding	5	516	/	0.50
GO:0016787	hydrolase activity	5	249	0.12	/
GO:0016491	oxidoreductase activity	3	77	0.048	/
GO:0005198	structural molecule activity	2	73	0.19	/
GO:0030234	enzyme regulator activity	1	89	0.63	/
GO:0016874	ligase activity	1	70	0.54	/
GO:0016740	transferase activity	1	215	/	0.31
GO:0030528	transcription regulator activity	0	227	/	0.074
GO:0004871	signal transducer activity	0	132	/	0.23
GO:0004872	receptor activity	0	82	/	0.41
GO:0016301	kinase activity	0	111	/	0.29
GO:0005215	transporter activity	0	79	/	0.42
Cellular Component					
GO:0005623	cell	18	1532	0.30	/
GO:0005622	intracellular	18	1312	0.041	/
GO:0005737	cytoplasm	12	640	0.015	/
GO:0005634	nucleus	7	737	/	0.43
GO:0016020	membrane	3	493	/	0.18
GO:0009986	cell surface	1	16	0.16	/
GO:0005694	chromosome	0	58	/	0.53
GO:0005576	extracellular region	0	76	/	0.43

The *P*-value is the raw value from the binomial test. *n*, # of positively selected genes; *N*, # of annotated genes; *PI*, *P*-value for over-representation; *P2*, *P*-value for under-representation.

Table S4. The SNPs with an extreme *iHS* value

SNP_ID	Population	his	Promoter_location	Gene_ID	Gene_Symbol
rs10796921	CEU	-2.653	chr1:35794895-35795014	79932	KIAA0319L
rs737381	CEU	2.763	chr1:36393739-36393858	55700	MAP7D1
rs6766480	ASN	3.034	chr3:197498468-197498587	5130	PCYT1A
rs3806980	ASN	-2.981	chr6:74287920-74288039	1915	EEF1A1
rs2278203	CEU	-2.507	chr8:82760774-82760893	3612	IMPA1
rs2275639	CEU	-2.523	chr9:110735899-110736018	8518	IKBKAP
rs1571856	CEU	3.202	chr9:134271896-134272015	7270	TTF1
rs2070627	YRI	4.893	chr12:123964605-123964724	7316	UBC
rs2273315	CEU	-2.570	chr14:38642473-38642592	10484	SEC23A
rs8066463	YRI	-2.536	chr17:59172099-59172218	92335	LYK5