SUPPLEMENTAL MATERIAL FOR

GC skew at the 5' and 3' Ends of Human Genes Links R-loop Formation to Epigenetic Regulation and Transcription Termination

Paul A. Ginno, Yoong Wearn Lim, Paul L. Lott, Ian Korf,

and Frédéric Chédin

Supplementary Table 1: Gene Ontologies associated with promoter GC skew classes.

Class I Strong skew genes

Category	**Term**	**Count**	**PValue**	**Benjamini**
GOTERM_BP_ALL	GO:0044260~cellular macromolecule metabolic process	2336	2.14E-68	1.23E-64
GOTERM_BP_ALL	GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1532	6.54E-39	5.37E-36
GOTERM_BP_ALL	GO:0019222~regulation of metabolic process	1592	2.92E-34	1.87E-31
GOTERM_BP_ALL	GO:0010467~gene expression	1349	1.58E-33	7.59E-31
GOTERM_BP_FAT	GO:0006350~transcription	966	3.88E-26	2.18E-22
GOTERM_BP_FAT	GO:0045449~regulation of transcription	1156	9.26E-25	2.60E-21
GOTERM_BP_ALL	GO:0006464~protein modification process	690	3.25E-23	6.44E-21
GOTERM_BP_ALL	GO:0006996~organelle organization	626	9.74E-20	1.81E-17
GOTERM_BP_FAT	GO:0016568~chromatin modification	165	2.53E-16	4.15E-13
GOTERM_BP_ALL	GO:0007399~nervous system development	501	1.42E-13	2.20E-11
GOTERM_BP_FAT	GO:0044265~cellular macromolecule catabolic process	354	1.51E-13	2.12E-10
GOTERM_BP_FAT	GO:0030163~protein catabolic process	310	2.37E-13	2.66E-10
GOTERM_BP_FAT	GO:0019941~modification-dependent protein catabolic process	289	3.59E-13	3.36E-10
GOTERM_BP_FAT	GO:0007049~cell cycle	371	1.51E-12	9.40E-10

Class II Weak skew genes

Category	**Term**	**Count**	**PValue**	**Benjamini**
GOTERM_BP_ALL	GO:0044237~cellular metabolic process	1733	3.42E-12	8.72E-09
GOTERM_BP_ALL	GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	931	1.76E-09	1.79E-06
GOTERM_BP_ALL	GO:0010467~gene expression	810	2.86E-07	1.83E-04
GOTERM_MF_ALL	GO:0016651~oxidoreductase activity, acting on NADH or NADPH	37	5.09E-05	2.42E-02
GOTERM_BP_FAT	GO:0045184~establishment of protein localization	231	1.62E-05	0.03952185
GOTERM_BP_FAT	GO:0015031~protein transport	228	2.44E-05	0.0396427
GOTERM_BP_FAT	GO:0008104~protein localization	261	1.45E-05	0.06970837
GOTERM_BP_FAT	GO:0045333~cellular respiration	41	7.73E-05	0.0915451
GOTERM_BP_FAT	GO:0055114~oxidation reduction	189	2.51E-04	0.18779336
GOTERM_BP_FAT	GO:0006091~generation of precursor metabolites and energy	102	2.22E-04	0.19839695
GOTERM_BP_FAT	GO:0043547~positive regulation of GTPase activity	14	3.65E-04	0.20308996
GOTERM_BP_FAT	GO:0034660~ncRNA metabolic process	78	3.48E-04	0.219067

GOTERM_BP_FAT	GO:0006289~nucleotide-excision repair	25	7.71E-04	0.34664583
GOTERM_BP_FAT	GO:0006396~RNA processing	161	9.42E-04	0.37399808

Class III No skew genes

Category	**Term**	**Count**	**PValue**	**Benjamini**
GOTERM_BP_FAT	GO:0006952~defense response	345	6.12E-55	3.14E-51
GOTERM_BP_FAT	GO:0006955~immune response	363	7.78E-49	2.00E-45
GOTERM_BP_FAT	GO:0006954~inflammatory response	190	2.26E-33	3.87E-30
GOTERM_BP_FAT	GO:0009611~response to wounding	263	7.32E-30	9.39E-27
GOTERM_BP_FAT	GO:0022610~biological adhesion	298	3.07E-20	3.15E-17
GOTERM_BP_FAT	GO:0007155~cell adhesion	297	5.01E-20	4.28E-17
GOTERM_BP_FAT	GO:0042742~defense response to bacterium	74	1.62E-17	1.19E-14
GOTERM_BP_FAT	GO:0045087~innate immune response	85	3.04E-17	1.95E-14
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	124	2.58E-16	1.27E-13
GOTERM_BP_FAT	GO:0006811~ion transport	307	5.93E-16	2.85E-13

Class IV Reverse skew genes

Category	**Term**	**Count**	**PValue**	**Benjamini**
GOTERM_MF_ALL	GO:0003700~transcription factor activity	169	6.47E-15	8.07E-12
GOTERM_MF_ALL	GO:0030528~transcription regulator activity	232	4.02E-14	2.52E-11
GOTERM_MF_ALL	GO:0043565~sequence-specific DNA binding	117	1.29E-13	5.38E-11
GOTERM_MF_ALL	GO:0003677~DNA binding	295	6.47E-08	1.62E-05
GOTERM_BP_FAT	GO:0006357~regulation of transcription from RNA polymerase II promoter	117	2.36E-08	8.47E-05
GOTERM_BP_FAT	GO:0048598~embryonic morphogenesis	61	7.23E-08	1.30E-04
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	238	1.91E-07	2.29E-04
GOTERM_BP_FAT	GO:0045165~cell fate commitment	34	7.83E-07	2.81E-04
GOTERM_BP_FAT	GO:0045449~regulation of transcription	319	7.59E-07	3.03E-04
GOTERM_BP_FAT	GO:0045892~negative regulation of transcription, DNA-dependent	65	6.32E-07	3.24E-04
GOTERM_BP_FAT	GO:0051254~positive regulation of RNA metabolic process	81	7.28E-07	3.27E-04
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	231	5.58E-07	3.34E-04
GOTERM_BP_FAT	GO:0051253~negative regulation of RNA metabolic process	66	5.42E-07	3.89E-04
GOTERM_BP_FAT	GO:0045893~positive regulation of transcription, DNA-dependent	81	5.14E-07	4.61E-04

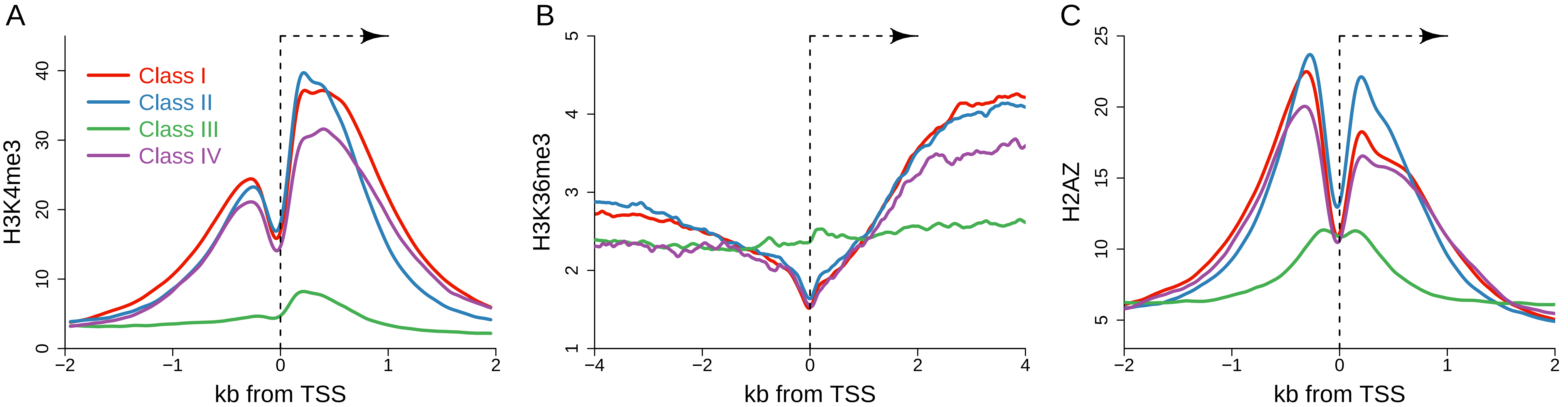
Supplementary Table 2: numbers of unique genes and unique TSSs in each GC skew class

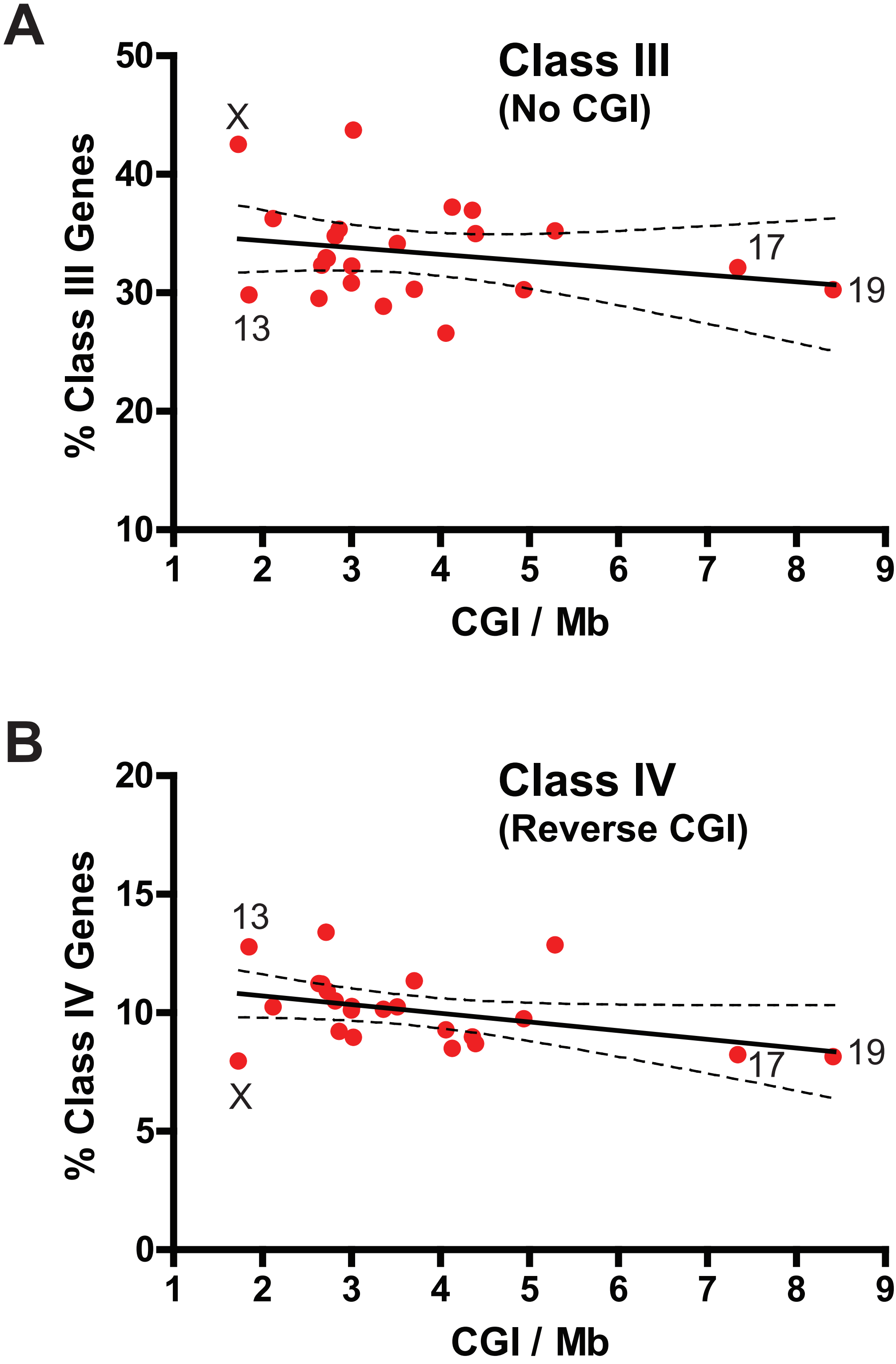
Class	# of Genes	# of TSS
I	7152	8332
II	4997	5799
III	6775	7968
IV	1918	2099

Legends for supplementary figures.

Supplementary Figure 1: **Promoter classes present distinct epigenetic signatures in hESCs.** Histone modification metaplots for each promoter class for H3K4me3 (A), H3K36me3 (B) and H2Az (C). Class-specific color codes are all identical and indicated in panel A. The y-axes represent arbitrary units.

Supplementary Figure 2. **Gene density does not affect the distribution of Class III and Class IV genes**. The distribution of Class III (A) and Class IV (B) genes on individual chromosomes is represented as a percentage of total RefSeq genes on that chromosome (y-axis) plotted against a measure of gene density. A few relevant chromosomes are indicated. The data was fit to a linear regression shown here with the corresponding 95% confidence interval.


A
B
C
H3K4me3
H3K36me3
H2AZ
Class I
Class II
Class III
Class IV
0
10
20
30
40
1
2
3
4
5
25
20
15
10
5
−2
−1
0
1
2
−4
−2
0
2
4
kb from TSS
kb from TSS
kb from TSS


A
Class III
(No CGI)
% Class III Genes
50
40
30
20
10
1
2
3
4
5
6
7
8
9
CGI / Mb
X
13
17
19
B
Class IV
(Reverse CGI)
% Class IV Genes
20
15
10
5
0
1
2
3
4
5
6
7
8
9
CGI / Mb
13
X
17
19