Stem Cell Reports, Volume *16*

Supplemental Information

Species-Specific Relationships between DNA and Chromatin Properties of CpG Islands in Embryonic Stem Cells and Differentiated Cells

Justin Langerman, David Lopez, Matteo Pellegrini, and Stephen T. Smale

Supplemental Figures for:

Species-Specific Relationships between the DNA and Chromatin Properties of CpG Islands in Embryonic Stem Cells and Differentiated Cells

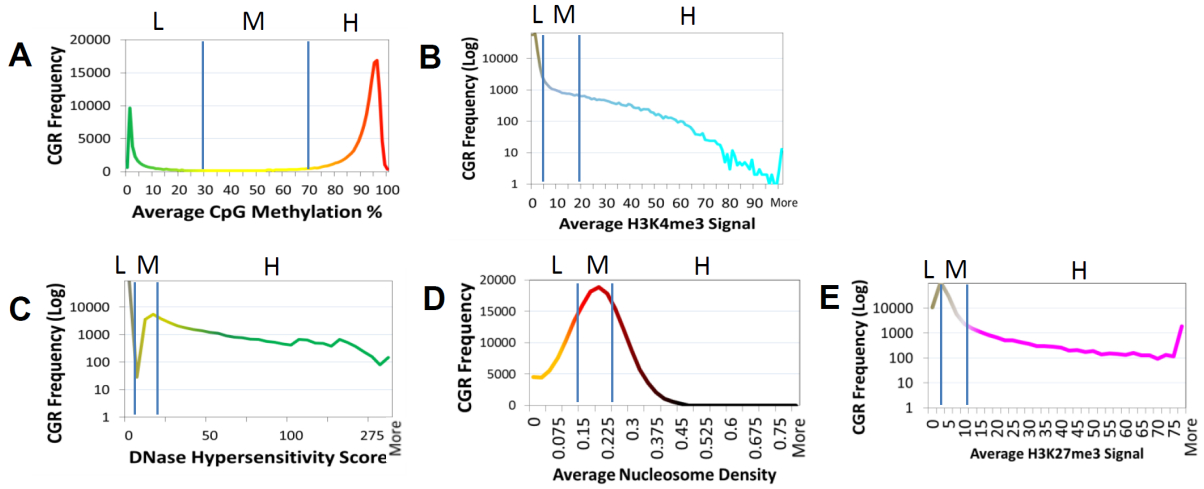

Figure S1. Distributions of chromatin properties at human ESC CGRs. Related to Figure 2.
Shown are graphs of the chromatin scores for all CGRs in human ESC. Lines on each graph demarcate the ranges used for low, medium, and high chromatin bins in subsequent figures, which are labeled with an L, M, or H respectively. Each line shows the count of CGRs at each integer value for chromatin scores pertaining to (A) average DNA methylation across the CGR, (B) average H3K4me3 ChIP-Seq signal over the CGR, (C) peak DNase HS score, (D) average MNase-seq nucleosome density, or (E) average H3K27me3 ChIP-Seq signal, in human ESC.

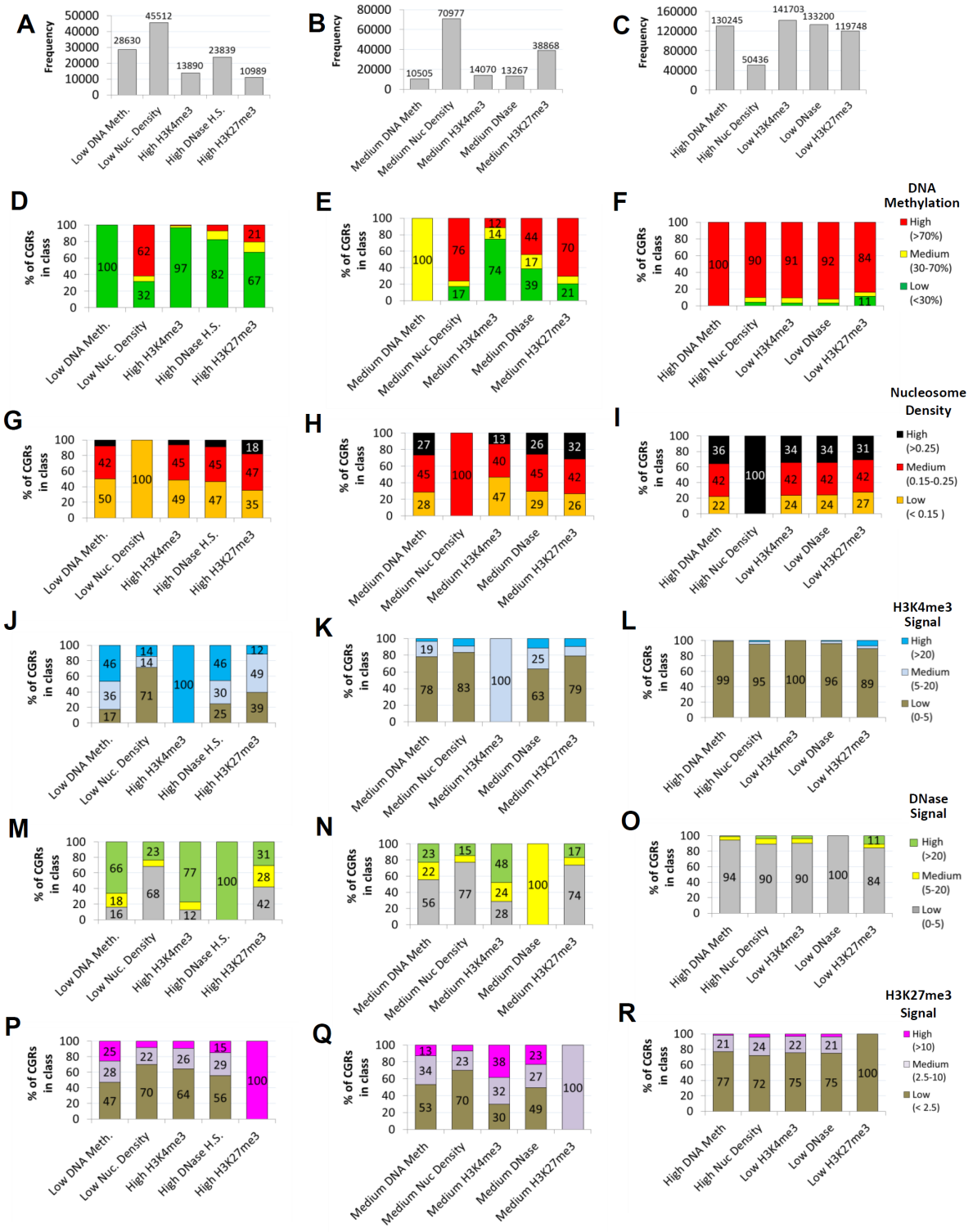

Figure S2. Intercorrelation of chromatin features at human ESC CGRs. Related to Figure 2.
To examine relationships between the chromatin features, co-segregation analyses were performed after assigning each human ESC CGR to one of three bins for each feature. The three bins for each chromatin feature are displayed in Figure S1. Bins for each chromatin feature were defined somewhat arbitrarily, but with the goal of creating three bins of approximately equal size, while looking for natural cutoffs based on the distributions shown in Figure S1.

(A-C) The numbers of CGRs (173,307 total CGRs) within each low, medium, or high bin for each chromatin feature are shown (see L, M, and H in Figure S1).
(D-R) The co-segregation analysis of chromatin features is shown. This analysis revealed a high degree of co-segregation for CGRs with low CpG methylation, high or medium H3K4me3, and high or medium DNase HS. For example, 97% of CGRs with high H3K4me3 exhibited low DNA methylation (panel D, bar 3; n=13,890 CGRs with high H3K4me3, as shown in panel A. In contrast, 91% of CGRs with low H3K4me3 exhibited high DNA methylation (panel F, bar 3; n=141,703, as shown in panel C) .
Strong correlations between CpG methylation, H3K4me3, and DNase HS were also observed when focusing attention on H3K4me3 levels (J-L) or DNase HS levels (M-O). Interestingly, nucleosome density displayed a relatively weak correlation to these three chromatin features. Although 90% of CGRs with high nucleosome density exhibited high DNA methylation (F), only 32% of CpGs with low nucleosome density exhibited low DNA methylation (D). Similarly, although 90% of CGRs with high nucleosome density exhibited low DNase HS (O), only 23% of CGRs with low nucleosome density exhibited high DNase HS (M). To be sure, CGRs with low DNA methylation, high H3K4me3, and high DNase HS exhibited moderately reduced nucleosome density distributions in comparison to CGRs with the opposite features (compare G to I). However, this relatively weak correlation between nucleosome density and the other chromatin features can be accounted for by differences in genome location. Importantly, the weak correlation between low nucleosome density and the other chromatin features could not be strengthened by using different bin thresholds (data not shown). Histone H3K27me3 also co-segregated only weakly with the other four chromatin properties (D-R), consistent with current knowledge that H3K27me3 is deposited at only a subset of inactive islands. These weak correlations are most readily apparent in panels P, Q, and R, which show that most CGRs exhibit low H3K27me3 levels, regardless of the other chromatin properties observed at the CGR.

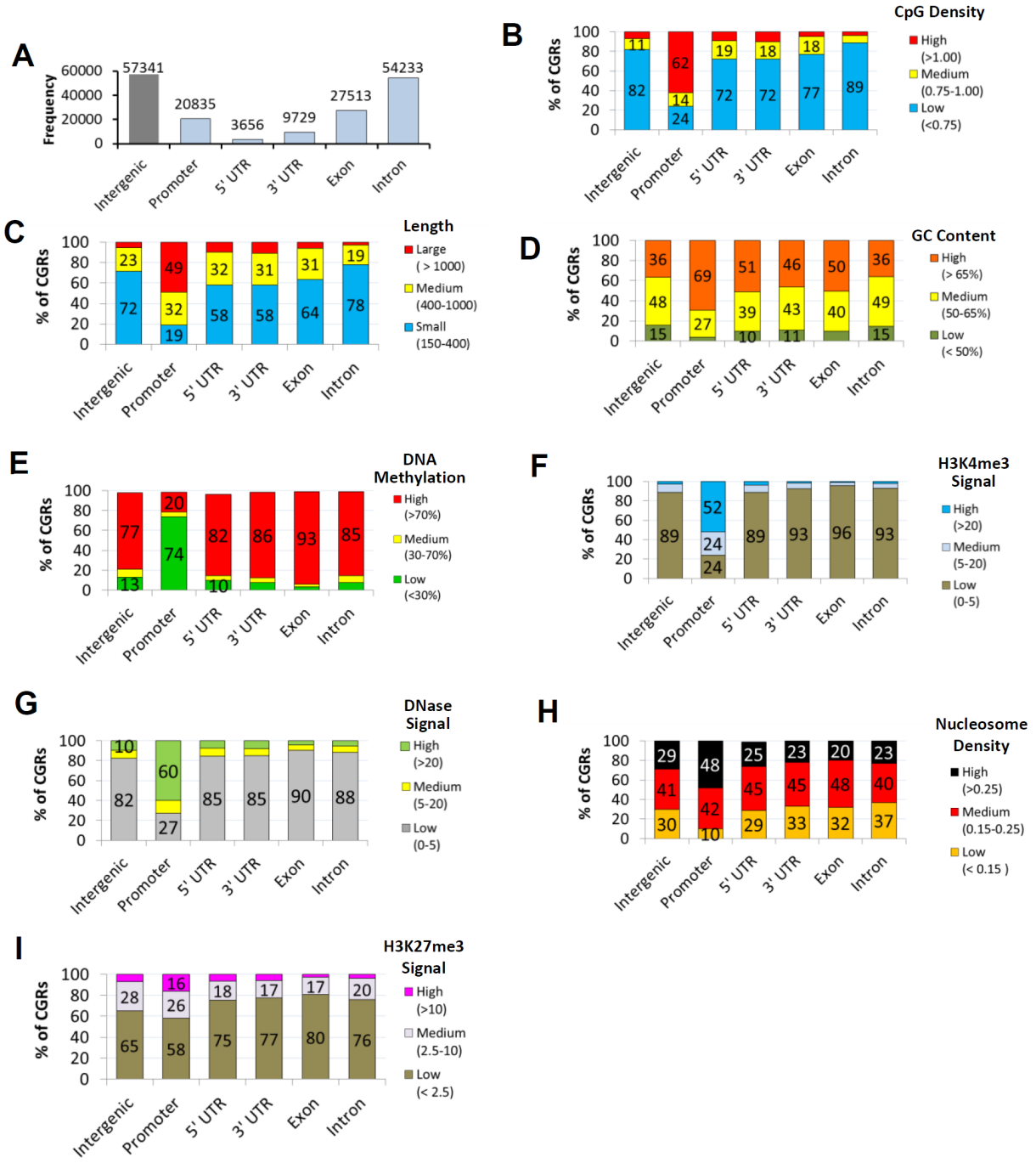

Figure S3. Features of human ESC CGRs are influenced by genomic location. Related to Figure 2.
The interrelationship between DNA properties and chromatin properties is shown. Importantly, we found that this analysis benefited greatly from simultaneous consideration of the genomic locations of the CGRs.
(A) 12% of all CGRs are located in promoter regions, with the remaining 88% distributed among 5' and 3' untranslated regions (UTRs), exons, introns, and intergenic regions. Multiple location matches are superseded in the following order: Promoter, UTR, Exon, and then Intron.
(B) After assigning CGRs to bins on the basis of defined DNA properties, we found that promoter CGRs exhibit a higher CpG density distribution than non-promoter CGRs.
(C-D) Promoter CGRs are also generally longer and possess higher GC percentages than non-promoter CGRs.

(E-H) Promoter CGRs often exhibited lower DNA methylation levels, higher H3K4me3, higher DNase HS, and higher nucleosome densities in human ESC than non-promoter CGRs.
(I) A slightly larger fraction of promoter CGRs exhibited higher H3K27me3 levels than non-promoter CGRs in human ESC.

	CpG Density (obs/ex)	CGR Length (bp)	Human ESC (Naive) # Meth	Human ESC (Naive) Total	Human ESC (Naive) % Meth	Human ESC (Primed) # Meth	Human ESC (Primed) Total	Human ESC (Primed) % Meth
Promoters	>1.0	>1,000	38	8079	0.5	151	8079	1.9
	0.8-1.0	>1,000	40	1773	2.3	153	1773	8.6
	>1.0	600-1,000	16	2441	0.7	81	2441	3.3
	0.8-1.0	600-1,000	46	1477	3.1	169	1480	11.4
Exons	>1.0	>1,000	109	1059	10.3	421	1059	39.8
	0.8-1.0	>1,000	221	979	22.6	700	979	71.5
	>1.0	600-1,000	65	550	11.8	238	550	43.3
	0.8-1.0	600-1,000	289	1166	24.8	858	1166	73.6
Introns	>1.0	>1,000	45	399	11.3	122	399	30.6
	0.8-1.0	>1,000	82	425	19.3	257	425	60.5
	>1.0	600-1,000	42	334	12.6	120	334	35.9
	0.8-1.0	600-1,000	115	584	19.7	327	584	56.0
Intergenic	>1.0	>1,000	75	1371	5.5	301	1371	22.0
	0.8-1.0	>1,000	109	890	12.2	358	890	40.2
	>1.0	600-1,000	47	901	5.2	198	901	22.0
	0.8-1.0	600-1,000	150	1271	11.8	517	1271	40.7

Figure S4. **Methylation of human CGRs differs dramatically between the naïve and primed ESC states even at strict criteria. Related to Figure 3.**
As in Figure 3, the tables show the counts of CGRs in human DNA methylation data sets that meet the DNA property criteria labeled at left. CGRs are separated by genomic location and by non-overlapping CpG density and length ranges, for human ESC naïve and primed condition methylomes (Guo et al., 2017). The number of CGRs with high DNA methylation (>70%) in each criteria range is shown under "# Meth" next to the number of regions that qualify for the criteria, "Total". The last column for each methylome group is the percent of CGRs methylated for each criterion, "% Meth". The tables are colored by frequency; increased grey indicates higher CGR numbers, while increased red indicates a higher percentage in the "% Meth" column. CGRs with insufficient bisulfite sequencing reads were discarded.