Supplementary Material

Authors: Ying, Epps, Williams, Huttley

Rate heterogeneity between DHS and Flank is not due to conserved sequences

We addressed the possibility that the reduced substitution rate of DHS compared with Flank regions was due to purifying natural selection on functional elements by eliminating sequence annotated as conserved. Constrained element classifications were obtained from the UCSC Table browser for human database hg18 with the table named "phastConsElements17ways" [1]. This table provides segment coordinates of high conservation scores from comparison of vertebrates by a two-state phylo-HMM. Conserved elements coincident with sampled DHS+Flank alignments range from 10 to hundreds of bp in length with more than half of them less than 40 bp. Any alignment with sequence positions annotated as conserved were eliminated, irrespective of the length of the annotation. This resulted in 2986 intergenic and 2936 intronic alignments respectively.

LR tests revealed the same results, although the statistical power was weakened due to the smaller number of loci examined. The excessive loci with lower substitution rate at DHS than Flank remained strong for both intergenic and intronic regions. The general transition and CpG transition substitution differences between DHS and Flank were also evident for intergenic region, but not in intron.

Table S1 Differences in substitutions between DHS and Flank regions with constrained element excluded

	Intronic			Intergenic		
Null Hypothesis	DHS $<$ Flank	DHS $>$ Flank	p	DHS $<$ Flank	DHS $>$ Flank	p
$K_{DHS} = K_{Flank}$	179	105	6.7×10^{-6}*	207	105	4.0×10^{-9}*
$\lambda_{DHS} = \lambda_{Flank}$	94	83	0.22	114	73	0.0017*
$CG.\lambda_{DHS} = CG.\lambda_{Flank}$, $\lambda_{DHS} \neq \lambda_{Flank}$	137	130	0.36	178	106	1.2×10^{-5}*
$\lambda_{DHS} = \lambda_{Flank}$, $CG.\lambda_{DHS} \neq CG.\lambda_{Flank}$	94	75	0.08	97	79	0.10

Loci with significant correlations between K and nucleosome score

Table S2 Loci with correlated K and nucleosome scores. p_{boot} is the probability that the estimate of the correlation coefficient $\hat{\rho}$ is not equal to 0, estimated using a bootstrap procedure with 2000 replicates. The listed loci were significant after correcting for multiple tests using the sequential Bonferroni.

Symbol	$\hat{\rho}$	p_{boot}
PRELP	-0.6740	1.0000
EYA1	-0.6682	1.0000
GIP	-0.6675	1.0000
GNAI2	-0.6517	1.0000
DAP	-0.6345	1.0000
PTP4A1	-0.6301	1.0000
PYGL	-0.6245	1.0000
GRAP2	-0.5757	1.0000
RHBDL2	-0.5755	1.0000
PPP5C	-0.5620	1.0000
ATG7	-0.5490	1.0000
TMEM103	-0.5485	1.0000
SMOX	-0.5188	1.0000
PRPF19	-0.5109	1.0000
MFGE8	-0.5017	1.0000
RGN	0.5899	0.0000
CACNA1G	0.6255	0.0000
PCDH8	0.6623	0.0000
GSTO1	0.6721	0.0000
ARG1	0.6762	0.0000

Figure S1 Quantile-Quantile plot of the probability distribution from bootstrap test against the quantiles from the uniform distribution showed departures at both ends of the distribution. The red line represents the expected relationship when the null hypothesis, no correlation between K and nucleosome score, is true.
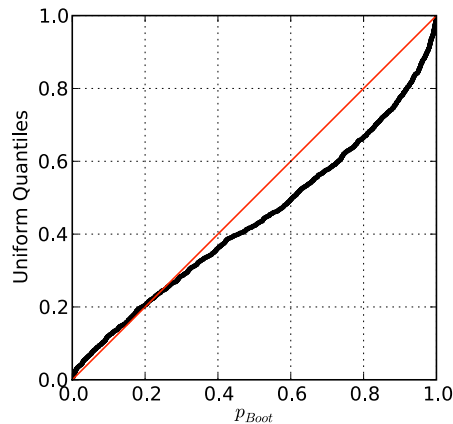
## Effect of Cramèr-Rao bound cutoff on estimated periodicity

We compared period distributions between the substitution rate (K) and nucleosome score signals. These were generally similar in terms of both the number of promoters and in particular the distribution of periods (Fig. S2). This observation held for both main and secondary periods, regardless of the CRB (Cramèr-Rao bound). As expected, the number of main and secondary periods observed decreased as the CRB threshold was reduced. In particular, fewer long periods were observed, as implied by equation (3). The attenuation in the number of significant periods in K for CRB thresholds 0.1 and 0.05 is expected given the increasing stringency of the period estimate variance.
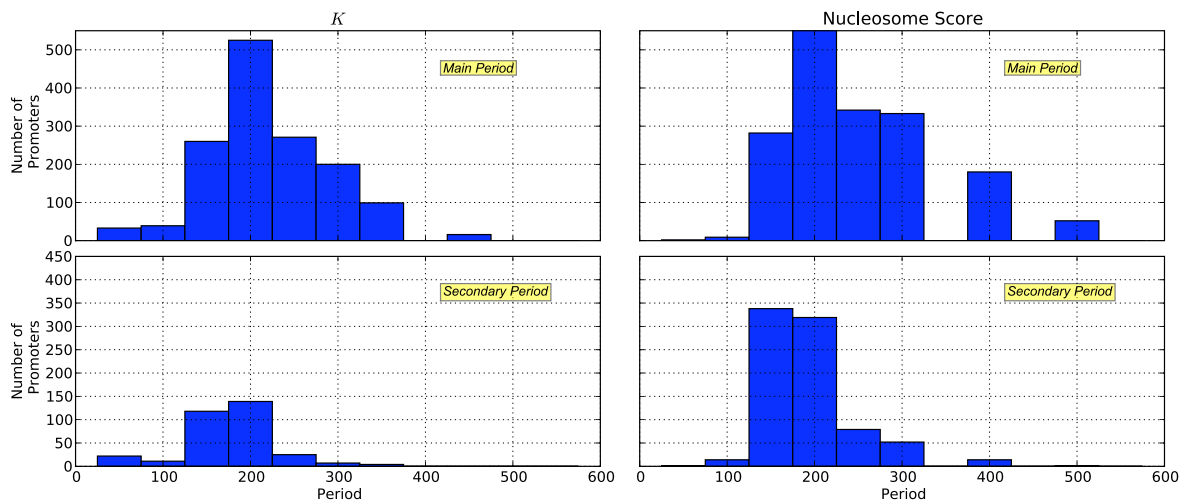
References

[1] Adam Siepel, Gill Bejerano, Jakob S. Pedersen, Angie S. Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, John Spieth, LaDeana W. Hillier, Stephen Richards, George M. Weinstock, Richard K. Wilson, Richard A. Gibbs, W. James Kent, Webb Miller, and David Haussler. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8):1034–1050, 2005.

Figure S2 Period distributions were consistent between K and nucleosome score signals determined under different CRB threshold.

(a) CRB threshold = 0.5
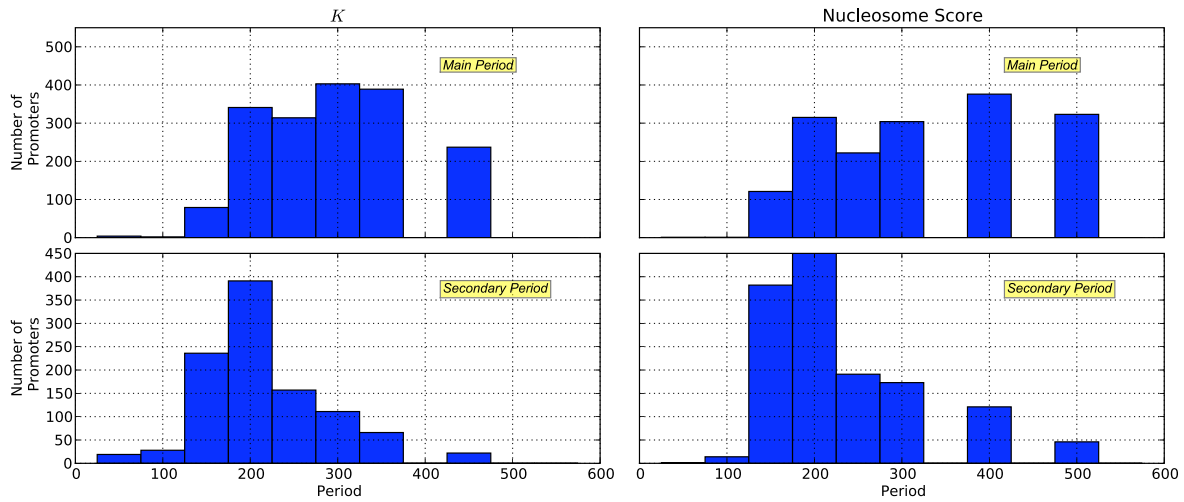
(b) CRB threshold = 0.1

(c) CRB threshold = 0.05