Supplementary Materials for

Systematic identification and annotation of human methylation marks based on bisulfite sequencing methylomes reveals distinct roles of cell type-specific hypomethylation in the regulation of cell identity genes

Hongbo Liu[1], Xiaojuan Liu[2], Shumei Zhang[1], Jie Lv[3], Song Li[1], Shipeng Shang[1], Shanshan Jia[1], Yanjun Wei[1], Fang Wang[1], Jianzhong Su[1], Qiong Wu[3], Yan Zhang[1]

[1] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China
[2] Department of Rehabilitation, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, China
[3] School of Life Science and Technology, State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150001, China

Overview of Supplementary Materials:

Supplementary Tables

Supplementary Figures

methylation.
Supplementary Figure 9: Features of LowSpe segments.
Supplementary Figure 10: UniHypo segments and ubiquitous enhancers.
Supplementary Figure 11: The methylation pattern and chromatin states of the promoter regions of CTCF.
Supplementary Figure 12: HighSpe segments identified by SMART in the human genome.
Supplementary Figure 13: K-means clustering of HighSpe segments.
Supplementary Figure 14: Genome location of different clusters of HighSpe segments.
Supplementary Figure 15: The enriched functions of genes related to HighSpe segments in each cluster.
Supplementary Figure 16: K-means clustering of InterSpe segments.
Supplementary Figure 17: K-means clustering of MethyMarks.
Supplementary Figure 18: Genome location and functional enrichment of cell type-specific MethyMarks.
Supplementary Figure 19: Cell type-specific MethyMarks that span large chromosomal regions.
Supplementary Figure 20: Example genes related to cell type-specific MethyMarks that span large chromosomal regions.
Supplementary Figure 21: Detailed epigenetic modifications in the iPSC-specific HyperMarks related to IRX2 across cell types.
Supplementary Figure 22: Heatmap of DNA methylation and H3K27ac in cell type-specific HyperMarks.
Supplementary Figure 23: Correlation between DNA methylation and H3K27ac in HypoMarks and HyperMarks of various cell types.
Supplementary Figure 24: Correlation between DNA methylation and two histone marks (H3K27ac and H3K27me3) in different categories of segments.
Supplementary Figure 25: hESC H1-specific HypoMarks and HyperMarks show distinct chromatin modifications.
Supplementary Figure 26: Sub-network of transcription factor-MethyMark collaboration network in hESC H1 cells as shown in Figure 4D.
Supplementary Figure 27: DNA methylation and H3K27ac state of cell type-specific SuperHypoMarks across 21 cell types.
Supplementary Figure 28: Detailed information about epigenetics and expression of H1-specific SuperHypoMark genes.

Supplementary Tables

Supplementary Table 1

Data type	source	Links	Reference
DNA methylation	the Human Epigenome Atlas	http://www.genboree.org/epigenomeatlas/	(1)
Histone modification	the Human Epigenome Atlas	http://www.genboree.org/epigenomeatlas/	(1)
Transcriptome	the Human Epigenome Atlas	http://www.genboree.org/epigenomeatlas/	(1)
Refseq genes	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
GENCODE gene	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
CpG islands	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
Repeats	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
Chromatin states	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
TFBSs	UCSC Table Browser	http://genome.ucsc.edu/cgi-bin/hgTables	(2)
Housekeeping genes	Eisenberg et al. Trends in Genetics 29, (2013).	http://www.tau.ac.il/~elieis/HKG/	(3)
Imprinted genes	MetaImprint database	http://bioinfo.hrbmu.edu.cn/MetaImprint	(4)
Enhancers	FANTOM Transcribed Enhancer Atlas	http://enhancer.binf.ku.dk/	(5)
Super-enhancers	Hnisz et al. Cell 155, (2013)	http://www.ncbi.nlm.nih.gov/pubmed/24119843	(6)

Supplementary Table 1: List of all datasets used in this study.

Supplementary Table 2

Cell type	CG coverage	Centre	GEO Accession	DevelopClass
H1	92.31%	UCSD	GSM429321	hESC
H9	95.63%	UCSD	GSM706059	hESC
HUES64	93.25%	BI	GSM1112840	hESC
UCSF-4* ESC	93.18%	UCSF-UBC	GSM1127122	hESC
H1-BMP4	95.62%	UCSD	GSM602251	hESC BMP4
H1-BMP4 Mesendoderm	94.62%	UCSD	GSM818003	hESC Derived Cells
H1 Mesenchymal SCs	89.73%	UCSD	GSM864032	hESC Derived Cells
H1 Derived NPCs	95.98%	UCSD	GSM675542	hESC Derived Cells
CD184+ Endoderm 1	94.89%	BI	GSM1112848	hESC Derived Cells
CD184+ Endoderm 2	96.01%	BI	GSM916051	hESC Derived Cells
CD56+ Ectoderm 1	94.90%	BI	GSM1112821	hESC Derived Cells
CD56+ Ectoderm 2	91.90%	BI	GSM1112849	hESC Derived Cells
CD56+ Mesoderm 1	93.69%	BI	GSM1112839	hESC Derived Cells
CD56+ Mesoderm 2	93.92%	BI	GSM1112842	hESC Derived Cells
iPS DF 19.11	95.98%	UCSD	GSM706053	iPSC
iPS DF 6.9	95.59%	UCSD	GSM706057	iPSC
IMR90	94.14%	UCSD	GSM432687	Cell Line
Neurosphere Cortex 1	95.23%	UCSF-UBC	GSM1127118	Neurosphere Cultured Cells
Neurosphere Cortex 2	94.05%	UCSF-UBC	GSM1127124	Neurosphere Cultured Cells
Neurosphere GanEmi 1	86.34%	UCSF-UBC	GSM1127055	Neurosphere Cultured Cells
Neurosphere GanEmi 2	94.71%	UCSF-UBC	GSM1127121	Neurosphere Cultured Cells
CD34 Primary Cells	94.79%	BI	GSM916052	Primary Cells
Keratinocyte Cells	93.53%	UCSF-UBC	GSM1127056	Primary Cells
Spermatozoa Cells	94.60%	UCSF-UBC	GSM1127117	Primary Cells
Breast Luminal Epithelial	94.31%	UCSF-UBC	GSM1127125	Primary Cells Breast
Breast Myoepithelial Cells	75.39%	UCSF-UBC	GSM1127054	Primary Cells Breast
Adipose Tissue	93.02%	UCSD	GSM1010983	Primary Tissue
Adrenal Gland	95.41%	UCSD	GSM1010981	Primary Tissue
Adult Liver	91.94%	BI	GSM916049	Primary Tissue
Aorta	96.34%	UCSD	GSM983648	Primary Tissue
Esophagus	96.11%	UCSD	GSM983649	Primary Tissue
Fetal Muscle Leg	90.73%	BI	GSM1172596	Primary Tissue
Fetal Thymus	92.41%	BI	GSM1172595	Primary Tissue
Gastric	95.78%	UCSD	GSM1010984	Primary Tissue
Left Ventricle 1	95.95%	UCSD	GSM1010978	Primary Tissue
Left Ventricle 2	96.29%	UCSD	GSM983650	Primary Tissue
Lung	96.10%	UCSD	GSM983647	Primary Tissue
Ovary	95.89%	UCSD	GSM1010980	Primary Tissue
Pancreas	96.29%	UCSD	GSM983651	Primary Tissue
Psoas Muscle	96.15%	UCSD	GSM1010986	Primary Tissue
Right Atrium	95.81%	UCSD	GSM1010987	Primary Tissue
Right Ventricle	95.58%	UCSD	GSM1010988	Primary Tissue
Sigmoid Colon 1	96.34%	UCSD	GSM1010989	Primary Tissue
Sigmoid Colon 2	96.59%	UCSD	GSM983645	Primary Tissue
Small Intestine	96.52%	UCSD	GSM983646	Primary Tissue
Spleen	96.34%	UCSD	GSM983652	Primary Tissue
Thymus	95.90%	UCSD	GSM1010979	Primary Tissue
Brain Germinal Matrix	85.67%	UCSF-UBC	GSM941747	Primary Tissue Brain
Brain Hippocampus Middle	92.77%	BI	GSM1112838	Primary Tissue Brain
Brain Hippocampus Middle	93.96%	BI	GSM916050	Primary Tissue Brain

Supplementary Table 2: DNA methylation data sets. DNA methylome data were obtained from the Human Epigenome Atlas produced by the NIH Epigenomics Roadmap Consortium (http://www.genboree.org/epigenomeatlas/).

Supplementary Table 3

Chrome	Num.	Total Length	Mean Length	Max Length	Mean CpG Num.	Max CpG Num.	CpG Num.	CpG Coverage
chr1	66,706	47,531,847	713	12,942	11	726	737,480	50%
chr2	59,219	39,489,344	667	15,792	11	608	627,756	48%
chr3	42,607	27,578,682	647	8,713	10	422	444,921	47%
chr4	33,347	21,106,331	633	12,636	11	544	357,439	45%
chr5	37,702	24,653,332	654	9,846	11	480	400,058	46%
chr6	39,684	25,652,523	646	12,455	11	626	429,012	48%
chr7	40,694	29,374,921	722	23,155	11	1,148	464,416	48%
chr8	33,871	23,069,599	681	25,104	11	632	360,061	46%
chr9	31,650	22,635,550	715	11,089	11	584	357,717	49%
chr10	39,269	27,839,976	709	27,357	11	898	422,365	48%
chr11	36,592	26,429,176	722	10,934	11	564	405,487	49%
chr12	35,888	25,726,933	717	21,843	11	1,066	408,537	50%
chr13	20,023	12,515,790	625	7,625	11	508	211,095	47%
chr14	23,979	16,797,827	701	9,113	11	446	266,403	49%
chr15	25,697	17,754,807	691	9,002	11	574	275,843	51%
chr16	31,724	25,566,930	806	16,474	12	706	386,926	51%
chr17	36,652	28,988,347	791	13,696	12	554	455,460	55%
chr18	17,811	11,715,884	658	9,303	10	548	186,993	46%
chr19	31,176	28,105,707	902	13,570	15	476	453,146	57%
chr20	22,479	17,766,532	790	12,152	11	396	253,910	49%
chr21	10,311	7,519,513	729	10,983	12	590	119,970	48%
chr22	18,376	15,558,444	847	14,481	12	626	228,204	53%
chrX	20,911	13,217,328	632	6,023	9	336	179,053	35%
chrY	1,519	1,502,482	989	22,704	15	526	22,586	51%
Total	757,887	538,097,805	710	27,357	11	1,148	8,454,838	49%

Supplementary Table 3: Length and CpG Number of the Segments identified in this study. Genome segmentation using 50 human DNA methylomes via the SMART algorithm identified 757,887 segments with more than 5 CpG sites and a length of longer than 20 bp. The length and CpG coverage of these segments in different chromosomes were calculated. In total, all of the identified segments covered ~538 million bp (mean length 710 bp) and ~8.5 million CpGs that consist of nearly 50% of all of the CpGs in the human genome.

Supplementary Table 4

HighSpe	Total	CGI	CGI shore	CGI desert	Up2kb	5'UTR	Coding Exon	Intron	3'UTR	Down 2kb	Refseq Gene [a]	Coding Gene [a]	ncRNA [a]
Total	95891	2713	15904	77274	6520	8760	3405	26556	3634	2195	51070 (16493)	44395 (13611)	6675 (2882)
Cluster1	9135	649	1561	6925	709	1087	361	2655	358	244	5414 (3805)	4680 (3208)	734 (597)
Cluster2	5804	46	299	5459	104	524	143	2133	257	94	3255 (2150)	2879 (1852)	376 (298)
Cluster3	4156	722	984	2450	411	516	225	1076	197	182	2607 (2162)	2260 (1867)	347 (295)
Cluster4	25281	91	1563	23627	515	2464	945	9792	1118	678	15512 (6967)	13877 (5996)	1635 (971)
Cluster5	2544	156	744	1644	221	320	102	809	140	93	1685 (1350)	1449 (1154)	236 (196)
Cluster6	33323	706	6314	26303	2837	2383	1068	6210	982	502	13982 (8578)	11776 (7158)	2206 (1420)
Cluster7	9972	295	3903	5774	1496	810	350	1764	333	255	5008 (4155)	4259 (3516)	749 (639)
Cluster8	5676	48	536	5092	227	656	211	2117	249	147	3607 (2504)	3215 (2176)	392 (328)

Supplementary Table 4: Number of HighSpe segments in different genome features and clusters. For each HighSpe segment, we determined its genome location relative to twelve genome features, including CpG island (CGI), CGI shore, CGI desert, Up2kb, 5'UTR, coding exon, intron, 3'UTR, Down2kb, Refseq gene, coding gene and ncRNA. The number in parentheses represents the number of genes related to the corresponding segments.

Supplementary Table 5

Cell-type	MethyMark	HypoMark		HyperMark	
	Num	Num	Gene	Num	Gene
Sigmoid_Colon	1,000	811	509	189	152
Small_Intestine	1,721	1,625	970	96	77
Lung	2,586	2,351	1,522	235	205
hESC_dCD56_Mesoderm	2,844	432	220	2,412	1,583
hESC_dCD184_Endoderm	2,930	371	175	2,559	1,645
hESC_H1_BMP4_dMesendoderm	3,335	1,343	761	1,992	1,306
hESC_dCD56_Ectoderm	3,429	1,207	718	2,222	1,451
hESC_UCSF4star	3,756	1,567	900	2,189	1,465
hESC_H1	3,758	1,241	662	2,517	1,618
iPSC_DF19.11	3,835	1,573	861	2,262	1,461
Left_Ventricle	3,881	3,834	2,594	47	37
hESC_HUES64	3,890	1,420	753	2,470	1,528
iPSC_DF6.9	4,005	1,577	888	2,428	1,535
hESC_H1_dNPCs	4,181	1,819	1,076	2,362	1,556
hESC_H9	4,266	1,436	753	2,830	1,878
hESC_H1_BMP4	4,830	2,586	1,400	2,244	1,462
Adipose_Tissue	4,832	4,612	2,986	220	180
Neurosphere_GanEmi	5,454	4,666	2,637	788	549
Adrenal_Gland	5,482	5,020	3,380	462	342
Right_Ventricle	5,620	5,542	3,599	78	60
hESC_H1_dMesenchymal_Stem_Cells	6,694	4,894	2,765	1,800	1,177
Right_Atrium	6,964	6,909	4,408	55	33
Neurosphere_Cortex	7,036	6,444	3,500	592	407
Spleen	8,033	7,314	4,736	719	551
Fetal_Thymus	8,290	6,795	4,321	1,495	1,142
Esophagus	8,609	8,353	5,141	256	167
Thymus	9,090	6,820	4,357	2,270	1,618
Liver	9,298	8,224	5,490	1,074	832
Mobilized_CD34_primary_Cells	9,814	7,766	4,968	2,048	1,457
Hippocampus_Middle	10,141	9,795	7,100	346	227
Brain_Germinal_Matrix	10,637	10,422	6,087	215	134
Gastric	11,761	11,584	7,254	177	134
Fetal_Muscle_Leg	14,781	14,576	8,153	205	166
Pancreas	16,254	15,416	8,614	838	619
Aorta	17,006	16,294	10,500	712	495
Ovary	18,780	18,363	10,892	417	311
Psoas_Muscle	21,241	20,766	12,925	475	344
Breast_Myoepithelial	27,334	24,581	15,512	2,753	1,680
Breast_Luminal_Epithelial	29,386	27,356	15,824	2,030	1,296
Keratinocyte_Cells	29,570	28,835	16,919	735	464
IMR90	35,703	35,276	14,396	427	311
Testis_Spermatozoa_Primary_Cells	68,381	61,569	27,117	6,812	3,754

Supplementary Table 5: Number of cell type-specific MethyMarks and related genes in different human cell types. For each cell type, the number of MethyMarks, HypoMarks and HyperMarks are shown. The number of genes related to HypoMarks and HyperMarks are also listed. Detailed information of these cell type-specific MethyMarks is available at http://fame.edbc.org/methymark.

Supplementary Table 7

Cell-type	SE2Hypo Num	SE2Hyper Num	Hypo2SE Num	Hyper2SE Num	HypoMark Num	HyperMark Num	Hypo2SE p value	Hypo2SE odds	Hypo VS Hyper p value	Hypo VS Hyper odds	SuperHypoMark Num
hESC_H1	145	20	204	25	1241	2517	1.33E-245	19.7173	3.15E-13	6.09659	175
Hippocampus_Middle	665	14	2040	17	9795	346	1.07E-45	2.14982	5.88E-56	23.9806	830
Adrenal_Gland	387	25	815	31	5020	462	5.31E-23	1.93516	4.62E-29	8.5075	137
Esophagus	660	17	1570	18	8353	256	1.72E-17	1.57157	1.15E-58	20.9432	372
Ovary	365	10	1344	17	18363	417	2.97E-07	1.45342	1.20E-30	18.8941	225
Pancreas	404	43	1190	57	15416	838	4.45E-14	1.67444	3.04E-21	4.89458	188
Fetal_Muscle_Leg	591	13	1576	15	14576	205	1.87E-05	1.26487	1.90E-53	23.9976	365
Psoas_Muscle	437	11	2058	16	20766	475	1.13E-07	1.4409	8.64E-39	22.0509	268
Right_Atrium	499	8	1180	12	6909	55	1.39E-09	1.43209	2.31E-47	33.2874	36
Left_Ventricle	461	4	960	7	3834	47	2.46E-22	1.79163	1.03E-61	58.1489	82
Right_Ventricle	177	2	358	3	5542	78	5.20E-04	1.39882	3.59E-24	46.4121	4
Aorta	626	25	3167	34	16294	712	4.01E-16	1.59684	3.89E-43	11.9657	637
Gastric	751	13	2243	19	11584	177	2.50E-17	1.54353	1.30E-71	32.5926	364
Sigmoid_Colon	138	23	166	26	811	189	1.04E-08	1.73739	7.39E-07	3.07071	14
Small_Intestine	219	13	297	15	1625	96	4.07E-11	1.69066	7.55E-18	8.64961	17
Lung	347	11	523	13	2351	235	4.33E-06	1.34953	6.03E-34	17.4232	31
Adipose_Tissue	8	0	8	0	4612	220	4.83E-01	0.707689	0.020255	6.060606	0
Spleen	545	73	1221	90	7314	719	4.78E-10	1.4205	2.21E-24	3.80436	174
Mobilized_CD34_primary_Cells	169	12	360	16	7766	2048	1.78E-07	1.69939	1.09E-07	5.08402	104
Thymus	317	61	714	72	6820	2270	4.26E-19	1.94644	0.000182	1.8844	73
Fetal_Thymus	279	33	727	47	6795	1495	2.93E-19	2.04824	1.22E-07	2.99337	126

Supplementary Table 7: Number and Fraction of HypoMarks and HyperMarks overlap with super-enhancers.

Details for each column are as follows:

Cell-type: 21 cell types were used for super-enhancer analysis in this study.

SE2Hypo Num: the number of super-enhancers overlapped with HypoMarks from the same cell type.

SE2Hyper Num: the number of super-enhancers overlapped with HyperMarks from the same cell type.

Hypo2SE Num: the number of HypoMarks overlapped with super-enhancers from the same cell type.

Hyper2SE Num: the number of HyperMarks overlapped with super-enhancers from the same cell type.

HypoMark Num: the number of HypoMarks identified in the given cell type.

HyperMark Num: the number of HyperMarks identified in the given cell type.

Hypo2SE p value: significance of the Chi-square test for overlap between HypoMarks and super-enhancers from the same cell type compared to HypoMarks of other cell types.

Hypo2SE odds: the odds ratio of HypoMarks overlapped with super-enhancers of the same cell type compared to HypoMarks of other cell types

Hypo VS Hyper p value: significance of the Chi-square test for overlap between HypoMarks and super-enhancers from the same cell type compared to HyperMarks of the same cell type.

Hypo VS Hyper odds: the odds ratio of HypoMarks overlapped with super-enhancers of the same cell type compared to HyperMarks of the same cell type.

SuperHypoMark Num: the number of cell type-specific SuperHypoMarks that were HypoMarks only overlapped with super-enhancers from the same cell type.

Supplementary Figure Legends

Supplementary Figure 1: Distribution of DNA methylation in 50 methylomes. Each sub-graph represents the distribution of CpG methylation in specific human tissues or cell lines. The DNA methylomes showed a bimodal distribution in all cell types, and most CpGs were hypermethylated.

Supplementary Figure 2: Distribution of DNA methylation corrected by one-step Tukey biweight in 50 methylomes. (A) Each sub-graph represents the distribution of CpG methylation corrected by one-step Tukey biweight in specific human tissues or cell lines. To determine methylation specificity, one-step Tukey biweight was calculated as a robust weighted mean using the methylation levels in the majority of cell types after discounting the outliers in the minority of cell types by a weight that was calculated by the bisquare function. (B) The distribution of one-step Tukey biweights for all CpG sites.

Supplementary Figure 3: Evaluation of the performance of an entropy-based algorithm in the quantification of methylation specificity across multiple samples. To determine the thresholds for methylation specificity, we modelled different methylation patterns by random sampling from different normal distributions with different mean and standard deviation values and studied the distribution of methylation specificity. For a given mean methylation level (mean, ranging from 0.0 to 1.0) and a given standard deviation (SD, ranging from 0.0 to 0.5), 50 values were randomly sampled as the methylation levels in 50 samples of a CpG site. This process was repeated 10,000 times to produce 10,000 CpG sites whose methylation specificity across 50 samples was quantified by our method as described in the manuscript. Then, the distribution of these methylation specificity values was used to evaluate the accuracy of our method in the quantification of methylation specificity and determine the thresholds for classification of degree of methylation specificity. (A) Boxplot of methylation specificity for random data produced by different mean and SD values. (B) Distribution of methylation specificity for random data produced by different mean and SD values. (C) Distribution of methylation specificity calculated using a different number of samples. The similar distribution of methylation specificity suggested our method should be applicable to datasets with different sample numbers.

Supplementary Figure 4: Methylation specificity across known gene regulatory regions. (A) Composite plot of methylation specificity quantified by normalized Shannon entropy across known regulatory elements, including CpG islands, Refseq genes, long noncoding RNAs (lncRNA), ubiquitous enhancers, cell type-specific enhancers and super-enhancers. Blue lines indicate the median of the methylation specificity across each element, and grey areas mark the twenty-fifth and seventy-fifth percentiles of methylation specificity. (B) Methylation specificity quantified by normalized Shannon entropy and methylation segments identified by SMART near the developmental gene *POU5F1* (also known as *OCT4*).

Supplementary Figure 5: Determination of thresholds for merging neighboring CpGs. (A) Distribution of Euclidean distance of DNA methylation levels between neighboring CpGs in real and random datasets. (B) Distribution of similar entropy of DNA

methylation levels between neighboring CpGs in real and random datasets. (C) Distribution of distance between neighboring CpGs. (D-J) To determine the threshold of max distance between neighboring CpGs in merging two neighboring CpGs, we performed genome segmentation, setting the same other parameters, but the max distance between two CpGs was set as 250 bp and 500 bp. The comparison of the results from two thresholds is shown in the following figures. (D) Distribution of methylation specificity of segments. (E) Distribution of CpG number of segments. (F) Distribution of mean methylation level of segments. (G) Distribution of length of segments. (H) Number of segments identified by 250 bp and 500 bp. Approximately 13.7% of the segments identified by 500 bp were not overlapped with any segment identified by 250 bp, while only 4.2% of the segments identified by 250 bp were not overlapped by any segment identified by 500 bp. This result suggests that some of the segments identified by the threshold of 250 bp were not lost but rather merged into larger segments by the threshold of 500 bp. (I) Number of segments with a length > 3.5 kb, which was used to identify long hypomethylated genome regions by Jeong et al. It is suggested that the threshold of 500 bp can be used to identify those segments not identified by 250 bp, especially those spanning large chromosomal regions. (J) The longest MethyMark (chr12:34489365-34507836) identified by only max distance=500. (K-P) To determine the threshold of interval CpG number between neighboring primary segments in merging two neighboring primary segments, we used different interval CpG numbers (1, 2, 3, 4, and 5) between two primary segments as thresholds for merging neighboring segments. The comparison of the results from five thresholds is shown in the following figures. (K) Number of segments identified by different thresholds. (L) Total length of segments identified by different thresholds. (M) Distribution of CpG number of segments. (N) Distribution of GC content of segments. (O) Distribution of CpG number per 100 bp of segments. (P) Distribution of Obs/Exp CpG of segments. These results indicate five interval CpG number should be useful for merging the primary segments separated by a few CpGs whose methylation levels may be distorted by potential random errors caused by incomplete bisulfite conversion and sequencing errors. In addition, this threshold has no effect on the features of segments, such as CpG density.

Supplementary Figure 6: The longest segment identified by SMART in the human genome. The longest segment was located at chr10:39103301-39130657, covers 27k bases, includes 449 CpGs, and is part of partially methylated domains (PMDs) identified in IMR90 by Lister et al. 2009.

Supplementary Figure 7: The features of human methylation segments identified by SMART. (A) The length of three types of segments. The length of segments ranges from 20 bp to 10 kb, including 5406 segments with a length of at least 3.5 kb. (B) The CpG number of three types of segments. The CpG number in these segments ranges from 5 to 1000, including 288 segments that have more than 150 CpGs. (C) Distribution of mean methylation of HighSpe, InterSpe and LowSpe segments. (D) Distribution of median methylation of HighSpe, InterSpe and LowSpe segments. (E) Distribution of methylation specificity of HighSpe, InterSpe and LowSpe segments. (F) Distribution of Obs/Exp CpG of HighSpe, InterSpe and LowSpe segments. (G) Distribution of Obs/Exp CpG of CpG

island, shore and desert segments. (H) Density scatterplot of Obs/Exp CpG and methylation specificity of all segments. (I) Density scatterplot of Obs/Exp CpG and mean methylation of all segments.

Supplementary Figure 8: Four classifications of LowSpe segments based on mean/median methylation. (A) Histogram of mean/median methylation across 50 samples of LowSpe segments. The distribution of methylation levels of these segments showed five peaks. Two peaks approximately 0.75 are close to each other and are smaller than other three. The methylation difference between these two peaks is approximately 0.05, which is usually regarded as meaningless in methylation analysis, thus we treated two peaks as the same methylation state: partial-high-methylation. (B) Heat map of methylation levels of 562,719 LowSpe segments in 50 methylomes. Each row represents a segment. The segments are ordered by their mean methylation levels in 50 samples from low to high. Seven methylation values were given in the right panel. For each segment, its cluster classification in K means (K=2, 3, 4, and 5) clustering shown in Figures C-F is given in the right panel. (C-F) The K means (K=2, 3, 4, and 5) clustering based on the 50 methylomes of LowSpe segments. The segments in Cluster1 and those in Cluster2 by 4-means clustering showed large methylation changes, but they were segmented into a cluster by the 3-means clustering. In addition, the greatest methylation difference among the segments in Cluster4 by 5-means clustering is only 0.07, which is usually regarded as meaningless in DNA methylation analysis. These results suggest the justifiability of four clusters, including UniHypo (0.00~0.25), UnipLow (0.25~0.60), UnipHigh (0.60~0.80) and UniHyper (0.80~1.00).

Supplementary Figure 9: Features of LowSpe segments. (A) Distribution of CpG number per 100 bp of segments in four groups of LowSpe segments including UniHypo, UnipLow, UnipHigh and UniHyper. (B) Distribution of Obs/Exp CpG in four groups of LowSpe segments. UniHypo segments showed higher CpG density, which was a typical feature of CpG islands (CGIs). (C) The base overlap rate between CpG islands and four groups of LowSpe segments. UniHypo segments showed higher overlap rate than other groups of LowSpe segments. (D) Overlap between CGIs UniHypo segments and promoter UniHypo segments. More than 7,000 UniHypo segments were overlapped with promoter CGIs. In addition, we found 88% of LowSpe segments that were located in promoters of housekeeping genes were significantly overlapped with uniformly hypomethylated CGIs ($p<10^{-282}$, Chi-square test). (E) Chromatin modification patterns of LowSpe segments. The chromatin modifications (H3K4me3, H3K27me3, H3K9me3, H2K27ac, EP300) of the segments in each LowSpe group in H1 cell line. Average enrichment profiles of log2 ratios of several histone marks and transcription factor vs. DNA input around ±3 Kb regions of different types of LowSpe segments. “L” and “R” represent the boundary of HypoMark/HyperMark. Ngs.plot (http://code.google.com/p/ngsplot/) was used to visualize the average profiles and heat maps with fragment length equal to 300 bp and other default parameters.

Supplementary Figure 10: UniHypo segments and ubiquitous enhancers. (A) Overlap between different types of segments and ubiquitous enhancers. It was revealed that ubiquitous enhancers were prone to overlap with UniHypo segments (p<10-10,

Chi-square test). (B) Number of associated genes per UniHypo segment overlapped with U-enhancer. (C) Genome location of UniHypo segment overlapped with U-enhancer relative to transcription start site (TSS) of an associated gene. (D) 312 genes related to UniHypo segment overlapped with U-enhancer. Among these genes, 66 genes have been reported as housekeeping genes, including the well-known *CTCF*. (E) Functional enrichment analysis of genes related to 223 UniHypo segments overlapped with ubiquitous enhancers. Top 10 biological processes and top 10 pathways were shown. It was revealed these genes were enriched in functional terms involving fundamental biological processes (such as macromolecule biosynthesis) and metabolic pathways (such as the mTOR signaling pathway).

Supplementary Figure 11: The methylation pattern and chromatin states of the promoter regions of *CTCF*. *CTCF* encodes a transcriptional regulator protein with 11 highly conserved zinc finger (ZF) domains and owns two UniHypo segments. Two UniHypo segments were overlapped with multiple active features including a CGI, ubiquitous enhancer and transcriptional factor binding sites, an active chromatin state (promoter-associated state represented by red color), and active histone modifications (H3K4me3 and H3K27ac) in its promoter region.

Supplementary Figure 12: HighSpe segments identified by SMART in the human genome. (A) Composite plot of methylation specificity across HighSpe and InterSpe segments and differentially methylated regions (DMRs) across human tissues/cells that were identified by previous studies based on the methylomes profiled by different technologies including BS-Seq, MeDIP-chip/seq, HumanMethylation450 and CHARM (7-11). The methylation specificity near HighSpe and InterSpe segments revealed a pattern of high methylation specificity in the body of HighSpe and InterSpe segments and low specificity in their flanking sequences. The DMRs across human tissues/cells that were identified by previous studies showed similar results with our study, confirming the accuracy of the methylation specificity quantified by SMART and the reliability of methylation segments identified in this study. (B) The principal component analysis of 50 methylomes. This figure revealed the specific methylation pattern in sperm and the clustering of pluripotent cell lines.

Supplementary Figure 13: K-means clustering of HighSpe segments. In each panel, methylation levels are represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of 50 samples that were classified into six main groups tagged by different colors and abbreviations: Pluripotent cells (P), Epithelial cells (E), Sperm cells (S), Neuronal cells (N), Thymocytes (T) and Others (O). (A) 6-means clustering of HighSpe segments. Six clusters of segments are differentially colored on the right. (B) 10-means clustering of HighSpe segments. Ten clusters of segments are differentially colored on the right. (C) A larger version of 8-means clustering of HighSpe segments shown in Figure 1H. On the left, the cluster of each segment in 6-means clustering and 10-means clustering are given as the cluster color defined in A and B. On the right, eight clusters are given, and examples of the related genes for each cluster are also listed.

Supplementary Figure 14: Genome location of different clusters of HighSpe segments. Radar plots showing the ratio of observed to expected HighSpe segments in different clusters and genome features including CpG islands (CGI), CpG island shores (CGIshore) and Refseq genes related seven categories including upstream 2 kb of transcription start site (Up2kb), 5'UTR, Coding Exon (CodingExon), Intron, 3'UTR, downstream 2 kb of transcription end site (Down2kb), and noncoding RNAs (ncRNAs). To examine whether the HighSpe segments in specific clusters are enriched in some specific genome features, we calculated the number of HighSpe segments in each cluster (Cluster HighSpe Num.), the number of HighSpe segments in each genome feature (Feature HighSpe Num.), the number of overlapped HighSpe segments between Cluster HighSpe and Feature HighSpe (Cluster&Feature HighSpe Num.), and the number of total HighSpe segments identified (Total HighSpe Num.). The ratio of observed to expected HighSpe segments (Obs/Exp HighSpe) in each cluster and genome feature was calculated as $Obs/Exp\ HighSpe = \frac{(Cluster\ \&\ Feature\ HighSpe\ Num.) \times (\text{Total}\ HighSpe\ Num.)}{(Cluster\ HighSpe\ Num.) \times (Feature\ HighSpe\ Num.)}$. The center of the plot was 0, and a colored dot on the respective axis indicates the Obs/Exp HighSpe of the HighSpe from specific cluster (colored line) in a specific genome feature (angle). It was obvious that the HighSpe segments in Clusters 1, 3 and 5 exhibit high Obs/Exp HighSpe in CGI, suggesting potential roles of methylation dynamics in CGIs in cell type identity. In addition, the HighSpe segments in Cluster 7 show enrichment in CGI shores and Up2kb.

Supplementary Figure 15: The enriched functions of genes related to HighSpe segments in each cluster. On the right, the representative and significant biological processes or KEGG pathways are shown. For the functional analysis of genes related to each cluster of HighSpe segments, the genes related to HighSpe segments with length ≥200 bp in each cluster were selected. Due to the limitation of gene number in DAVID, we adopted more stringent standards for selection of genes in cluster 4 and cluster 6, both of which were related to more than 3,000 genes. For cluster 4, only genes with promoter HighSpe segments with length ≥200 bp were selected, and for cluster 6, only genes with HighSpe segments with length ≥200 bp in the regions from upstream 2 kb to transcription start site (TSS) were selected. Then, the selected genes in each cluster were imported into DAVID to perform functional enrichment analysis of these genes in biological process and the KEGG pathway. Finally, 371 enriched function terms were clustered and visualized by R.

Supplementary Figure 16: K-means clustering of InterSpe segments. In each panel, methylation levels were represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of the 50 samples that were classified into five main groups tagged by different color and abbreviation: Pluripotent cells (P), IMR90 (I), Sperm cells (S), Neuro cells (N) and Others (O). (A) 4-means clustering of InterSpe segments. Four clusters of segments are differentially colored on the right. (B) 8-means clustering of InterSpe segments. Eight clusters of segments are differentially colored on the right. (C) 6-means clustering of InterSpe segments. On the left, the cluster of each segments in 4-means clustering and 8-means clustering are given as the cluster color defined in A and B. On the right, six clusters are listed.

Supplementary Figure 17: K-means clustering of MethyMarks. In each panel, methylation levels were represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of the 50 samples that were classified into six main groups tagged by different color and abbreviation: Pluripotent cells (P), Epithelial cells (E), Sperm cells (S), Neuro cells (N), Thymocytes (T) and Others (O). (A) 6-means clustering of MethyMarks. Six clusters of MethyMarks are differentially colored on the right. (B) 10-means clustering of MethyMarks. Ten clusters of MethyMarks are differentially colored on the right. (C) 8-means clustering of MethyMarks. On the left, the cluster of each MethyMark in 6-means clustering and 10-means clustering are given as the cluster color defined in A and B.

Supplementary Figure 18: Genome location and functional enrichment of cell type-specific MethyMarks. (A) Percentage of HypoMarks and HyperMarks overlapped with CpG islands (CGI), CGI shores and CGI deserts. (B) Percentage of HypoMarks and HyperMarks overlapped with different genome features including Up2kb, 5'UTR, CodingExon, Intron, 3'UTR, Down2kb and Intergenic. (C) Percentage of MethyMarks overlapped with cell type-specific active enhancer. (D) High expression and functional enrichment of HypoMark genes. For the functional analysis of genes related to cell-specific HypoMarks, the genes with promoter HypoMarks in each cell type were selected. For testis spermatozoa primary cells, only the genes with promoter HypoMarks with length ≥700 bp were selected. Then, the selected genes in each cell type were imported into DAVID to perform functional enrichment analysis in over-expressed tissue, biological process and the KEGG pathway. For each type of analysis, the three most significant terms were selected and visualized by R. The grids colored from white (0) to dark red (28) represent the –log10 of the p value for the enrichment of HypoMark genes in each cell type (Column) in each function term (Row). The function terms that were related to the cell types in this study were bolded and colored blue (over-expression), red (biological processes) and purple (KEGG pathway).

Supplementary Figure 19: Cell type-specific MethyMarks that span large chromosomal regions. (A) The length of cell type-specific MethyMarks identified by SMART. (B) Genome location and epigenomic features of the longest MethyMarks identified. Each methylation track represents a cell type, and the height of the bar represents the methylation level. Super-enhancer, chromatin states (the bar colored in blue was for insulator, red for active promoter, orange for strong enhancer, yellow for weak enhancer, and light green for weak transcribed), and H3K27ac (the height of bar represents the number of reads overlapping each 25 bp bin) are shown at the bottom.

Supplementary Figure 20: Example genes related to cell type-specific MethyMarks that span large chromosomal regions. (A) An example for large HyperMark genes in pluripotent cells. This HyperMark was specifically hypermethylated in pluripotent cells and overlapped with *PCHHB11*. (B) An example for large HypoMark genes in pluripotent cells. This HypoMark was specifically hypomethylated in pluripotent cells and overlapped with *Ac005062.2*. (C) iPSC cells DF 19.11 showed a different methylation pattern compared to H1 hESC cells in the MethyMarks, which overlapped with a large CpG island and *IRX1*. The histone modifications and mRNA were obtained from

http://www.genboree.org/EpigenomeAtlasBrowser/. (D) The DNA methylation and expression pattern of imprinted gene, *MEG3*. MRE and MeDIP tracks represent the un-methylated and methylated CpGs in the brain, respectively. The 100 vertebrates' basewise conservation by PhyloP is shown in the bottom track.

Supplementary Figure 21: Detailed epigenetic modifications in the iPSC-specific HyperMarks related to *IRX2* across cell types. Each whole-genome bisulfite sequencing (WGBS) methylation track shows the methylation of a cell type, and the height of the bar represents the methylation level. The histone modifications, chromatin states and methylation level by reduced representation bisulfite sequencing (RRBS) in various samples including cancer were shown. Each RRBS methylation track represents a cell type, and the color of bar represents the methylation level from unmethylated (green) to full methylation (red). This HyperMark showed specific hypermethylation in iPSC cells DF 19.11 but hypomethylation in other cell types. As shown by RRBS methylation, the aberrant hypermethylation of this mark may cause the deactivation of *IRX2* in cancer. For instance, this mark showed specific hypermethylation levels and inactive chromatin states in human cancer cell lines, including K562 and HepG2.

Supplementary Figure 22: Heatmap of DNA methylation and H3K27ac in cell type-specific HyperMarks. Each row denotes a HyperMark, and each column a cell type. DNA methylation levels are represented by a gradient from green (ummethylated) to red (full methylation) and H3K27ac from white (lowest) to red (highest). The density of H3K27ac was represented by the read count per million mapped reads (RPKM).

Supplementary Figure 23: Correlation between DNA methylation and H3K27ac in HypoMarks and HyperMarks of various cell types. (A) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in HypoMarks of a cell type. The Spearman's rank correlation coefficient (SCC) between DNA methylation and H3K27ac in HypoMarks was calculated for each cell type, respectively. P represents the significance of the coefficient. (B) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in HyperMarks of a cell type. The SCC between DNA methylation and H3K27ac in HyperMarks was calculated for each cell type, respectively. P represents the significance of the coefficient.

Supplementary Figure 24: Correlation between DNA methylation and two histone marks (H3K27ac and H3K27me3) in different categories of segments. (A) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in different categories of segments including all segments, HighSpe, InterSpe, UniHypo, UnipLow, UnipHigh UniHyper segments, and H1 specific MethyMarks. Spearman's rank correlation coefficient (SCC) between DNA methylation and H3K27ac in each category of segments was calculated by the R function "cor.test". P represents the significance of the coefficient. (B) Each sub-figure shows the density scatterplot of DNA methylation and H3K27me3 in different categories of segments including all segments, HighSpe, InterSpe, UniHypo, UnipLow, UnipHigh UniHyper segments, and H1 specific MethyMarks. SCC represents Spearman's rank correlation coefficient between DNA methylation and H3K27me3 in each category of segments. *P* represents the significance of the coefficient.

Supplementary Figure 25: hESC H1-specific HypoMarks and HyperMarks show distinct chromatin modifications. (A) Heatmap of log2 enrichment ratios of several histone marks and transcription factors vs. DNA input at HypoMark/HyperMark ±3 Kb regions mapped by ngs.plot (12). “L” and “R” represent the boundary of HypoMark/HyperMark. The log2 enrichment ratios were represented by colors from green (low) to red (high). (B) Average profiles of log2 enrichment ratios of several histone marks and transcription factors vs. DNA input at promoter HypoMark/HyperMark ±3 Kb regions.

Supplementary Figure 26: Sub-network of transcription factor-MethyMark collaboration network in hESC H1 cells as shown in Figure 4D. (A) Sub-network derived by transcriptional factors from transcription factor-MethyMark collaboration network in hESC H1 cells. The size of the transcription factor (TF) node represents the number of the MethyMarks bound by it, and the width of the TF-TF line represents the number of MethyMarks co-targeted by two TFs. (B) Sub-network derived by H1 MethyMarks from TF-MethyMark collaboration network in the hESC H1 cell line. The width of the MethyMark-MethyMark line represents the number of TFs binding to both MethyMarks. Only the lines with more than ten TFs are shown. (C) Sub-network derived by transcriptional factors NANOG and POU5F1 from TF-MethyMark collaboration network in the hESC H1 cell line. From TF-MethyMark collaboration network in the hESC H1 cell line of Fig. 4, NANOG and POU5F1 and their one-step neighboring nodes and the lines between these nodes were extracted to construct this sub-network. It was shown that most methylated segments in this sub-network were H1-specific HypoMarks, and these HypoMarks were prone to be bound by the same active TFs. (D) Functional enrichment of H1-specific HypoMarks in NANOG and POU5F1 related sub-network. GREAT (http://bejerano.stanford.edu/great/public/html/) was used to perform the functional enrichment of H1-specific HypoMarks in the sub-network. H1-specific HypoMarks were assigned to nearby protein-coding genes based on GREAT’s basal plus extension rule for regulatory regions (proximal: 5 kb upstream, 1 kb downstream, plus distal up to 1 Mb). Significant annotated terms from the enrichment analysis were selected by both hypergeometric and binomial tests (P<0.05). Four enriched functions were found as targets of TF NANOG, POU5F1 and SOX2, overexpression in human ESC, functions related with embryonic development, and abnormal developmental phenotype.

Supplementary Figure 27: DNA methylation and H3K27ac state of cell type-specific SuperHypoMarks across 21 cell types. Each row denotes a SuperHypoMark, and each column denotes a cell type. DNA methylation level was represented by a gradient from green (ummethylated) to red (full methylation), and H3K27ac from white (lowest) to red (highest). RPKM represents the H3K27ac reads per kilobase per million mapped reads in a given segment.

Supplementary Figure 28: Detailed information about epigenetics and expression of H1-specific SuperHypoMark genes. Shown in this figure are the example genes related to H1-specific SuperHypoMarks. For each gene, the epigenetic pattern was visualized by our local UCSC genome browser, and the expression patterns of the genes were visualized by Epigenome Atlas Browser (http://www.genboree.org/EpigenomeAtlasBrowser). The detailed analysis for each gene

is listed as followings:

***POU5F1*:** The promoter region of POU5F1 is H1-specific hypomethylated and bound by mediator coactivators including RNA polymerase II, mediator and transcription factors (such as POU5F1 and NANOG), which form a super-enhancer in this region. We also found the *POU5F1* promoter was enriched by histone H3K27ac, a surrogate mark of a super-enhancer, and H3K4me3, an active mark for gene expression that has been identified as an H1-specific promoter or enhancer state by a hidden Markov model (Ernst et al. 2011). Furthermore, *POU5F1* was specifically expressed at extremely high levels in the H1 cell line.

***NANOG*:** We found *NANOG* showed very similar epigenetic and expression patterns to *POU5F1*.

***DNMT3B*:** Interestingly, the DNA methyltransferase DNMT3B that was essential for de novo methylation and mammalian development (Okano et al. 1999) showed H1-specific hypomethylation and extremely high expression levels, which is consistent with its downregulation after ES cell differentiation as reported previously (Okano et al. 1998; Watanabe et al. 2002). The unmethylated status of the promoter regions facilitates the formation of a super-enhancer, which accounts for the extremely high expression of *DNMT3B*. In stem cells, the high expression of *DNMT3B* induces high levels of DNA methyltransferase, which further methylates most genome CpGs except those related to pluripotency maintenance.

***NSD1*:** Another H1-specific hypomethylated gene *NSD1* (also known as *KMT3B*) encodes a histone methyltransferase that preferentially methylates H3 lysine 36. The methylation data by another technology, Infinium Methylation 450K, confirms low methylation levels of this region in the H1 cell line but high methylation levels in adult tissues and other cell types. Furthermore, *NSD1* shows higher expression in the H1 cell line than other cell types.

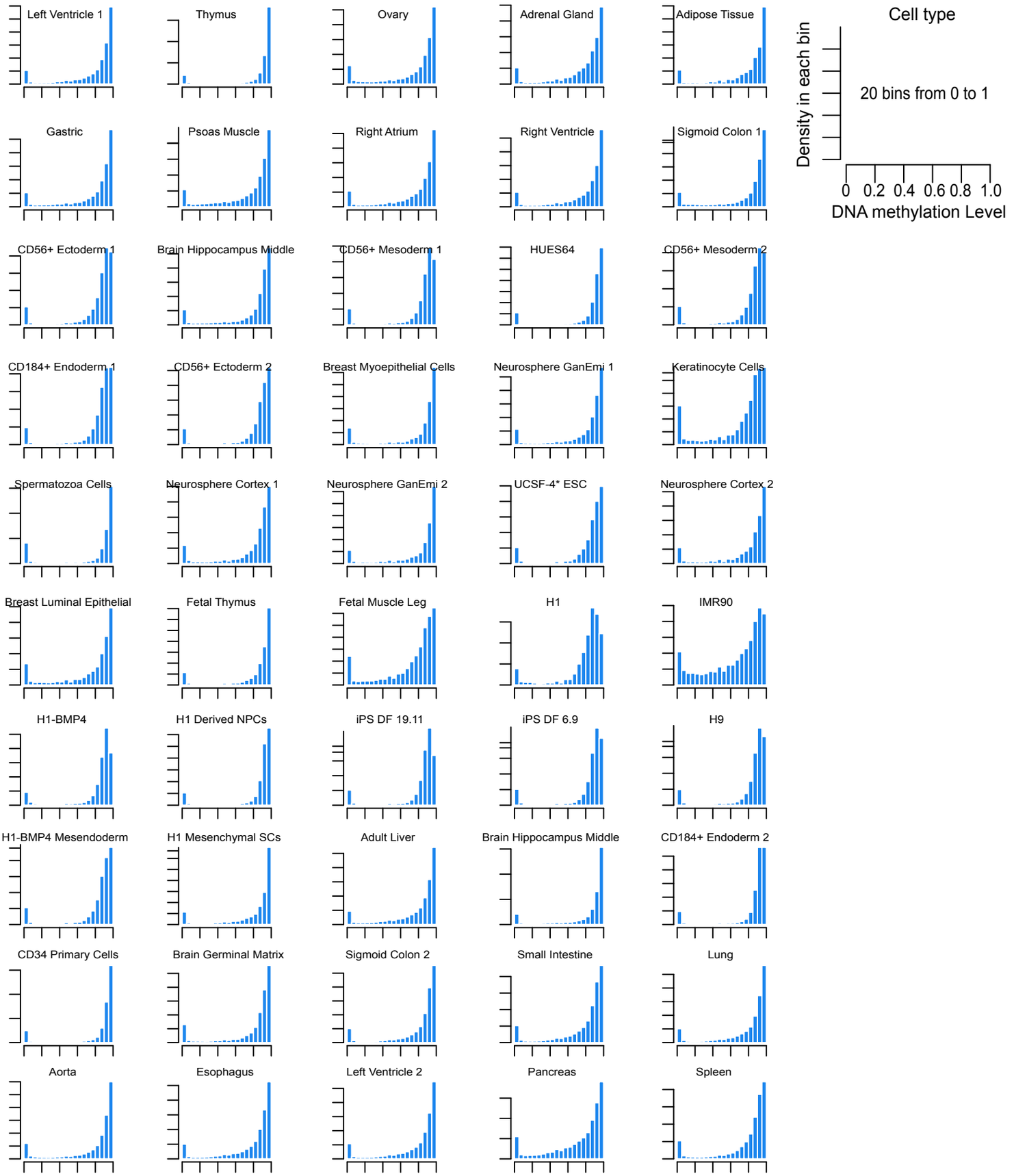
***LINC00678*:** This gene was one of 22 lncRNA genes that overlapped with H1-specific SuperHypoMarks. It was shown that the promoter region of this gene was specifically hypomethylated and extremely highly expressed in ESC and iPSC cell lines. However, the expression level of this gene in extremely low, which is consistent with a previous finding of its down-regulation during the transition from iPSCs to NPCs (Chen et al. 2013). As far as we know, there were no more reports about the functions of *LINC00678* in stem cells, suggesting it may be a novel mark of stem cells.

***miR-6130*:** A microRNA gene *miR-6130* overlapped with two ESC-specific SuperHypoMarks. We found *miR-6130* was the longest microRNA gene (836, 530 bp) in the list of Refseq genes. It was specifically hypomethylated in ESCs and iPSCs and overlapped with a super enhancer that was only found in the H1 cell line. As far as we know, there were no more reports about the functions of *miR-6130* in stem cells, suggesting it may be a novel mark of stem cells.

Reference

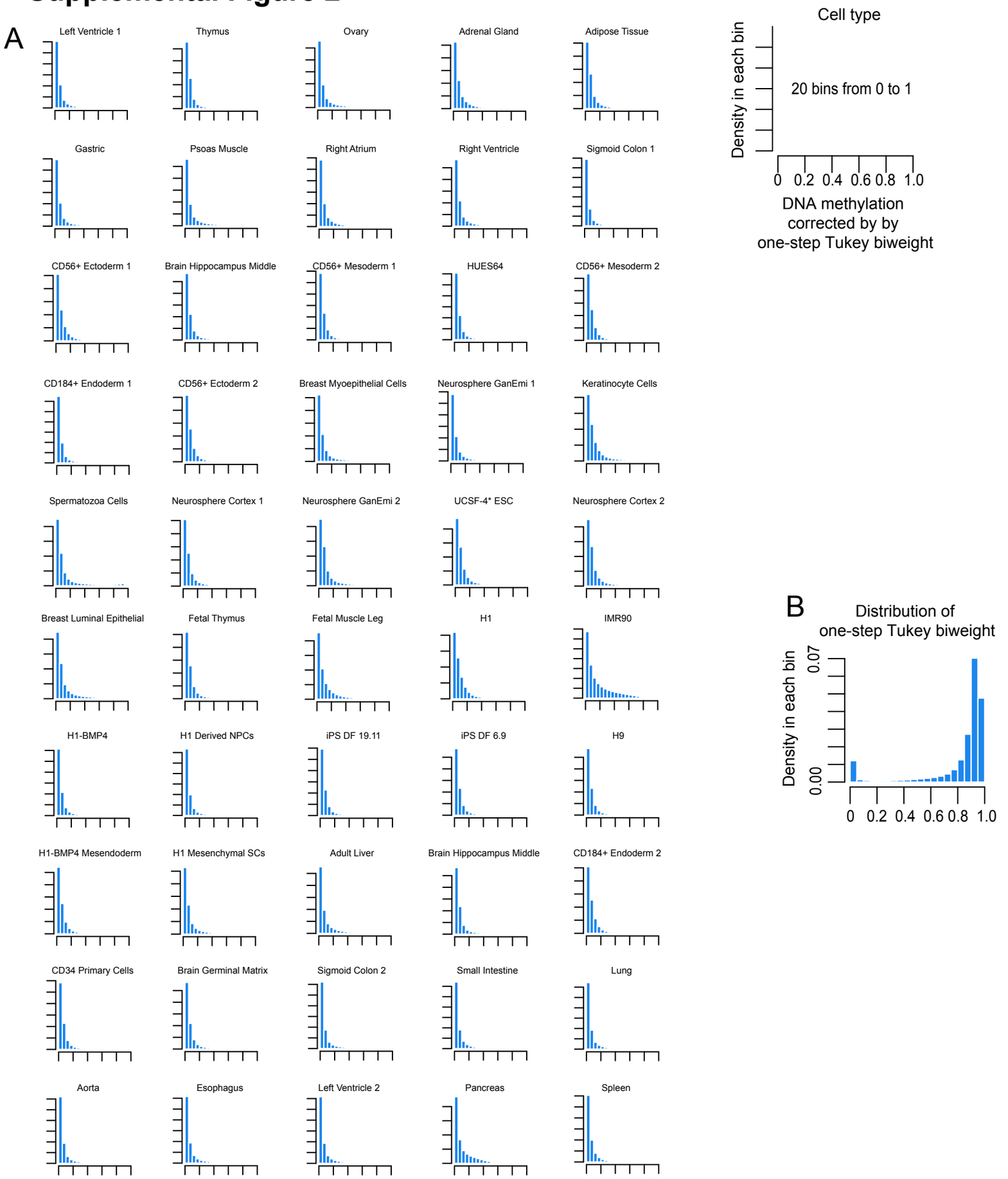
1. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. *et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*, **28**, 1045-1048.
2. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*, **32**, D493-496.
3. Eisenberg, E. and Levanon, E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet*, **29**, 569-574.
4. Wei, Y., Su, J., Liu, H., Lv, J., Wang, F., Yan, H., Wen, Y., Liu, H., Wu, Q. and Zhang, Y. (2014) MetaImprint: an information repository of mammalian imprinted genes. *Development*, **141**, 2516-2523.
5. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455-461.
6. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934-947.
7. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*, **41**, 178-186.
8. Zhang, B., Zhou, Y., Lin, N., Lowdon, R.F., Hong, C., Nagarajan, R.P., Cheng, J.B., Li, D., Stevens, M., Lee, H.J. *et al.* (2013) Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome research*, **23**, 1522-1540.
9. Zhang, Y., Liu, H., Lv, J., Xiao, X., Zhu, J., Liu, X., Su, J., Li, X., Wu, Q., Wang, F. *et al.* (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res*, **39**, e58.
10. Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477-481.
11. Lokk, K., Modhukur, V., Rajashekar, B., Martens, K., Magi, R., Kolde, R., Kolt Ina, M., Nilsson, T.K., Vilo, J., Salumets, A. *et al.* (2014) DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol*, **15**, R54.
12. Shen, L., Shao, N., Liu, X. and Nestler, E. (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics*, **15**, 284.

Supplemental Figure 1



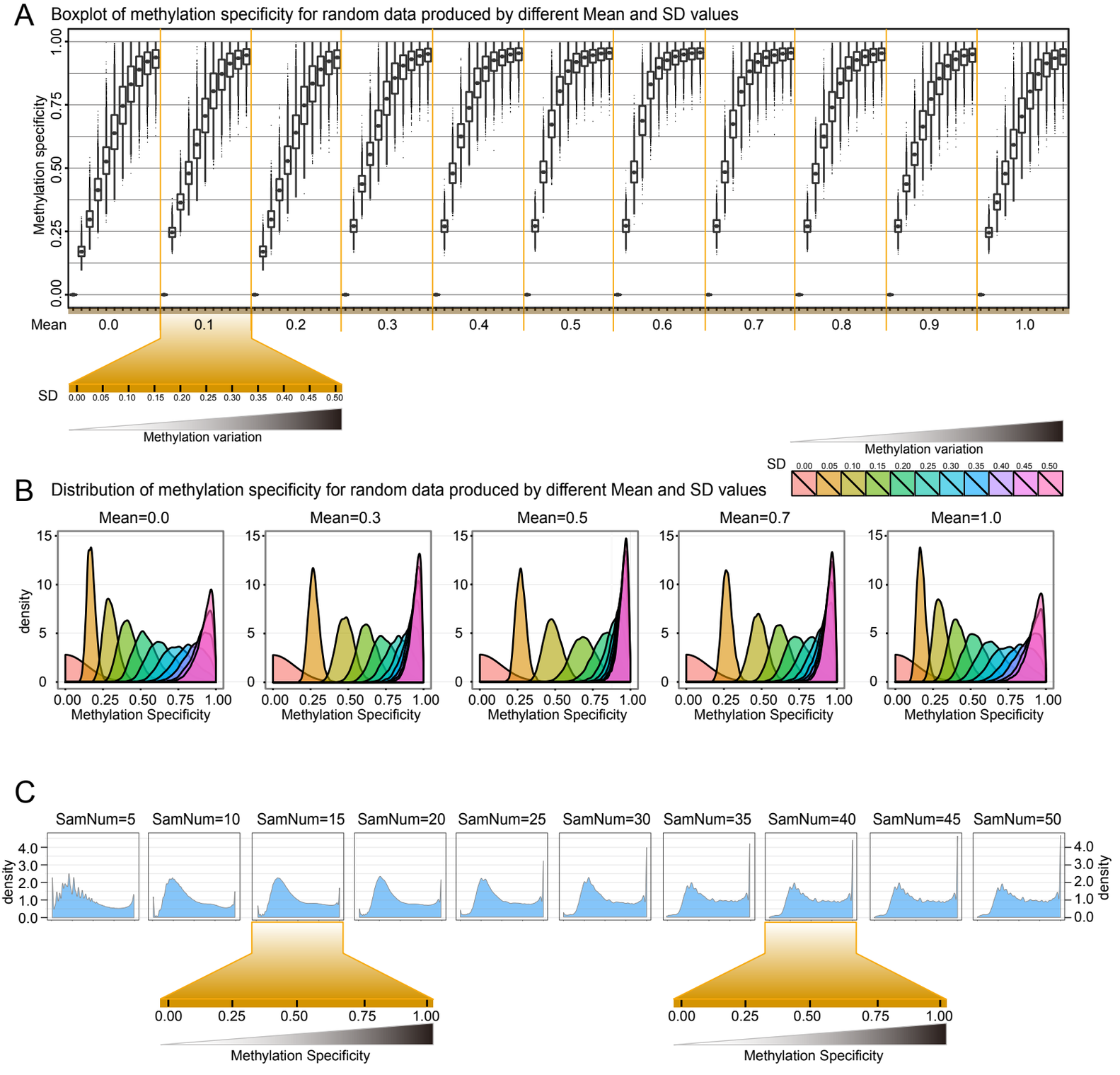
Supplementary Figure 1: Distribution of DNA methylation in 50 methylomes. Each sub-graph represents the distribution of CpG methylation in specific human tissues or cell lines. The DNA methylomes showed a bimodal distribution in all cell types, and most CpGs were hypermethylated.

Supplemental Figure 2



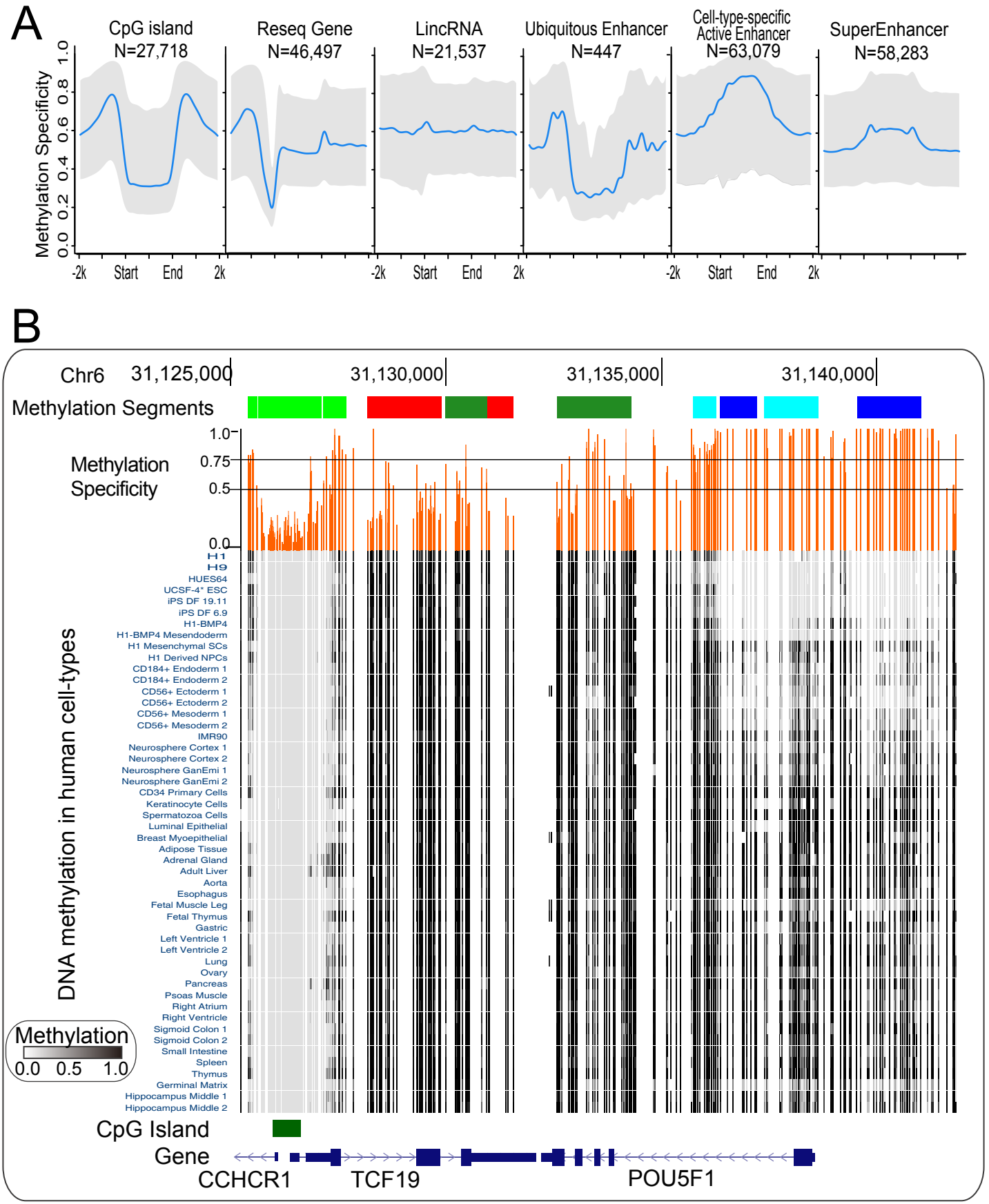
Supplementary Figure 2: Distribution of DNA methylation corrected by one-step Tukey biweight in 50 methylomes. (A) Each sub-graph represents the distribution of CpG methylation corrected by one-step Tukey biweight in specific human tissues or cell lines. To determine methylation specificity, one-step Tukey biweight was calculated as a robust weighted mean using the methylation levels in the majority of cell types after discounting the outliers in the minority of cell types by a weight that was calculated by the bisquare function. (B) The distribution of one-step Tukey biweights for all CpG sites.

Supplemental Figure 3



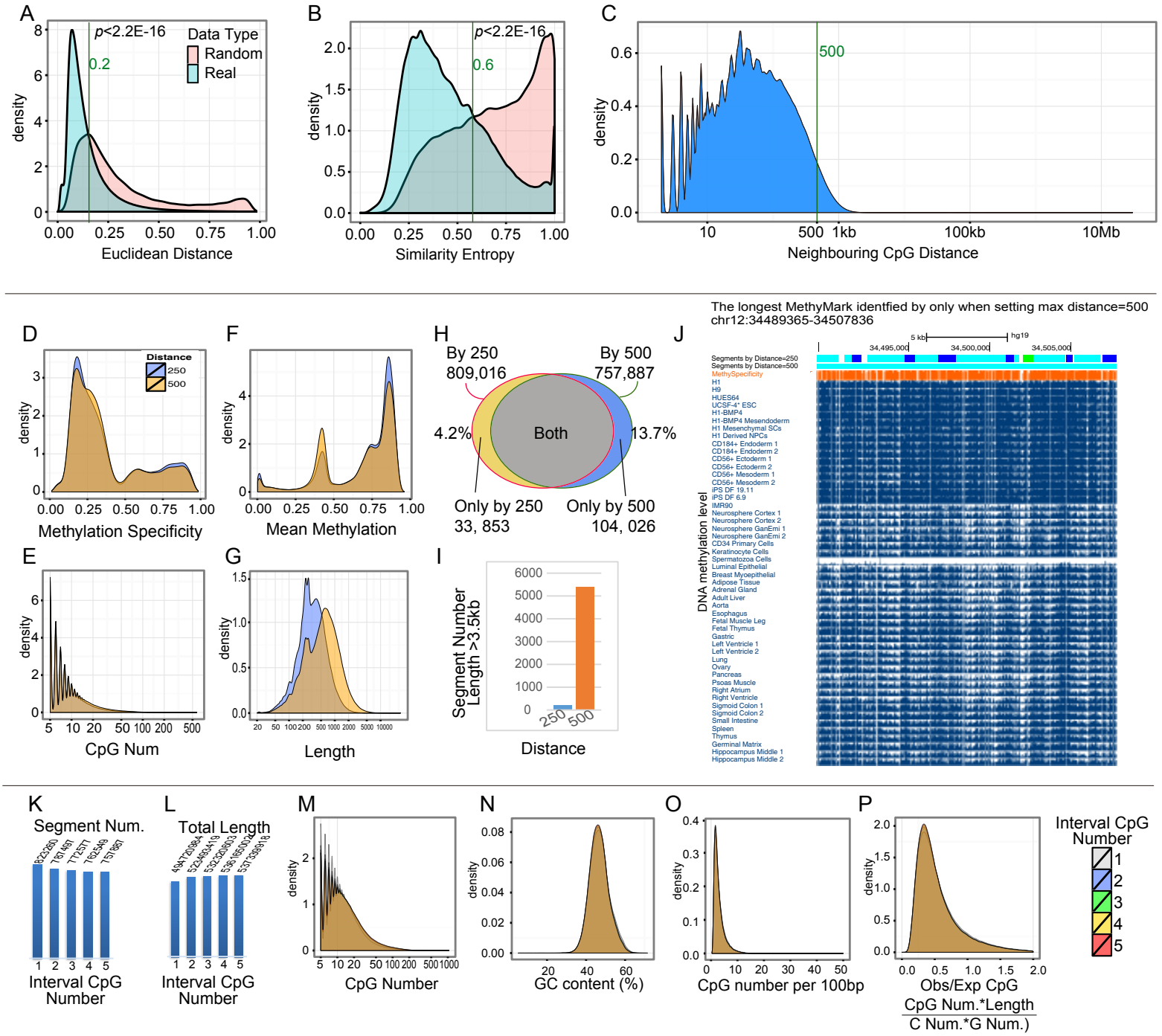
Supplementary Figure 3: Evaluation of the performance of an entropy-based algorithm in the quantification of methylation specificity across multiple samples. To determine the thresholds for methylation specificity, we modelled different methylation patterns by random sampling from different normal distributions with different mean and standard deviation values and studied the distribution of methylation specificity. For a given mean methylation level (mean, ranging from 0.0 to 1.0) and a given standard deviation (SD, ranging from 0.0 to 0.5), 50 values were randomly sampled as the methylation levels in 50 samples of a CpG site. This process was repeated 10,000 times to produce 10,000 CpG sites whose methylation specificity across 50 samples was quantified by our method as described in the manuscript. Then, the distribution of these methylation specificity values was used to evaluate the accuracy of our method in the quantification of methylation specificity and determine the thresholds for classification of degree of methylation specificity. (A) Boxplot of methylation specificity for random data produced by different mean and SD values. (B) Distribution of methylation specificity for random data produced by different mean and SD values. (C) Distribution of methylation specificity calculated using a different number of samples. The similar distribution of methylation specificity suggested our method should be applicable to datasets with different sample numbers.

Supplementary Figure 4



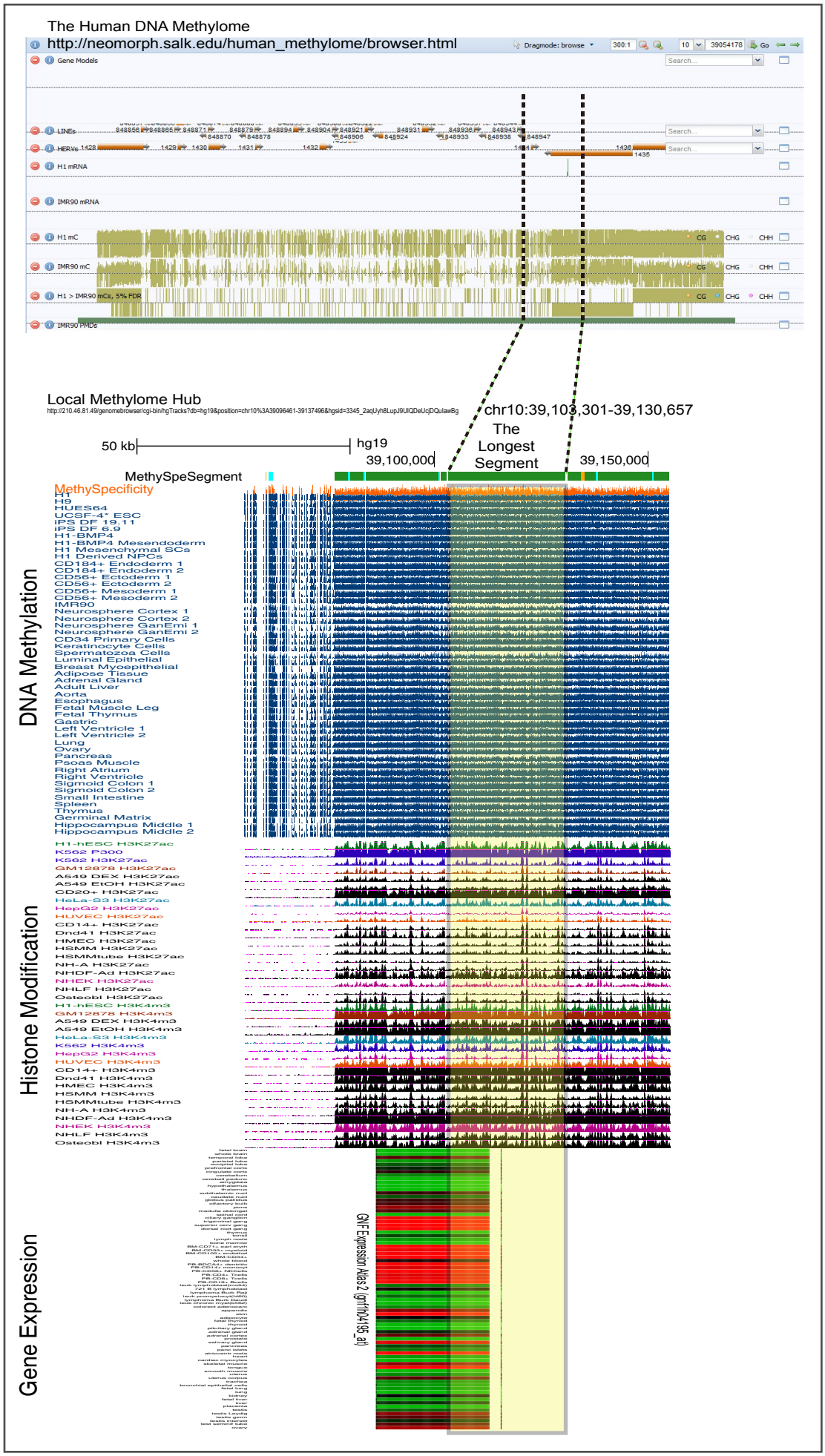
Supplementary Figure 4: Methylation specificity across known gene regulatory regions. (A) Composite plot of methylation specificity quantified by normalized Shannon entropy across known regulatory elements, including CpG islands, Refseq genes, long noncoding RNAs (lncRNA), ubiquitous enhancers, cell type-specific enhancers and super-enhancers. Blue lines indicate the median of the methylation specificity across each element, and grey areas mark the twenty-fifth and seventy-fifth percentiles of methylation specificity. (B) Methylation specificity quantified by normalized Shannon entropy and methylation segments identified by SMART near the developmental gene *POU5F1* (also known as *OCT4*).

Supplementary Figure 5



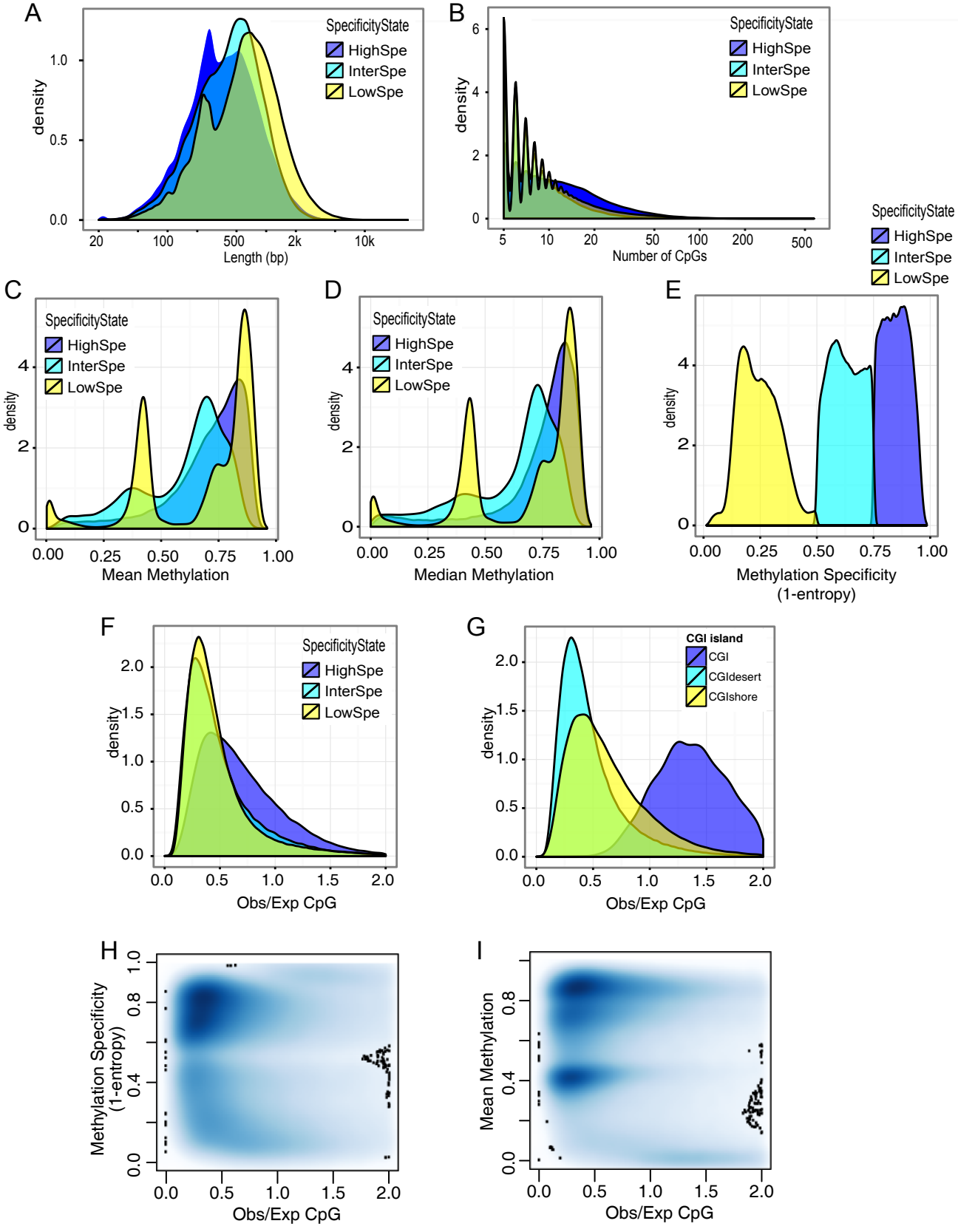
Supplementary Figure 5: Determination of thresholds for merging neighboring CpGs. (A) Distribution of Euclidean distance of DNA methylation levels between neighboring CpGs in real and random datasets. (B) Distribution of similar entropy of DNA methylation levels between neighboring CpGs in real and random datasets. (C) Distribution of distance between neighboring CpGs. (D-J) To determine the threshold of max distance between neighboring CpGs in merging two neighboring CpGs, we performed genome segmentation, setting the same other parameters, but the max distance between two CpGs was set as 250 bp and 500 bp. The comparison of the results from two thresholds is shown in the following figures. (D) Distribution of methylation specificity of segments. (E) Distribution of CpG number of segments. (F) Distribution of mean methylation level of segments. (G) Distribution of length of segments. (H) Number of segments identified by 250 bp and 500 bp. Approximately 13.7% of the segments identified by 500 bp were not overlapped with any segment identified by 250 bp, while only 4.2% of the segments identified by 250 bp were not overlapped by any segment identified by 500 bp. This result suggests that some of the segments identified by the threshold of 250 bp were not lost but rather merged into larger segments by the threshold of 500 bp. (I) Number of segments with a length > 3.5 kb, which was used to identify long hypomethylated genome regions by Jeong et al. It is suggested that the threshold of 500 bp can be used to identify those segments not identified by 250 bp, especially those spanning large chromosomal regions. (J) The longest MethyMark (chr12:34489365-34507836) identified by only max distance=500. (K-P) To determine the threshold of interval CpG number between neighboring primary segments in merging two neighboring primary segments, we used different interval CpG numbers (1, 2, 3, 4, and 5) between two primary segments as thresholds for merging neighboring segments. The comparison of the results from five thresholds is shown in the following figures. (K) Number of segments identified by different thresholds. (L) Total length of segments identified by different thresholds. (M) Distribution of CpG number of segments. (N) Distribution of GC content of segments. (O) Distribution of CpG number per 100 bp of segments. (P) Distribution of Obs/Exp CpG of segments. These results indicate five interval CpG number should be useful for merging the primary segments separated by a few CpGs whose methylation levels may be distorted by potential random errors caused by incomplete bisulfite conversion and sequencing errors. In addition, this threshold has no effect on the features of segments, such as CpG density.

Supplementary Figure 6



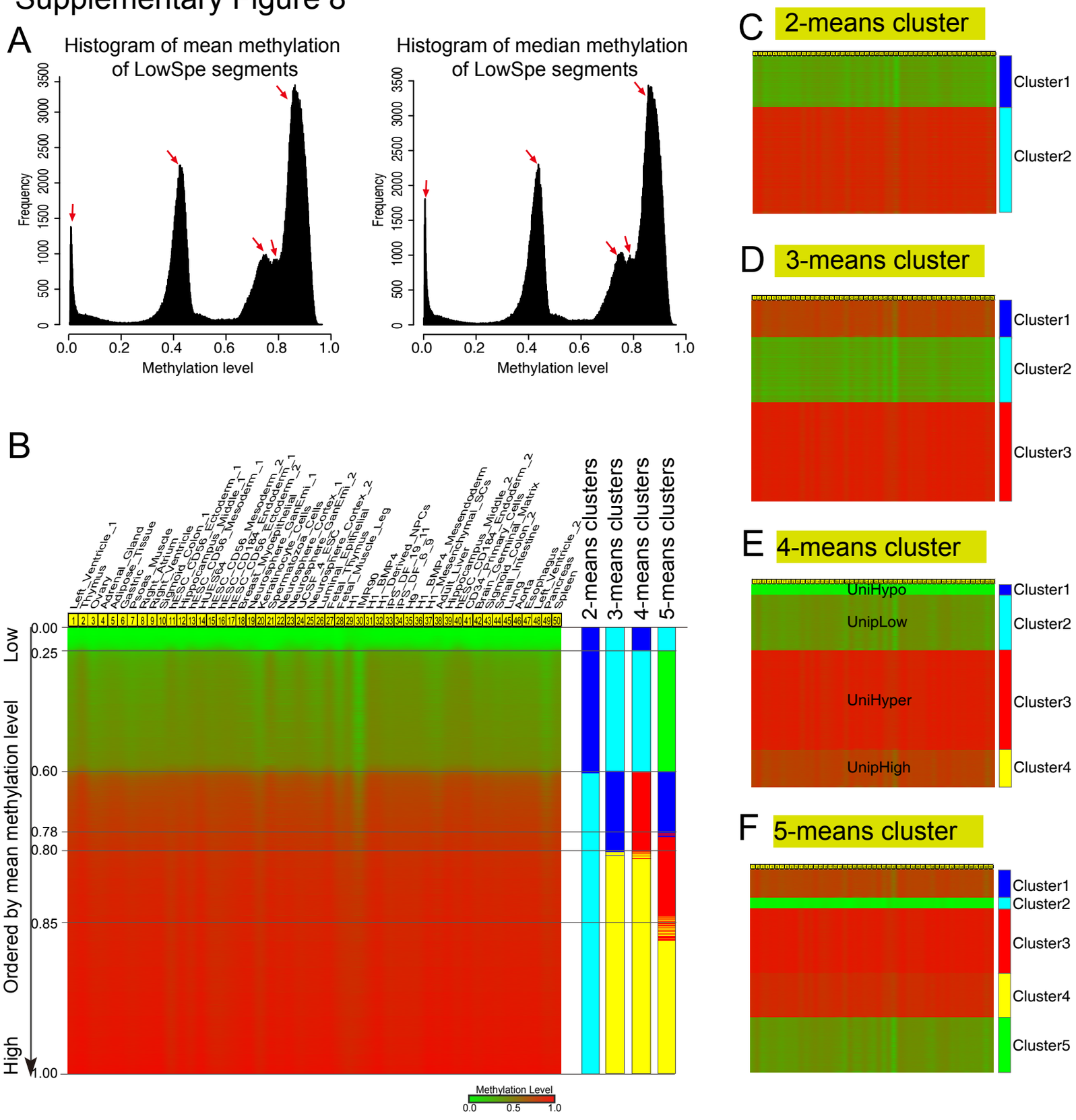
Supplementary Figure 6: The longest segment identified by SMART in the human genome. The longest segment was located at chr10:39103301-39130657, covers 27k bases, includes 449 CpGs, and is part of partially methylated domains (PMDs) identified in IMR90 by Lister et al. 2009.

Supplementary Figure 7



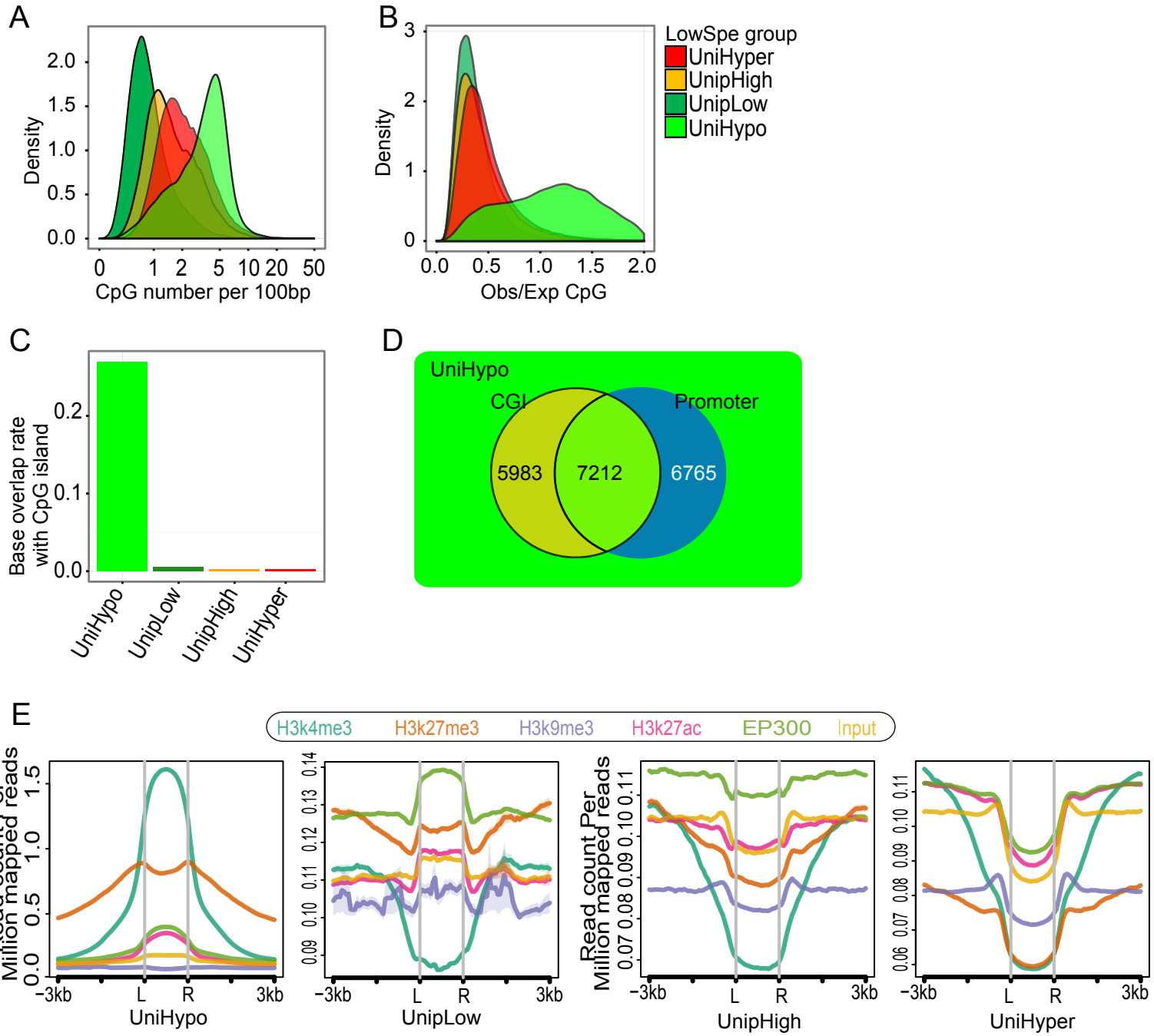
Supplementary Figure 7: The features of human methylation segments identified by SMART. (A) The length of three types of segments. The length of segments ranges from 20 bp to 10 kb, including 5406 segments with a length of at least 3.5 kb. (B) The CpG number of three types of segments. The CpG number in these segments ranges from 5 to 1000, including 288 segments that have more than 150 CpGs. (C) Distribution of mean methylation of HighSpe, InterSpe and LowSpe segments. (D) Distribution of median methylation of HighSpe, InterSpe and LowSpe segments. (E) Distribution of methylation specificity of HighSpe, InterSpe and LowSpe segments. (F) Distribution of Obs/Exp CpG of HighSpe, InterSpe and LowSpe segments. (G) Distribution of Obs/Exp CpG of CpG island, shore and desert segments. (H) Density scatterplot of Obs/Exp CpG and methylation specificity of all segments. (I) Density scatterplot of Obs/Exp CpG and mean methylation of all segments.

Supplementary Figure 8



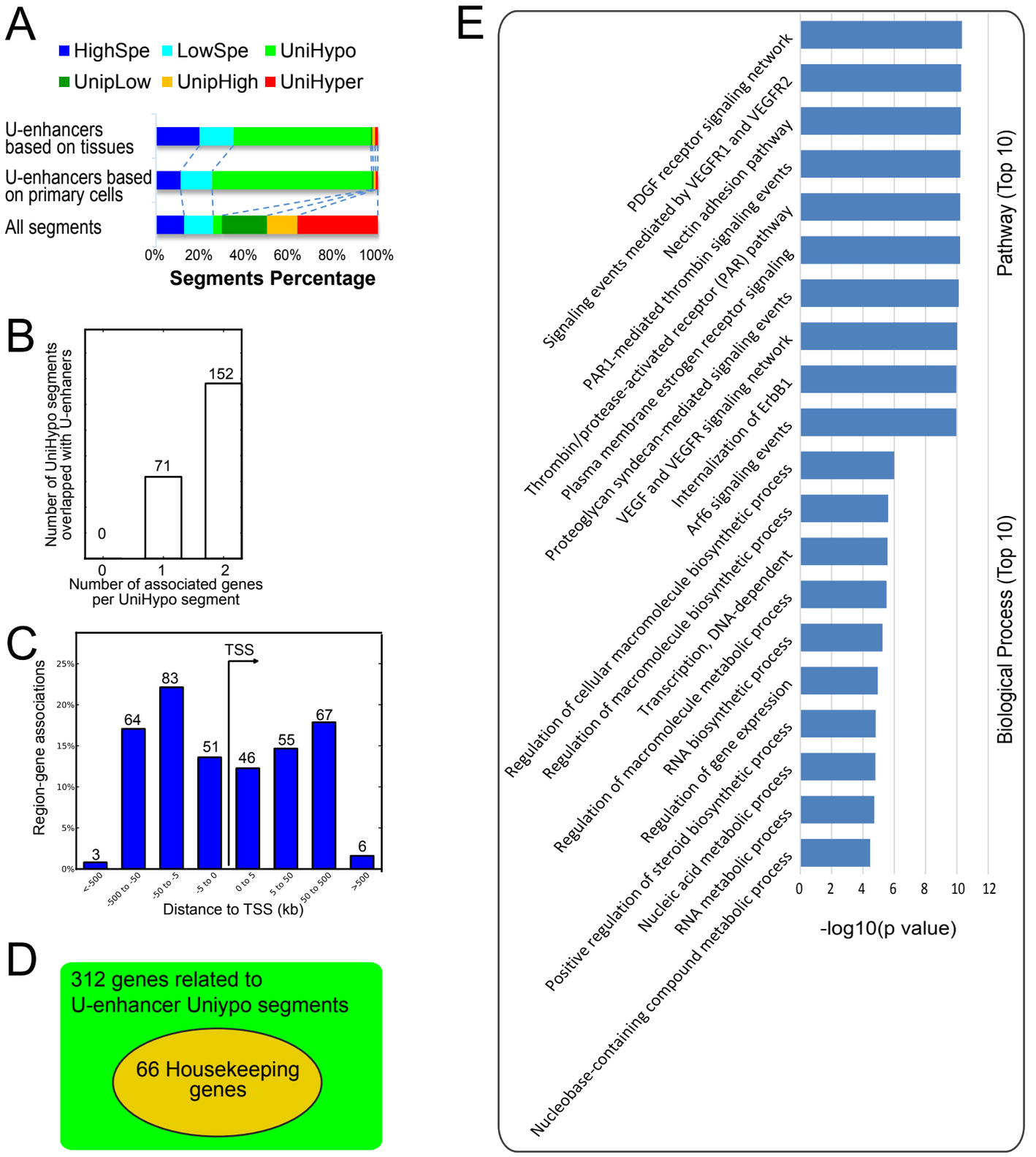
Supplementary Figure 8: Four classifications of LowSpe segments based on mean/median methylation. (A) Histogram of mean/median methylation across 50 samples of LowSpe segments. The distribution of methylation levels of these segments showed five peaks. Two peaks approximately 0.75 are close to each other and are smaller than other three. The methylation difference between these two peaks is approximately 0.05, which is usually regarded as meaningless in methylation analysis, thus we treated two peaks as the same methylation state: partial-high-methylation. (B) Heat map of methylation levels of 562,719 LowSpe segments in 50 methylomes. Each row represents a segment. The segments are ordered by their mean methylation levels in 50 samples from low to high. Seven methylation values were given in the right panel. For each segment, its cluster classification in K means (K=2, 3, 4, and 5) clustering shown in Figures C-F is given in the right panel. (C-F) The K means (K=2, 3, 4, and 5) clustering based on the 50 methylomes of LowSpe segments. The segments in Cluster1 and those in Cluster2 by 4-means clustering showed large methylation changes, but they were segmented into a cluster by the 3-means clustering. In addition, the greatest methylation difference among the segments in Cluster4 by 5-means clustering is only 0.07, which is usually regarded as meaningless in DNA methylation analysis. These results suggest the justifiability of four clusters, including UniHypo (0.00~0.25), UnipLow (0.25~0.60), UniphHigh (0.60~0.80) and UniHyper (0.80~1.00).

Supplementary Figure 9



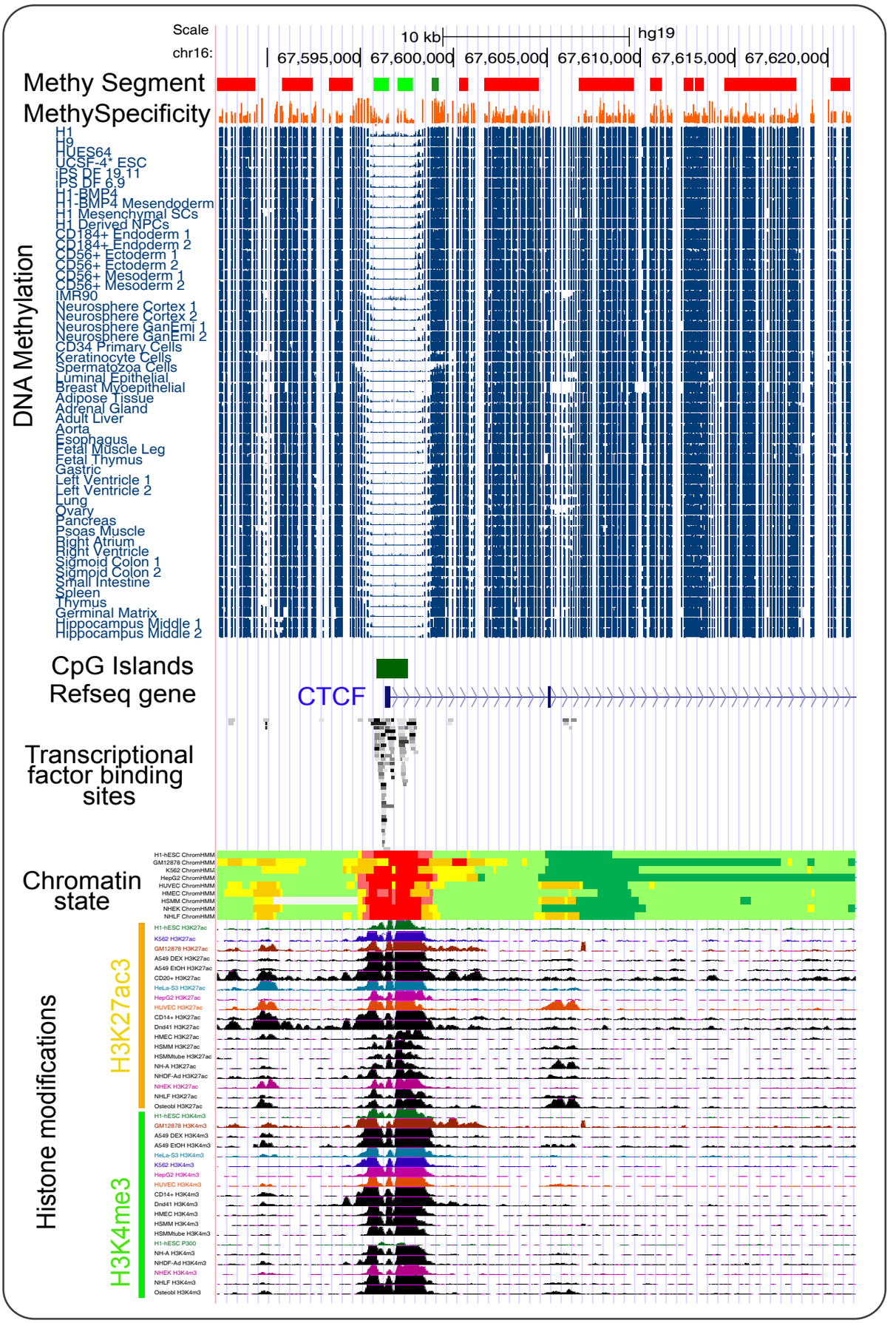
Supplementary Figure 9: Features of LowSpe segments. (A) Distribution of CpG number per 100 bp of segments in four groups of LowSpe segments including UniHypo, UnipLow, UnipHigh and UniHyper. (B) Distribution of Obs/Exp CpG in four groups of LowSpe segments. UniHypo segments showed higher CpG density, which was a typical feature of CpG islands (CGIs). (C) The base overlap rate between CpG islands and four groups of LowSpe segments. UniHypo segments showed higher overlap rate than other groups of LowSpe segments. (D) Overlap between CGIs UniHypo segments and promoter UniHypo segments. More than 7,000 UniHypo segments were overlapped with promoter CGIs. In addition, we found 88% of LowSpe segments that were located in promoters of housekeeping genes were significantly overlapped with uniformly hypomethylated CGIs (p<10-282, Chi-square test). (E) Chromatin modification patterns of LowSpe segments. The chromatin modifications (H3K4me3, H3K27me3, H3K9me3, H2K27ac, EP300) of the segments in each LowSpe group in H1 cell line. Average enrichment profiles of log2 ratios of several histone marks and transcription factor vs. DNA input around ±3 Kb regions of different types of LowSpe segments. “L” and “R” represent the boundary of HypoMark/HyperMark. Ngs.plot (http://code.-google.com/p/ngsplot/) was used to visualize the average profiles and heat maps with fragment length equal to 300 bp and other default parameters.

Supplementary Figure 10



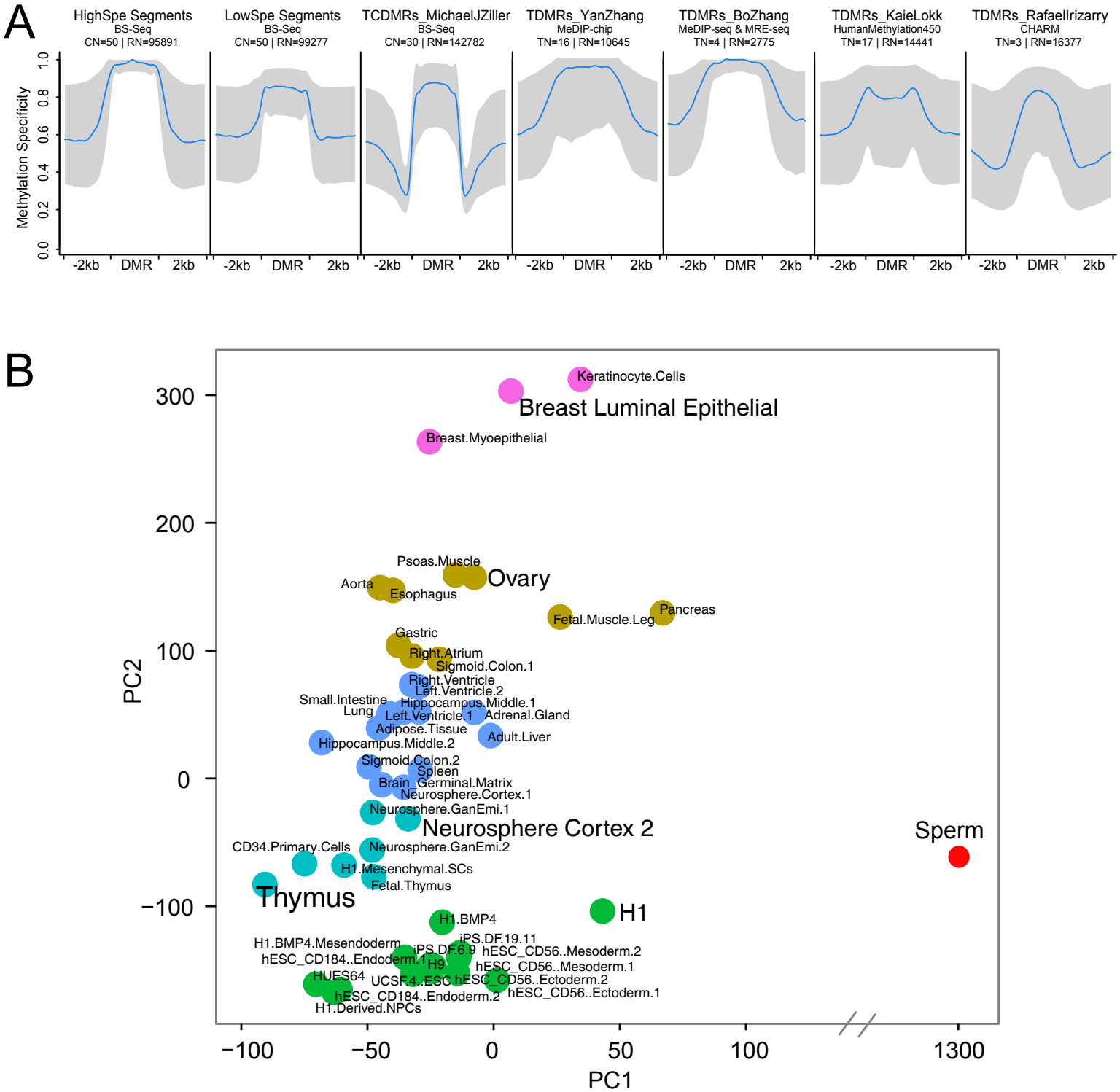
Supplementary Figure 10: UniHypo segments and ubiquitous enhancers. (A) Overlap between different types of segments and ubiquitous enhancers. It was revealed that ubiquitous enhancers were prone to overlap with UniHypo segments ($p<10^{-10}$, Chi-square test). (B) Number of associated genes per UniHypo segment overlapped with U-enhancer. (C) Genome location of UniHypo segment overlapped with U-enhancer relative to transcription start site (TSS) of an associated gene. (D) 312 genes related to UniHypo segment overlapped with U-enhancer. Among these genes, 66 genes have been reported as housekeeping genes, including the well-known CTCF. (E) Functional enrichment analysis of genes related to 223 UniHypo segments overlapped with ubiquitous enhancers. Top 10 biological processes and top 10 pathways were shown. It was revealed these genes were enriched in functional terms involving fundamental biological processes (such as macromolecule biosynthesis) and metabolic pathways (such as the mTOR signaling pathway).

Supplementary Figure 11



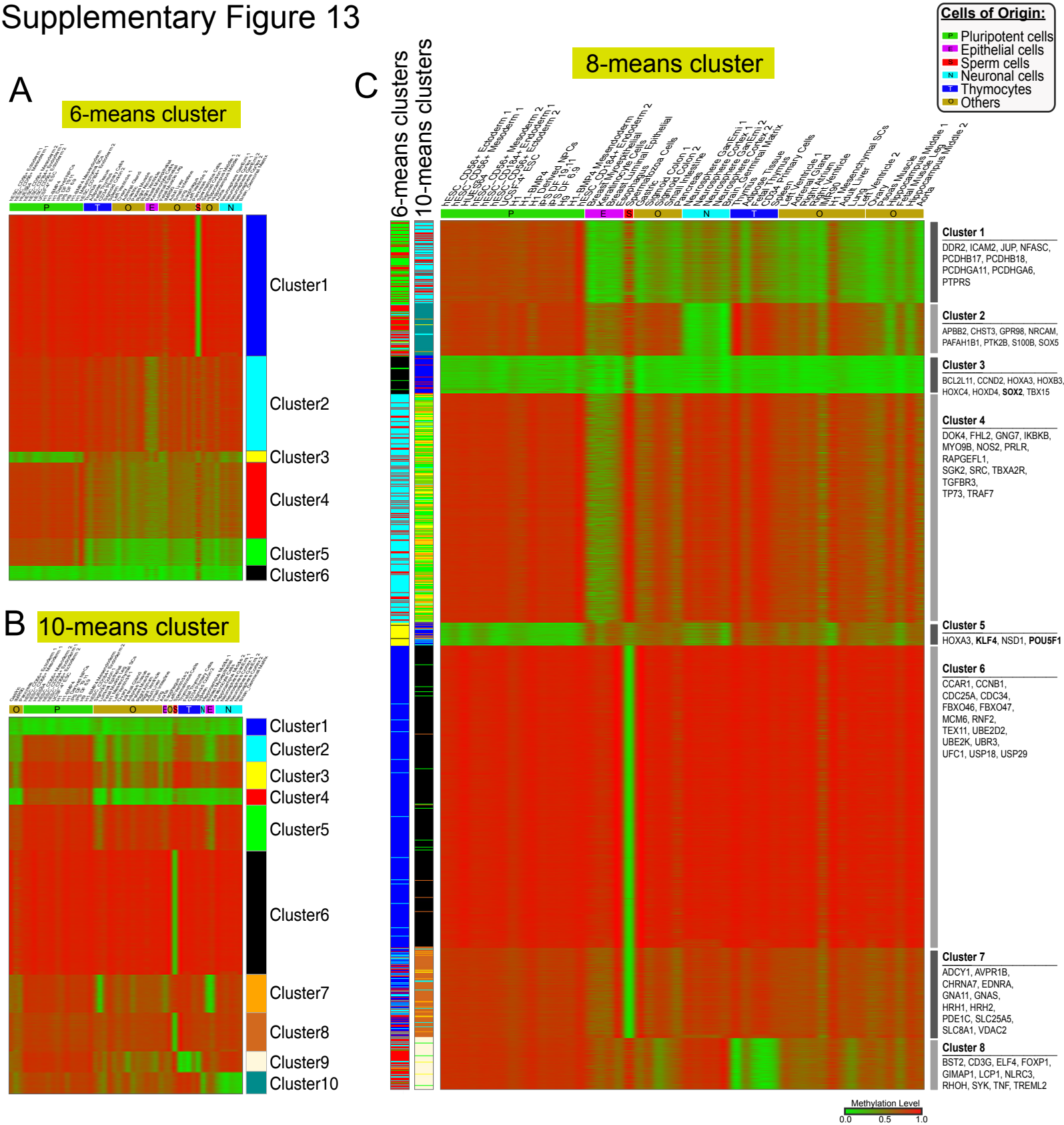
Supplementary Figure 11: The methylation pattern and chromatin states of the promoter regions of CTCF. CTCF encodes a transcriptional regulator protein with 11 highly conserved zinc finger (ZF) domains and owns two UniHypo segments. Two UniHypo segments were overlapped with multiple active features including a CGI, ubiquitous enhancer and transcriptional factor binding sites, an active chromatin state (promoter-associated state represented by red color), and active histone modifications (H3K4me3 and H3K27ac) in its promoter region.

Supplementary Figure 12



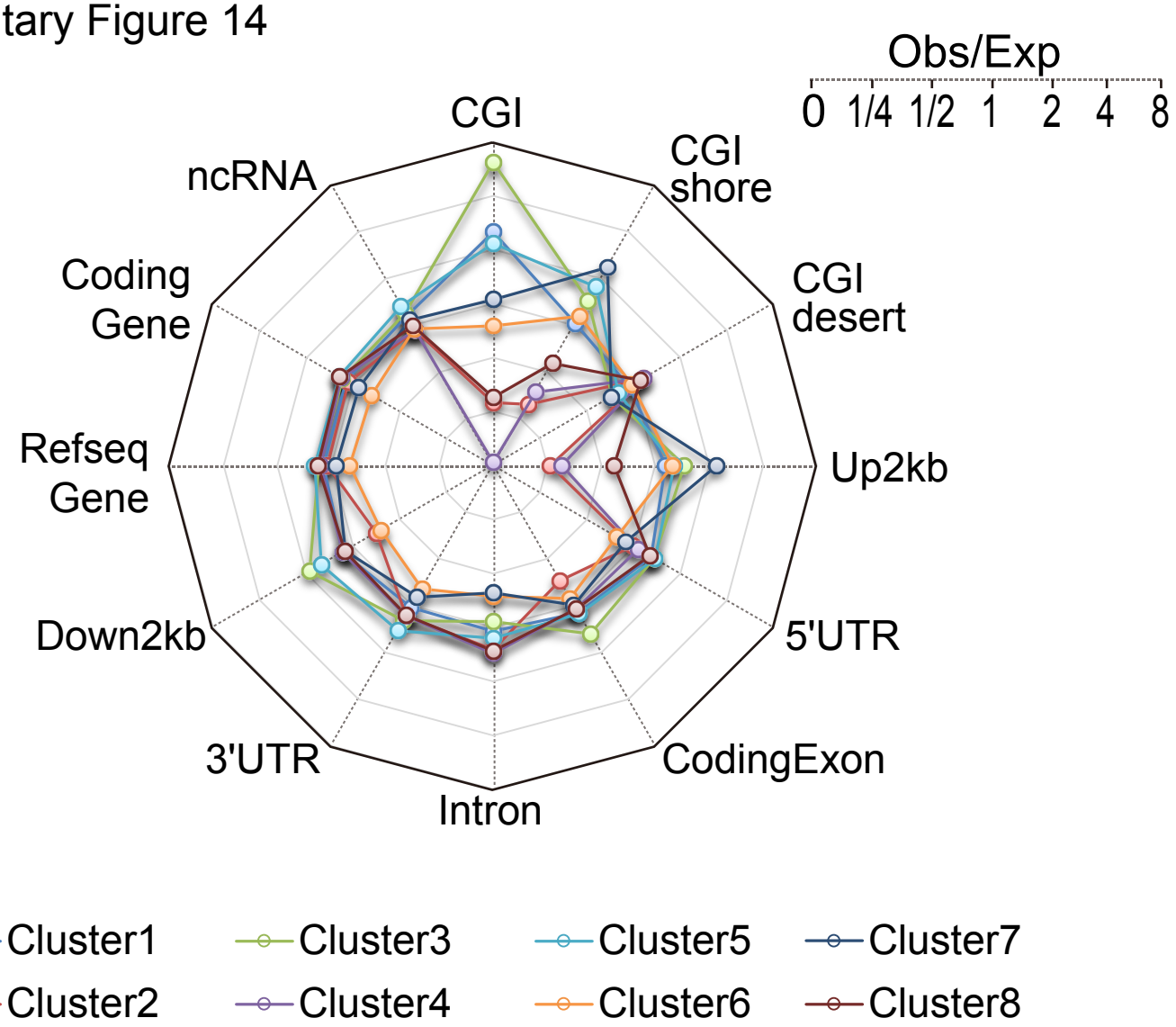
Supplementary Figure 12: HighSpe segments identified by SMART in the human genome. (A) Composite plot of methylation specificity across HighSpe and InterSpe segments and differentially methylated regions (DMRs) across human tissues/cells that were identified by previous studies based on the methylomes profiled by different technologies including BS-Seq, MeDIP-chip/seq, HumanMethylation450 and CHARM (7-11). The methylation specificity near HighSpe and InterSpe segments revealed a pattern of high methylation specificity in the body of HighSpe and InterSpe segments and low specificity in their flanking sequences. The DMRs across human tissues/cells that were identified by previous studies showed similar results with our study, confirming the accuracy of the methylation specificity quantified by SMART and the reliability of methylation segments identified in this study. (B) The principal component analysis of 50 methylomes. This figure revealed the specific methylation pattern in sperm and the clustering of pluripotent cell lines.

Supplementary Figure 13



Supplementary Figure 13: K-means clustering of HighSpe segments. In each panel, methylation levels are represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of 50 samples that were classified into six main groups tagged by different colors and abbreviations: Pluripotent cells (P), Epithelial cells (E), Sperm cells (S), Neuronal cells (N), Thymocytes (T) and Others (O). (A) 6-means clustering of HighSpe segments. Six clusters of segments are differentially colored on the right. (B) 10-means clustering of HighSpe segments. Ten clusters of segments are differentially colored on the right. (C) A larger version of 8-means clustering of HighSpe segments shown in Figure 1H. On the left, the cluster of each segment in 6-means clustering and 10-means clustering are given as the cluster color defined in A and B. On the right, eight clusters are given, and examples of the related genes for each cluster are also listed.

Supplementary Figure 14

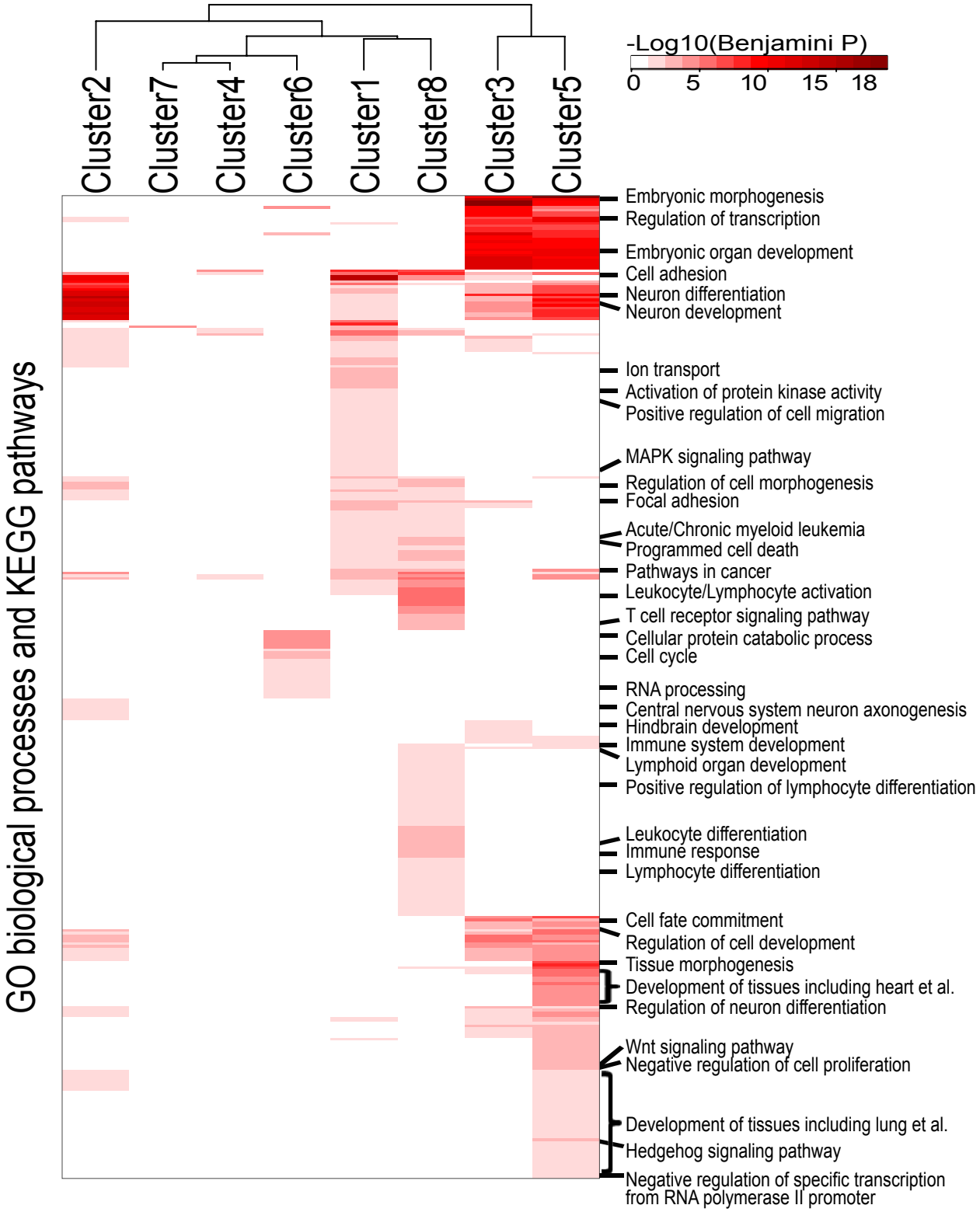


Supplementary Figure 14: Genome location of different clusters of HighSpe segments. Radar plots showing the ratio of observed to expected HighSpe segments in different clusters and genome features including CpG islands (CGI), CpG island shores (CGIshore) and Refseq genes related seven categories including upstream 2 kb of transcription start site (Up2kb), 5'UTR, Coding Exon (CodingExon), Intron, 3'UTR, downstream 2 kb of transcription end site (Down2kb), and noncoding RNAs (ncRNAs). To examine whether the HighSpe segments in specific clusters are enriched in some specific genome features, we calculated the number of HighSpe segments in each cluster (Cluster HighSpe Num.), the number of HighSpe segments in each genome feature (Feature HighSpe Num.), the number of overlapped HighSpe segments between Cluster HighSpe and Feature HighSpe (Cluster&Feature HighSpe Num.), and the number of total HighSpe segments identified (Total HighSpe Num.). The ratio of observed to expected HighSpe segments (Obs/Exp HighSpe) in each cluster and genome feature was calculated as

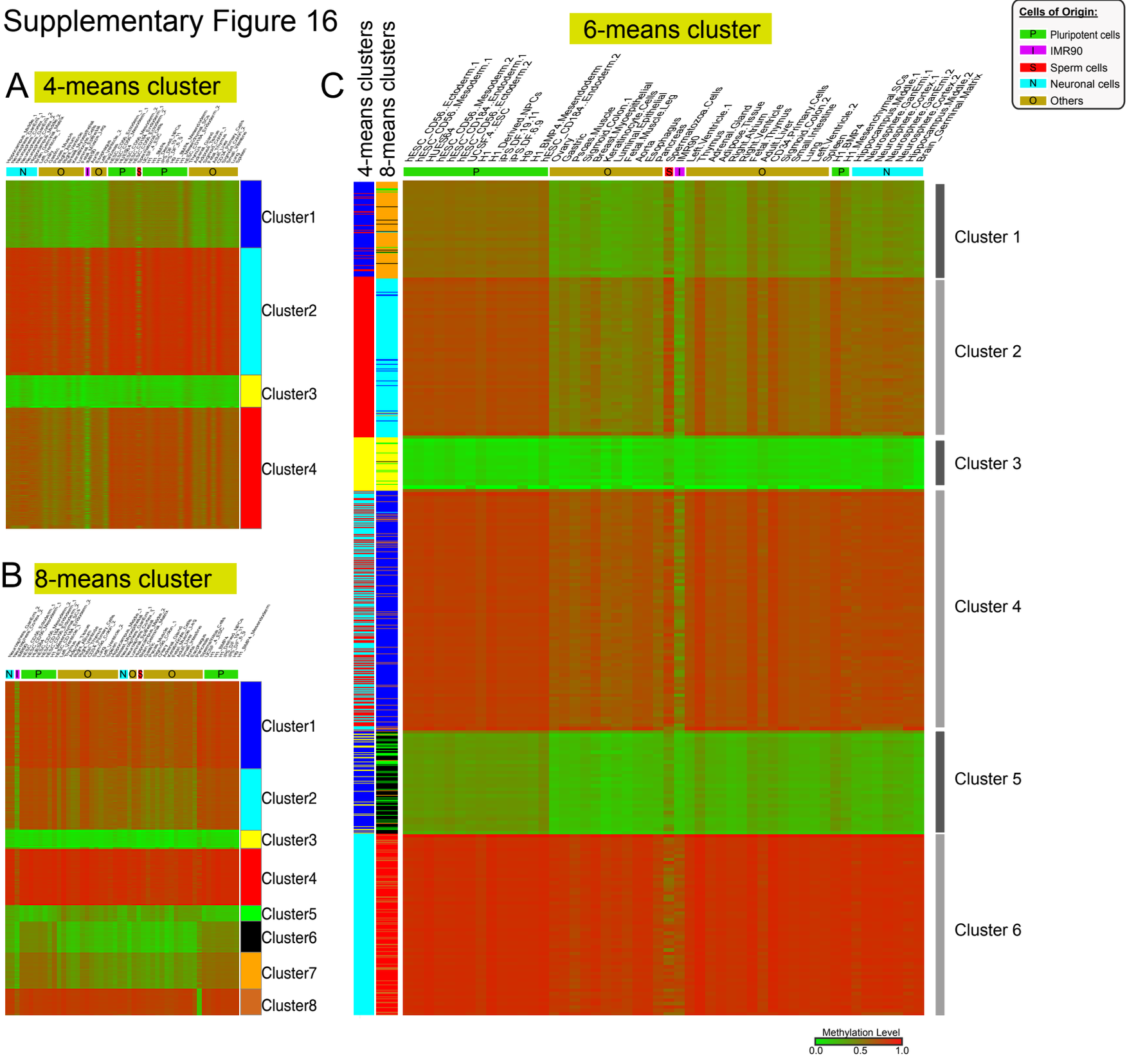
$$Obs/Exp\ HighSpe = \frac{(Cluster\ \&\ Feature\ HighSpe\ Num.) \times (\text{Total}\ HighSpe\ Num.)}{(Cluster\ HighSpe\ Num.) \times (Feature\ HighSpe\ Num.)}$$

The center of the plot was 0, and a colored dot on the respective axis indicates the Obs/Exp HighSpe of the HighSpe from specific cluster (colored line) in a specific genome feature (angle). It was obvious that the HighSpe segments in Clusters 1, 3 and 5 exhibit high Obs/Exp HighSpe in CGI, suggesting potential roles of methylation dynamics in CGIs in cell type identity. In addition, the HighSpe segments in Cluster 7 show enrichment in CGI shores and Up2kb.

Supplementary Figure 15

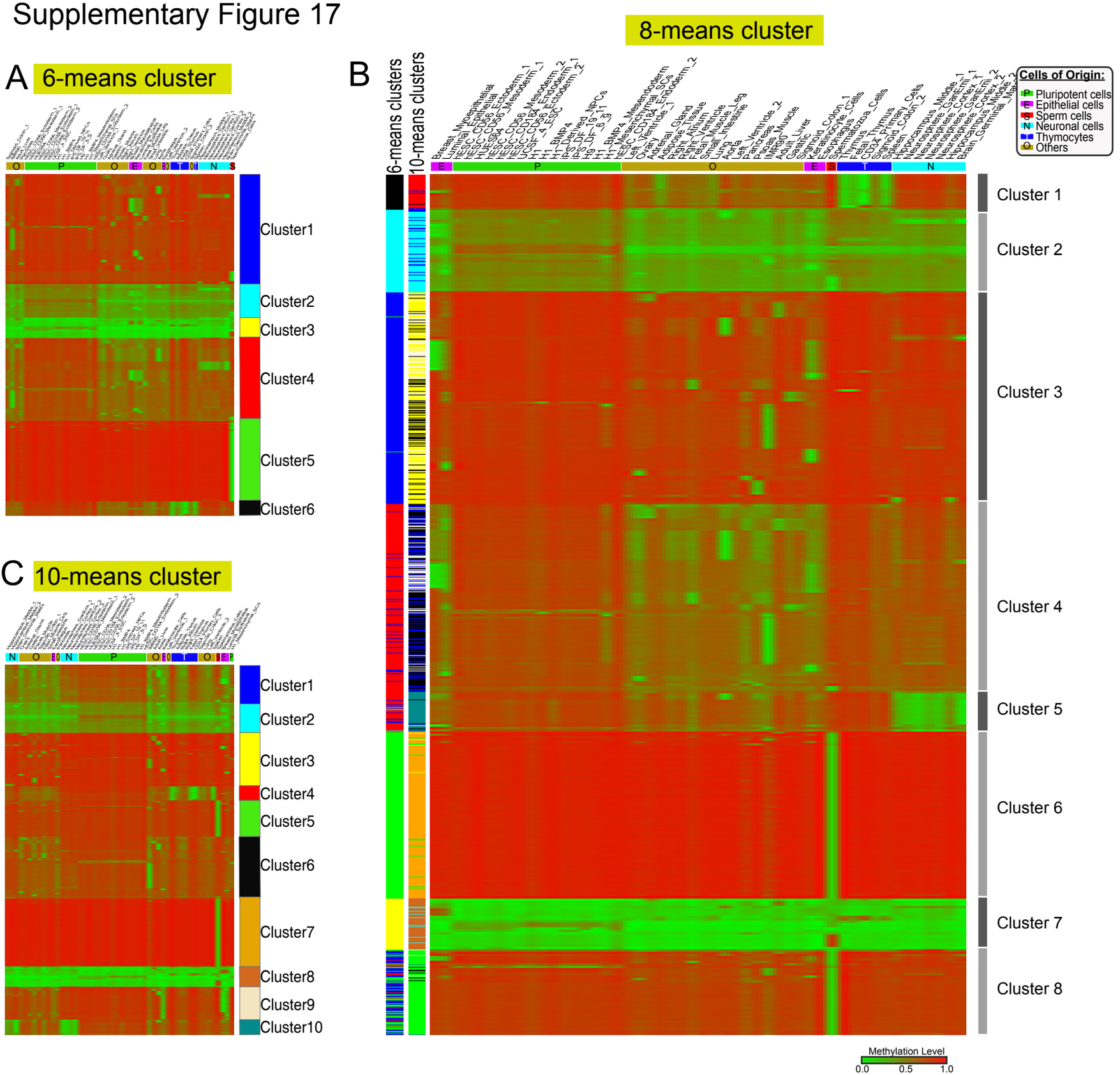


Supplementary Figure 15: The enriched functions of genes related to HighSpe segments in each cluster. On the right, the representative and significant biological processes or KEGG pathways are shown. For the functional analysis of genes related to each cluster of HighSpe segments, the genes related to HighSpe segments with length ≥200 bp in each cluster were selected. Due to the limitation of gene number in DAVID, we adopted more stringent standards for selection of genes in cluster 4 and cluster 6, both of which were related to more than 3,000 genes. For cluster 4, only genes with promoter HighSpe segments with length ≥200 bp were selected, and for cluster 6, only genes with HighSpe segments with length ≥200 bp in the regions from upstream 2 kb to transcription start site (TSS) were selected. Then, the selected genes in each cluster were imported into DAVID to perform functional enrichment analysis of these genes in biological process and the KEGG pathway. Finally, 371 enriched function terms were clustered and visualized by R.



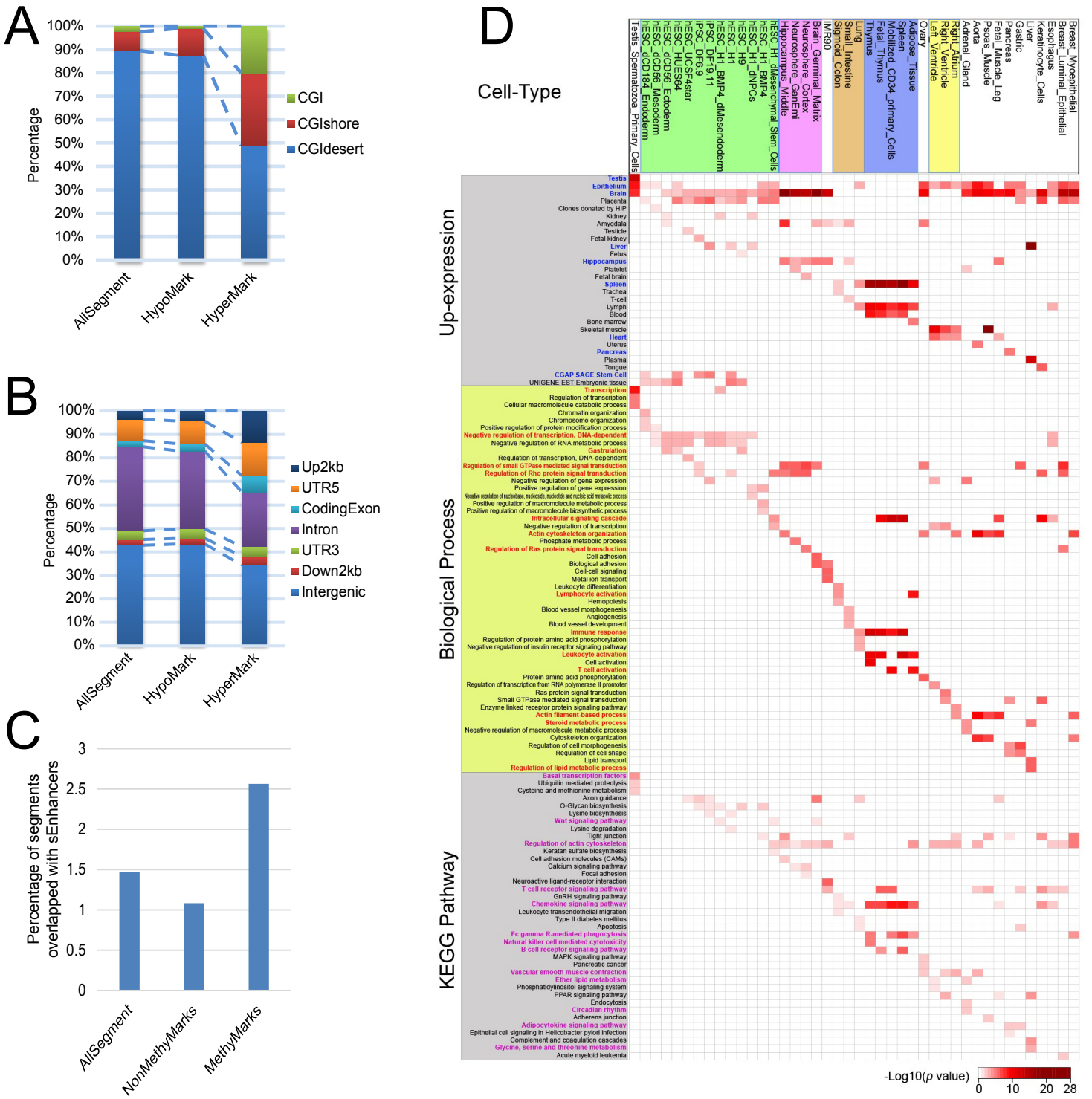
Supplementary Figure 16: K-means clustering of InterSpe segments. In each panel, methylation levels were represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of the 50 samples that were classified into five main groups tagged by different color and abbreviation: Pluripotent cells (P), IMR90 (I), Sperm cells (S), Neuro cells (N) and Others (O). (A) 4-means clustering of InterSpe segments. Four clusters of segments are differentially colored on the right. (B) 8-means clustering of InterSpe segments. Eight clusters of segments are differentially colored on the right. (C) 6-means clustering of InterSpe segments. On the left, the cluster of each segments in 4-means clustering and 8-means clustering are given as the cluster color defined in A and B. On the right, six clusters are listed.

Supplementary Figure 17



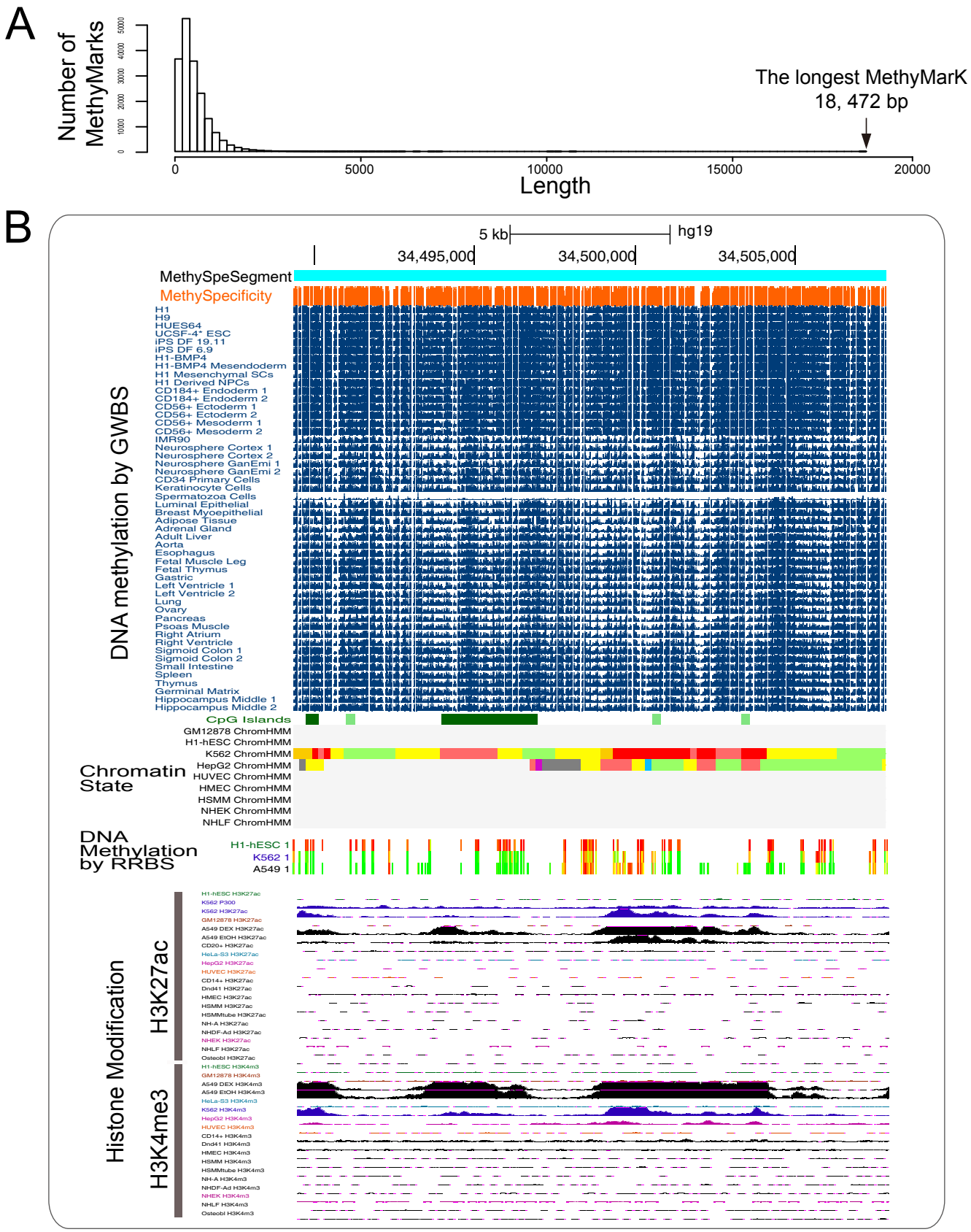
Supplementary Figure 17: K-means clustering of MethyMarks. In each panel, methylation levels were represented by a gradient from green (unmethylation) to red (full methylation). Each column represents one of the 50 samples that were classified into six main groups tagged by different color and abbreviation: Pluripotent cells (P), Epithelial cells (E), Sperm cells (S), Neuro cells (N), Thymocytes (T) and Others (O). (A) 6-means clustering of MethyMarks. Six clusters of MethyMarks are differentially colored on the right. (B) 10-means clustering of MethyMarks. Ten clusters of MethyMarks are differentially colored on the right. (C) 8-means clustering of MethyMarks. On the left, the cluster of each MethyMark in 6-means clustering and 10-means clustering are given as the cluster color defined in A and B.

Supplementary Figure 18



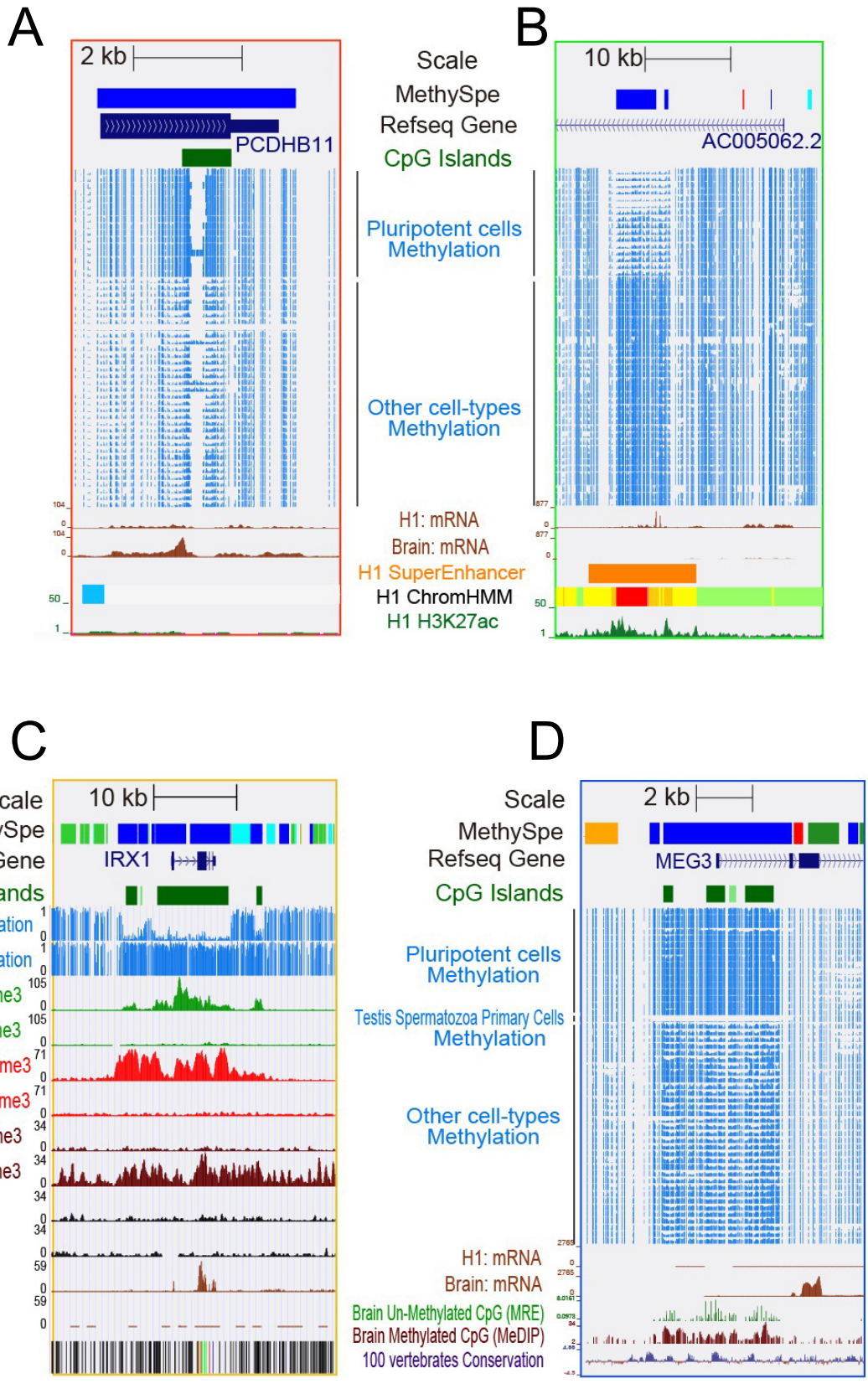
Supplementary Figure 18: Genome location and functional enrichment of cell type-specific Methy-Marks. (A) Percentage of HypoMarks and HyperMarks overlapped with CpG islands (CGI), CGI shores and CGI deserts. (B) Percentage of HypoMarks and HyperMarks overlapped with different genome features including Up2kb, 5'UTR, CodingExon, Intron, 3'UTR, Down2kb and Intergenic. (C) Percentage of Methy-Marks overlapped with cell type-specific active enhancer. (D) High expression and functional enrichment of HypoMark genes. For the functional analysis of genes related to cell-specific HypoMarks, the genes with promoter HypoMarks in each cell type were selected. For testis spermatozoa primary cells, only the genes with promoter HypoMarks with length ≥700 bp were selected. Then, the selected genes in each cell type were imported into DAVID to perform functional enrichment analysis in over-expressed tissue, biological process and the KEGG pathway. For each type of analysis, the three most significant terms were selected and visualized by R. The grids colored from white (0) to dark red (28) represent the –log10 of the p value for the enrichment of HypoMark genes in each cell type (Column) in each function term (Row). The function terms that were related to the cell types in this study were bolded and colored blue (over-expression), red (biological processes) and purple (KEGG pathway).

Supplementary Figure 19



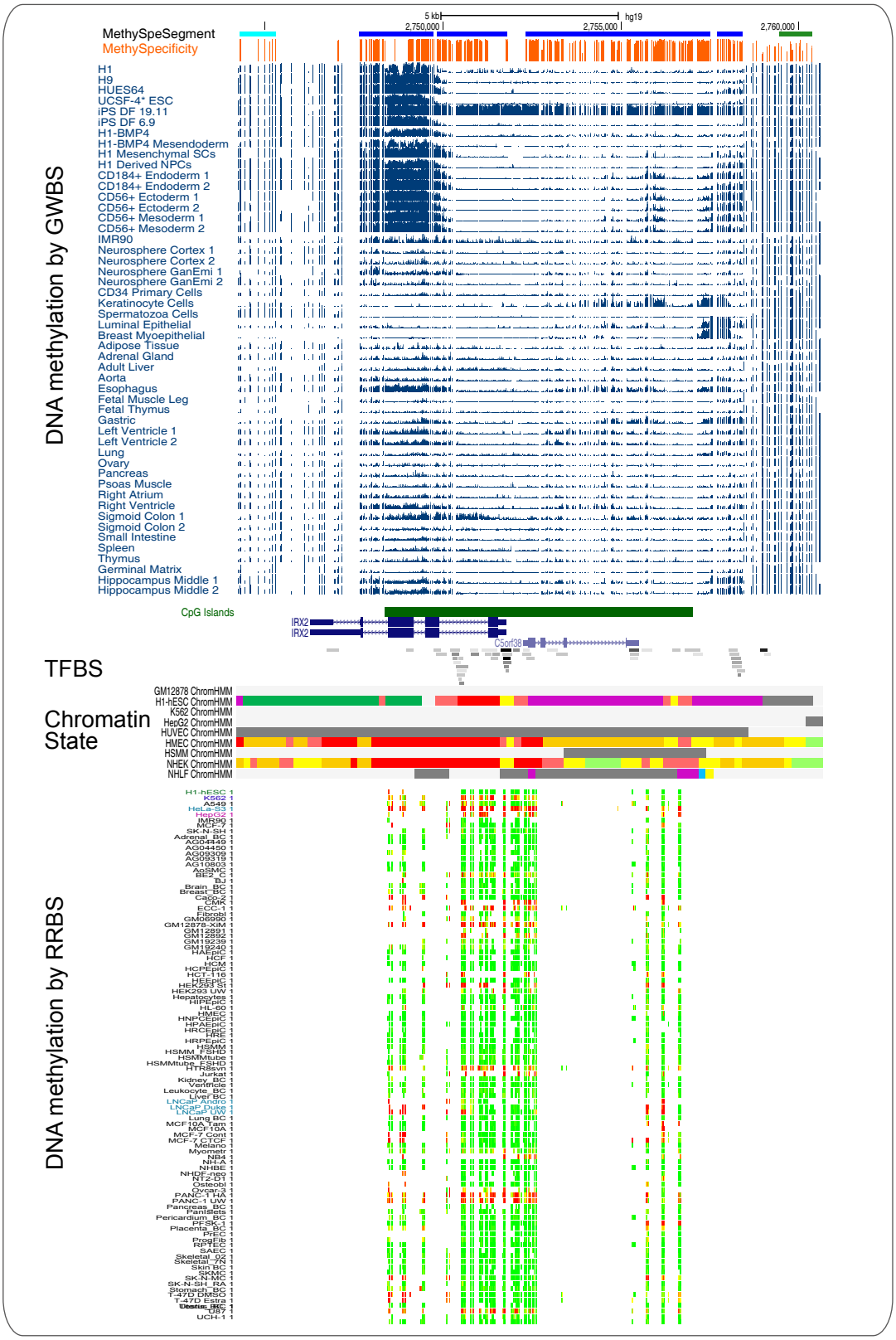
Supplementary Figure 19: Cell type-specific MethyMarks that span large chromosomal regions. (A) The length of cell type-specific MethyMarks identified by SMART. (B) Genome location and epigenomic features of the longest MethyMarks identified. Each methylation track represents a cell type, and the height of the bar represents the methylation level. Super-enhancer, chromatin states (the bar colored in blue was for insulator, red for active promoter, orange for strong enhancer, yellow for weak enhancer, and light green for weak transcribed), and H3K27ac (the height of bar represents the number of reads overlapping each 25 bp bin) are shown at the bottom.

Supplementary Figure 20



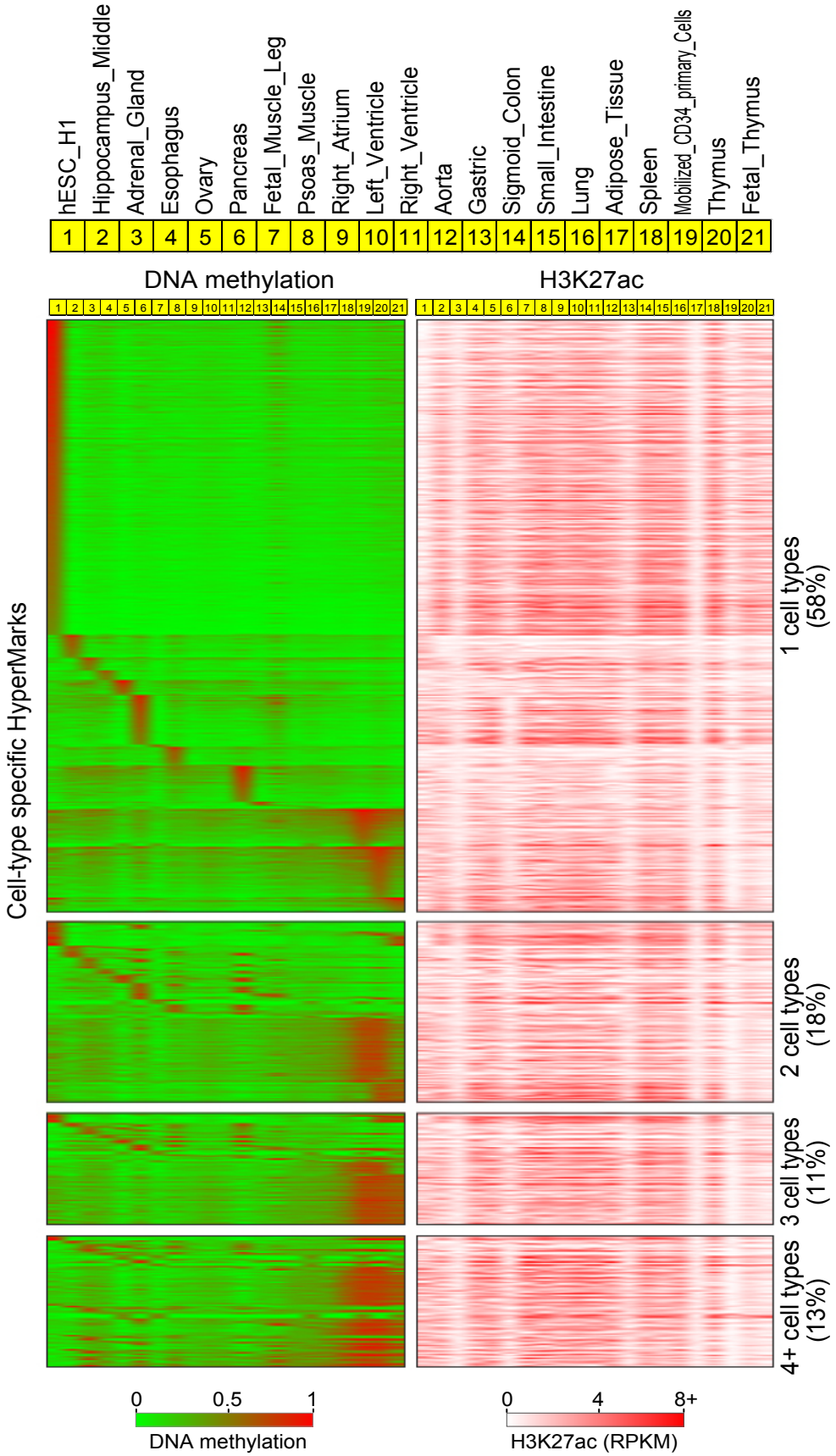
Supplementary Figure 20: Example genes related to cell type-specific MethyMarks that span large chromosomal regions. (A) An example for large HyperMark genes in pluripotent cells. This HyperMark was specifically hypermethylated in pluripotent cells and overlapped with *PCHHB11*. (B) An example for large HypoMark genes in pluripotent cells. This HypoMark was specifically hypomethylated in pluripotent cells and overlapped with *Ac005062.2*. (C) iPSC cells DF 19.11 showed a different methylation pattern compared to H1 hESC cells in the MethyMarks, which overlapped with a large CpG island and *IRX1*. The histone modifications and mRNA were obtained from http://www.genboree.org/EpigenomeAtlasBrowser/. (D) The DNA methylation and expression pattern of imprinted gene, *MEG3*. MRE and MeDIP tracks represent the un-methylated and methylated CpGs in the brain, respectively. The 100 vertebrates' basewise conservation by PhyloP is shown in the bottom track.

Supplementary Figure 21



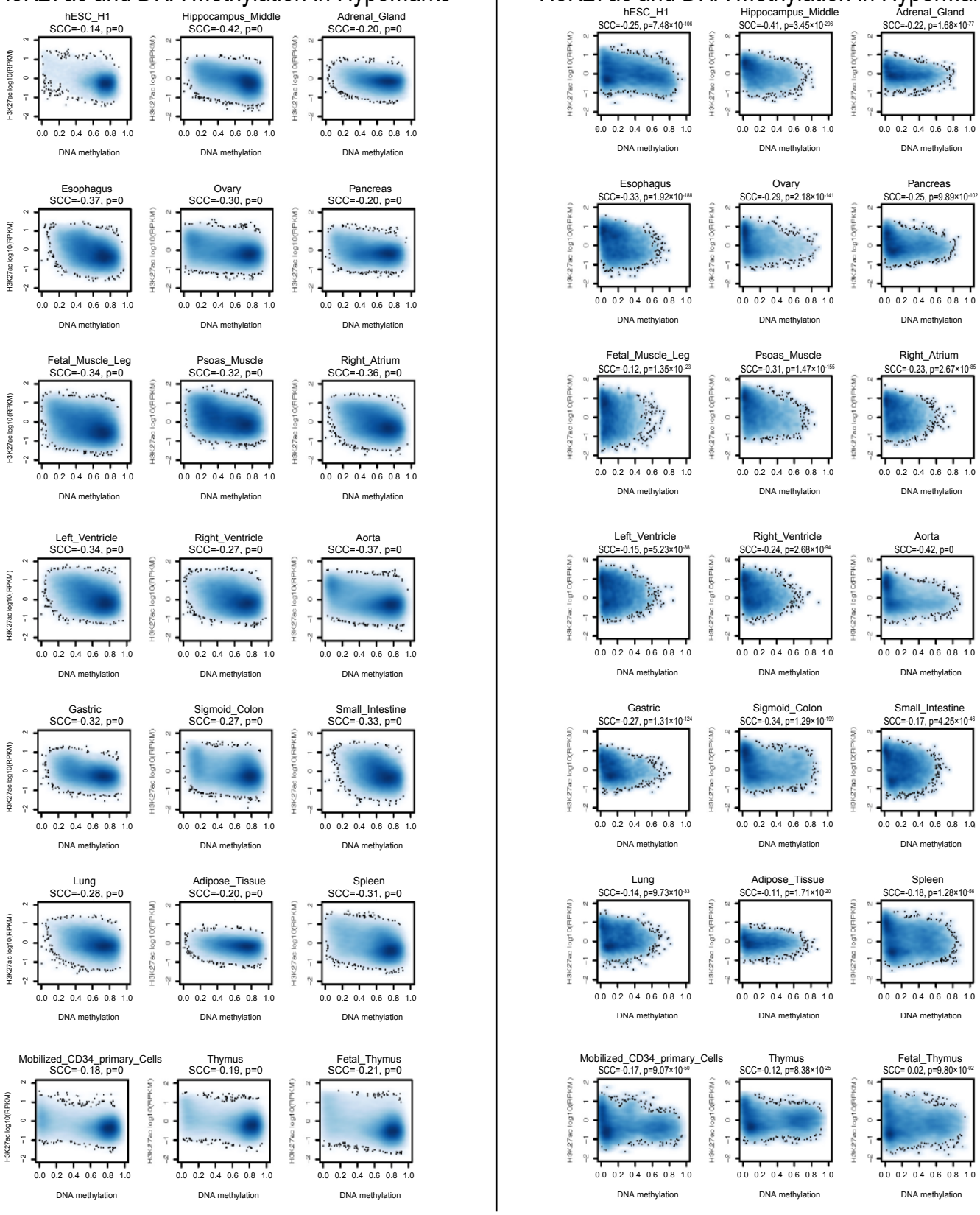
Supplementary Figure 21: Detailed epigenetic modifications in the iPSC-specific HyperMarks related to *IRX2* across cell types. Each whole-genome bisulfite sequencing (WGBS) methylation track shows the methylation of a cell type, and the height of the bar represents the methylation level. The histone modifications, chromatin states and methylation level by reduced representation bisulfite sequencing (RRBS) in various samples including cancer were shown. Each RRBS methylation track represents a cell type, and the color of bar represents the methylation level from unmethylated (green) to full methylation (red). This HyperMark showed specific hypermethylation in iPSC cells DF 19.11 but hypomethylation in other cell types. As shown by RRBS methylation, the aberrant hypermethylation of this mark may cause the deactivation of *IRX2* in cancer. For instance, this mark showed specific hypermethylation levels and inactive chromatin states in human cancer cell lines, including K562 and HepG2.

Supplementary Figure 22



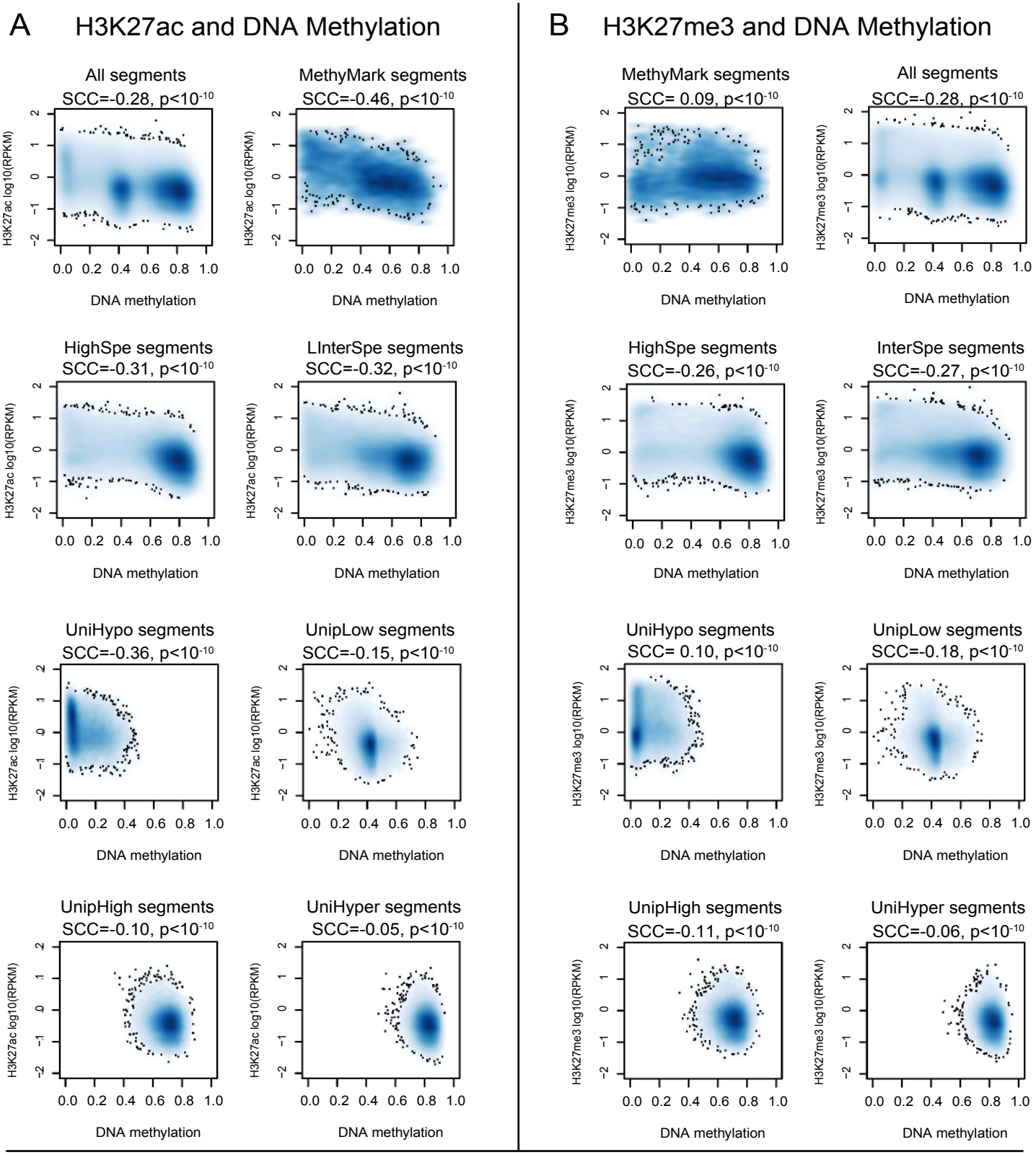
Supplementary Figure 22: Heatmap of DNA methylation and H3K27ac in cell type-specific HyperMarks. Each row denotes a HyperMark, and each column a cell type. DNA methylation levels are represented by a gradient from green (ummethylated) to red (full methylation) and H3K27ac from white (lowest) to red (highest). The density of H3K27ac was represented by the read count per million mapped reads (RPKM).

Supplementary Figure 23



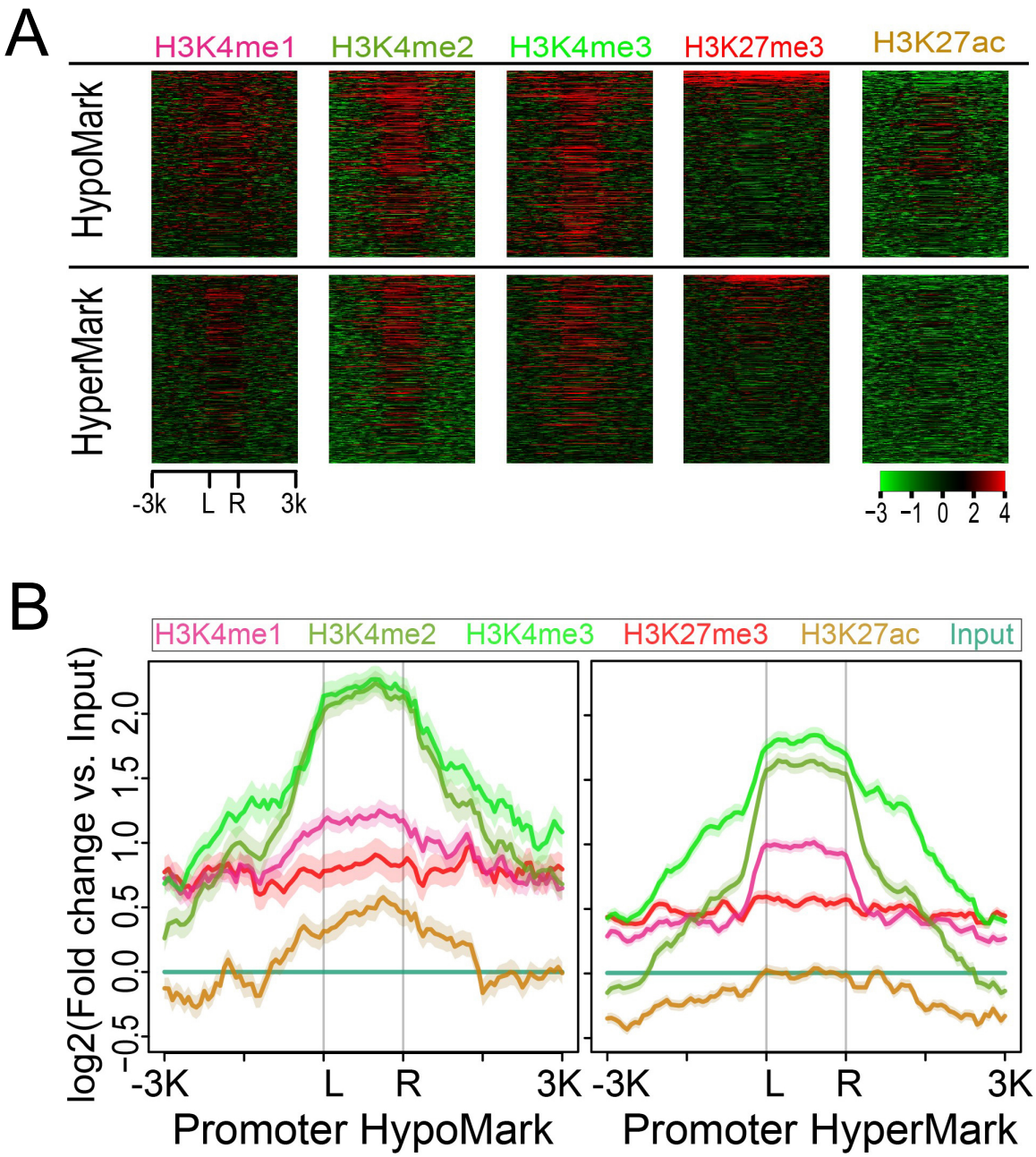
Supplementary Figure 23: Correlation between DNA methylation and H3K27ac in HypoMarks and HyperMarks of various cell types. (A) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in HypoMarks of a cell type. The Spearman's rank correlation coefficient (SCC) between DNA methylation and H3K27ac in HypoMarks was calculated for each cell type, respectively. P represents the significance of the coefficient. (B) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in HyperMarks of a cell type. The SCC between DNA methylation and H3K27ac in HyperMarks was calculated for each cell type, respectively. P represents the significance of the coefficient.

Supplementary Figure 24



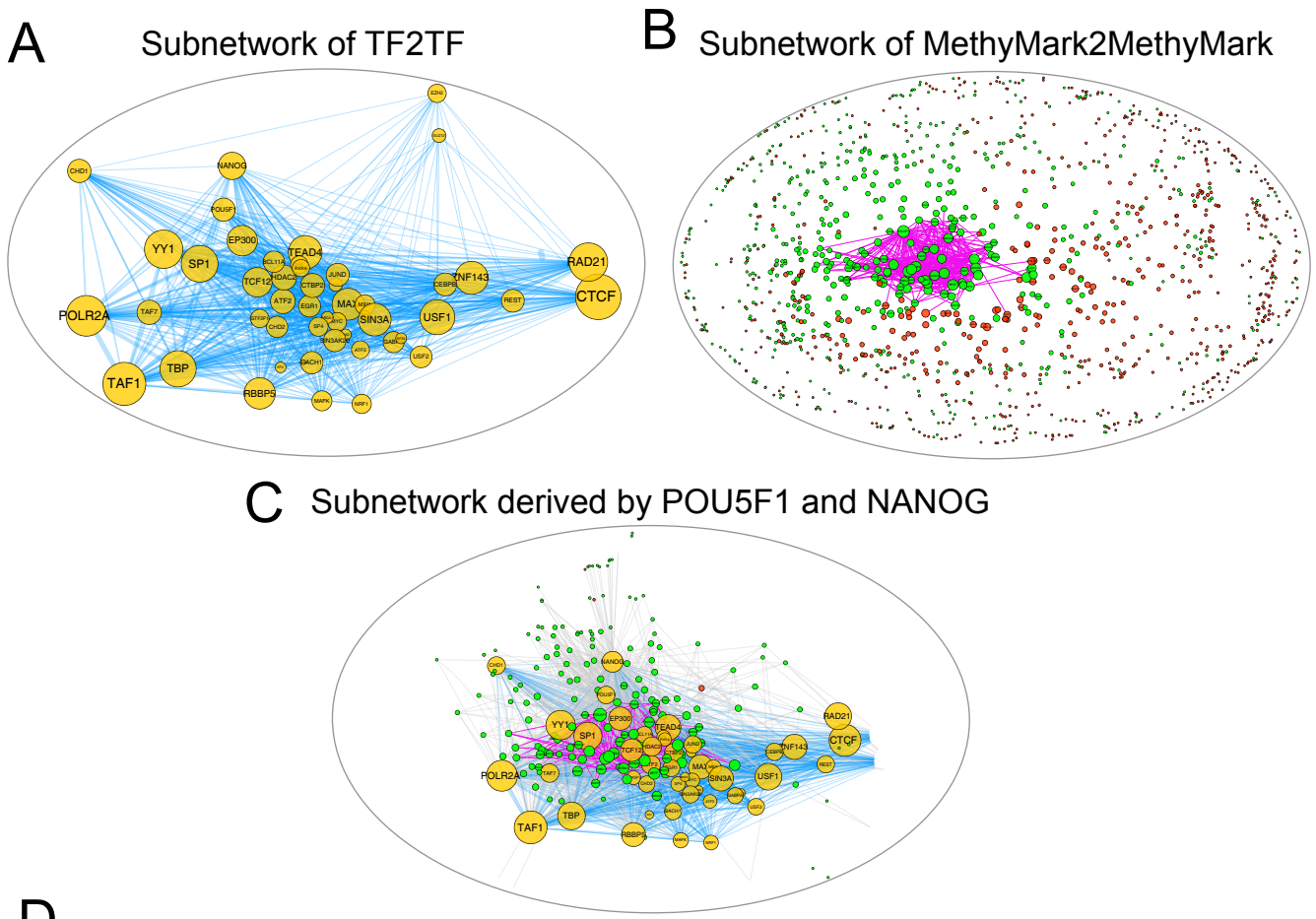
Supplementary Figure 24: Correlation between DNA methylation and two histone marks (H3K27ac and H3K27me3) in different categories of segments. (A) Each sub-figure shows the density scatterplot of DNA methylation and H3K27ac in different categories of segments including all segments, HighSpe, InterSpe, UniHypo, UnipLow, UnipHigh UniHyper segments, and H1 specific MethyMarks. Spearman's rank correlation coefficient (SCC) between DNA methylation and H3K27ac in each category of segments was calculated by the R function “cor.test”. P represents the significance of the coefficient. (B) Each sub-figure shows the density scatterplot of DNA methylation and H3K27me3 in different categories of segments including all segments, HighSpe, InterSpe, UniHypo, UnipLow, UnipHigh UniHyper segments, and H1 specific MethyMarks. SCC represents Spearman's rank correlation coefficient between DNA methylation and H3K27me3 in each category of segments. P represents the significance of the coefficient.

Supplementary Figure 25



Supplementary Figure 25: hESC H1-specific HypoMarks and HyperMarks show distinct chromatin modifications. (A) Heatmap of log2 enrichment ratios of several histone marks and transcription factors vs. DNA input at HypoMark/HyperMark ±3 Kb regions mapped by ngs.plot (12). “L” and “R” represent the boundary of HypoMark/HyperMark. The log2 enrichment ratios were represented by colors from green (low) to red (high). (B) Average profiles of log2 enrichment ratios of several histone marks and transcription factors vs. DNA input at promoter HypoMark/HyperMark ±3 Kb regions.

Supplementary Figure 26

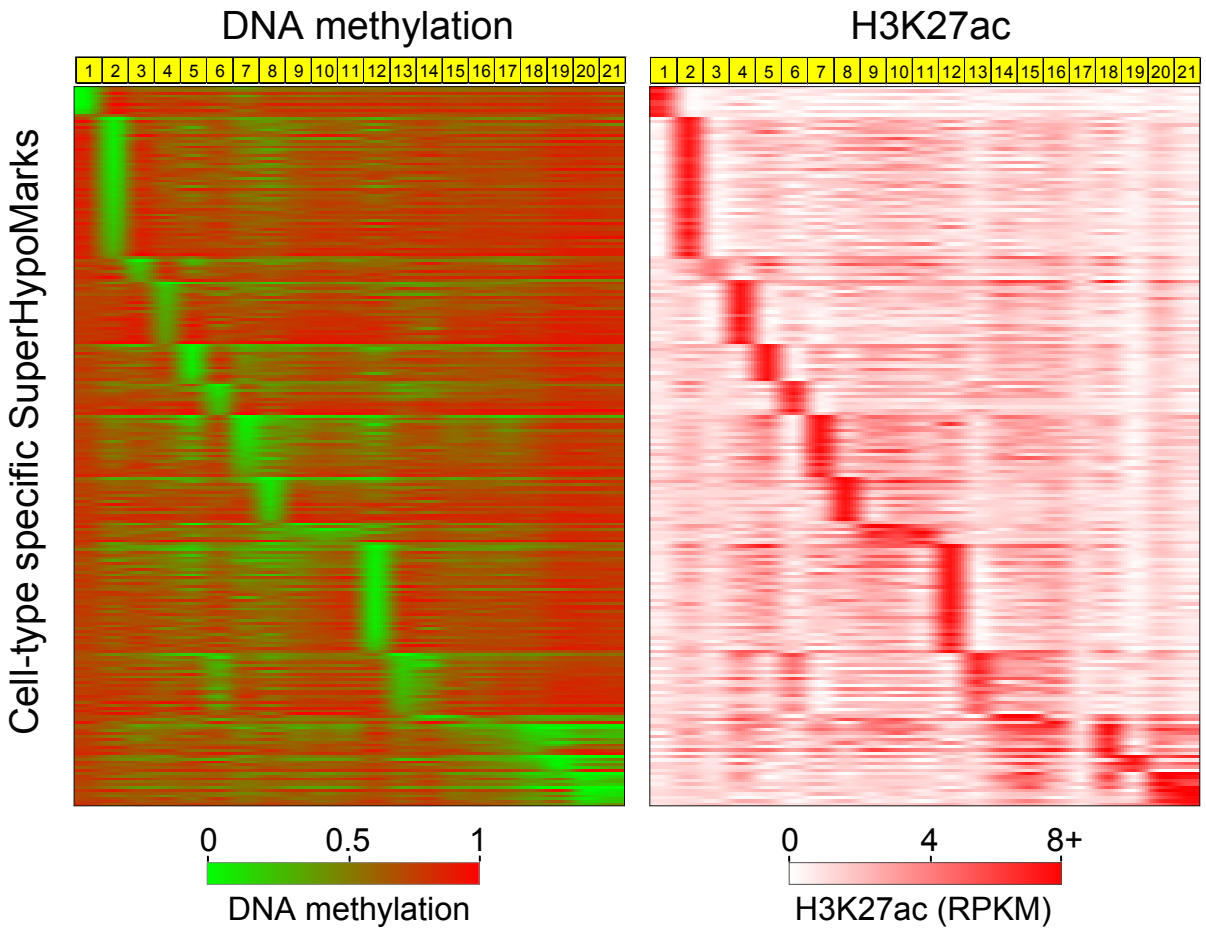


D Function enrichment of H1-specific HypoMarks binded by NANOG or POU5F1

Ontology	ID	Description	P value
GO Biological Process	GO:0048646	anatomical structure formation involved in morphogenesis	1.08E-06
GO Biological Process	GO:0010468	regulation of gene expression	1.87E-06
GO Biological Process	GO:0009653	anatomical structure morphogenesis	2.41E-05
GO Biological Process	GO:0009888	tissue development	1.05E-04
GO Biological Process	GO:0010453	regulation of cell fate commitment	1.29E-04
O Biological Process	GO:0009790	embryo development	3.95E-04
Biological Process	GO:0043045	DNA methylation involved in embryo development	4.02E-04
GO Biological Process	GO:0042659	regulation of cell fate specification	6.84E-04
GO Biological Process	GO:0051093	negative regulation of developmental process	6.94E-04
Mouse Phenotype	MP:0002084	abnormal developmental patterning	3.69E-07
Mouse Phenotype	MP:0001672	abnormal embryogenesis/ development	6.99E-07
Mouse Phenotype	MP:0001730	embryonic growth arrest	4.08E-06
Mouse Phenotype	MP:0001698	decreased embryo size	1.07E-05
Mouse Phenotype	MP:0003984	embryonic growth retardation	1.67E-05
MSigDB Perturbation	BENPORATH_ES_1	Set 'ES exp1': genes overexpressed in human embryonic stem cells according to 5 or more out of 20 profiling studies.	2.65E-12
MSigDB Perturbation	BENPORATH_SOX2_TARGETS	Set 'Sox2 targets': genes upregulated and identified by ChIP on chip as SOX2 [Gene ID=6657] transcription factor targets in human embryonic stem cells.	6.73E-08
MSigDB Perturbation	BENPORATH_NANOG_TARGETS	Set 'Nanog targets': genes upregulated and identified by ChIP on chip as Nanog [Gene ID=79923] transcription factor targets in human embryonic stem cells.	6.60E-07
MSigDB Perturbation	BENPORATH_NOS_TARGETS	Set 'NOS targets': genes upregulated and identified by ChIP on chip as targets of the transcription factors NANOG [Gene ID=79923], OCT4[Gene ID=5460], and Sox2 [Gene ID=6657] (NOS) in human embryonic stem cells.	8.71E-06
MSigDB Perturbation	WONG_EMBRYONIC_STEM_CELL_CORE	The 'core ESC-like gene module': genes coordinately up-regulated in a compendium of mouse embryonic stem cells (ESC) which are shared with the human ESC-like module.	6.22E-04
Transcription Factor Targets	Nanog_Boyer	Targets of Nanog, identifed by ChIP-chip in embryonic stem cells	1.60E-03

Supplementary Figure 26: Sub-network of transcription factor-MethyMark collaboration network in hESC H1 cells as shown in Figure 4D. (A) Sub-network derived by transcriptional factors from transcription factor-MethyMark collaboration network in hESC H1 cells. The size of the transcription factor (TF) node represents the number of the MethyMarks bound by it, and the width of the TF-TF line represents the number of MethyMarks co-targeted by two TFs. (B) Sub-network derived by H1 MethyMarks from TF-MethyMark collaboration network in the hESC H1 cell line. The width of the MethyMark-MethyMark line represents the number of TFs binding to both MethyMarks. Only the lines with more than ten TFs are shown. (C) Sub-network derived by transcriptional factors NANOG and POU5F1 from TF-MethyMark collaboration network in the hESC H1 cell line. From TF-MethyMark collaboration network in the hESC H1 cell line of Fig. 4, NANOG and POU5F1 and their one-step neighboring nodes and the lines between these nodes were extracted to construct this sub-network. It was shown that most methylated segments in this sub-network were H1-specific HypoMarks, and these HypoMarks were prone to be bound by the same active TFs. (D) Functional enrichment of H1-specific HypoMarks in NANOG and POU5F1 related sub-network. GREAT (http://bejerano.stanford.edu/great/public/html/) was used to perform the functional enrichment of H1-specific HypoMarks in the sub-network. H1-specific HypoMarks were assigned to nearby protein-coding genes based on GREAT's basal plus extension rule for regulatory regions (proximal: 5 kb upstream, 1 kb downstream, plus distal up to 1 Mb). Significant annotated terms from the enrichment analysis were selected by both hypergeometric and binomial tests (P<0.05). Four enriched functions were found as targets of TF NANOG, POU5F1 and SOX2, overexpression in human ESC, functions related with embryonic development, and abnormal developmental phenotype.

Supplementary Figure 27



Supplementary Figure 27: DNA methylation and H3K27ac state of cell type-specific SuperHypoMarks across 21 cell types. Each row denotes a SuperHypoMark, and each column denotes a cell type. DNA methylation level was represented by a gradient from green (ummethylated) to red (full methylation), and H3K27ac from white (lowest) to red (highest). RPKM represents the H3K27ac reads per kilobase per million mapped reads in a given segment.

Supplementary Figure 28

Supplementary Figure 28: Detailed information about epigenetics and expression of H1-specific SuperHypoMark genes. Shown in this figure are the example genes related to H1-specific SuperHypoMarks. For each gene, the epigenetic pattern was visualized by our local UCSC genome browser, and the expression patterns of the genes were visualized by Epigenome Atlas Browser (http://www.genboree.org/EpigenomeAtlasBrowser). The detailed analysis for each gene is listed as followings:

POU5F1: The promoter region of *POU5F1* is H1-specific hypomethylated and bound by mediator coactivators including RNA polymerase II, mediator and transcription factors (such as POU5F1 and NANOG), which form a super-enhancer in this region. We also found the *POU5F1* promoter was enriched by histone H3K27ac, a surrogate mark of a super-enhancer, and H3K4me3, an active mark for gene expression that has been identified as an H1-specific promoter or enhancer state by a hidden Markov model (Ernst et al. 2011). Furthermore, *POU5F1* was specifically expressed at extremely high levels in the H1 cell line.

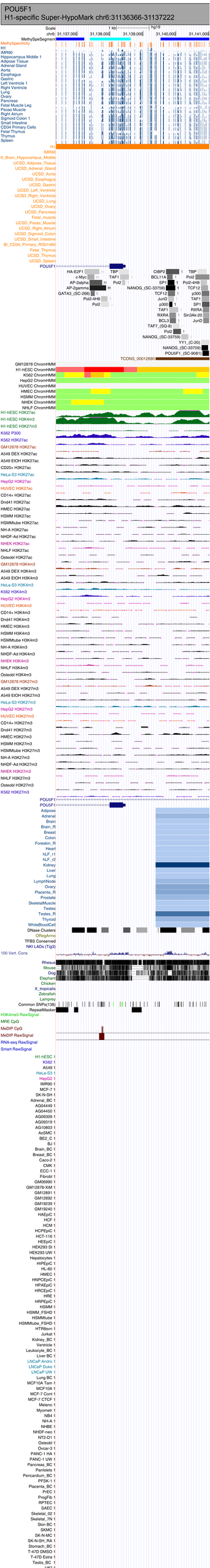
NANOG: We found *NANOG* showed very similar epigenetic and expression patterns to *POU5F1*.

DNMT3B: Interestingly, the DNA methyltransferase DNMT3B that was essential for de novo methylation and mammalian development (Okano et al. 1999) showed H1-specific hypomethylation and extremely high expression levels, which is consistent with its downregulation after ES cell differentiation as reported previously (Okano et al. 1998; Watanabe et al. 2002). The unmethylated status of the promoter regions facilitates the formation of a super-enhancer, which accounts for the extremely high expression of *DNMT3B*. In stem cells, the high expression of *DNMT3B* induces high levels of DNA methyltransferase, which further methylates most genome CpGs except those related to pluripotency maintenance.

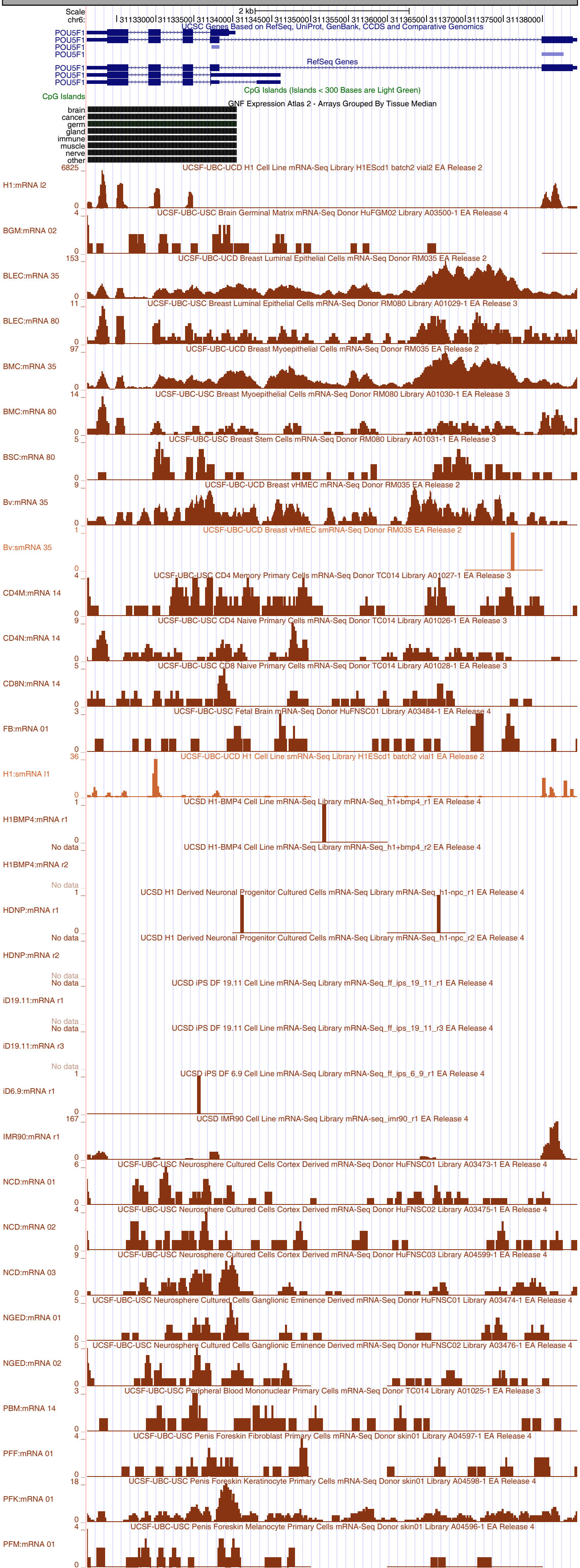
NSD1: Another H1-specific hypomethylated gene *NSD1* (also known as *KMT3B*) encodes a histone methyltransferase that preferentially methylates H3 lysine 36. The methylation data by another technology, Infinium Methylation 450K, confirms low methylation levels of this region in the H1 cell line but high methylation levels in adult tissues and other cell types. Furthermore, *NSD1* shows higher expression in the H1 cell line than other cell types.

LINC00678: This gene was one of 22 lncRNA genes that overlapped with H1-specific SuperHypoMarks. It was shown that the promoter region of this gene was specifically hypomethylated and extremely highly expressed in ESC and iPSC cell lines. However, the expression level of this gene in extremely low, which is consistent with a previous finding of its down-regulation during the transition from iPSCs to NPCs (Chen et al. 2013). As far as we know, there were no more reports about the functions of *LINC00678* in stem cells, suggesting it may be a novel mark of stem cells.

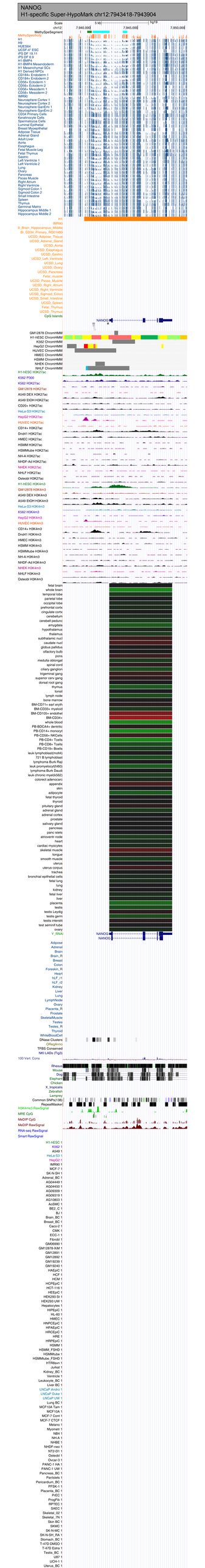
miR-6130: A microRNA gene *miR-6130* overlapped with two ESC-specific SuperHypoMarks. We found *miR-6130* was the longest microRNA gene (836, 530 bp) in the list of Refseq genes. It was specifically hypomethylated in ESCs and iPSCs and overlapped with a super enhancer that was only found in the H1 cell line. As far as we know, there were no more reports about the functions of *miR-6130* in stem cells, suggesting it may be a novel mark of stem cells.


POU5F1
H1-specific Super-HypoMark chr6:31136366-31137222
Scale
chr6: 31,137,000
1 kb
hg19
31,138,000
31,139,000
31,140,000
31,141,000
MethySpecificity
MethySeSegment
H1
IMR90
Hippocampus Middle 1
Adipose Tissue
Adrenal Gland
Aorta
Esophagus
Gastric
Left Ventricle 1
Right Ventricle
Lung
Ovary
Pancreas
Fetal Muscle Leg
Psoas Muscle
Right Atrium
Sigmoid Colon 1
Small Intestine
CD34 Primary Cells
Fetal Thymus
Thymus
Spleen
H1
IMR90
8I_Brain_Hippocampus_Middle
UCSD_Adipose_Tissue
UCSD_Adrenal_Gland
UCSD_Aorta
UCSD_Esophagus
UCSD_Gastric
UCSD_Left_Ventricle
UCSD_Right_Ventricle
UCSD_Lung
UCSD_Ovary
UCSD_Pancreas
Fetal_muscle
UCSD_Psoas_Muscle
UCSD_Right_Atrium
UCSD_Sigmoid_Colon
UCSD_Small_Intestine
BI_CD34_Primary_RO01480
Fetal_Thymus
UCSD_Thymus
UCSD_Spleen
POU5F1
HA-E2F1
c-Myc
AP-2alpha
AP-2gamma
GATA3_(SC-268)
TBP
TAF1
Pol2
Pol2-4H8
Pol2
H
H
H
CtBP2
BCL11A
SP1
NANOG_(SC-33759)
TCF12
JunD
p300
TAF1
RXRA
BCL3
TAF7_(SC-8)
Pol2
NANOG_(SC-33759)
TBP
Pol2
Pol2-4H8
TCF12
p300
TAF1
SP1
RXRA
Sin3Ak-20
JunD
YY1_(C-20)
NANOG_(SC-33759)
POU5F1_(SC-9081)
TCONS_00012690
GM12878 ChromHMM
H1-hESC ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HUVEC ChromHMM
HMEC ChromHMM
HSMM ChromHMM
NHEK ChromHMM
NHLF ChromHMM
H1-hESC H3K27ac
H1-hESC H3K4m3
H1-hESC H3K27m3
K562 P300
K562 H3K27ac
GM12878 H3K27ac
A549 DEX H3K27ac
A549 EtOH H3K27ac
CD20+ H3K27ac
HeLa-S3 H3K27ac
HepG2 H3K27ac
HUVEC H3K27ac
CD14+ H3K27ac
Dnd41 H3K27ac
HMEC H3K27ac
HSMM H3K27ac
HSMMtube H3K27ac
NH-A H3K27ac
NHDF-Ad H3K27ac
NHEK H3K27ac
NHLF H3K27ac
Osteobl H3K27ac
GM12878 H3K4m3
A549 DEX H3K4m3
A549 EtOH H3K4m3
HeLa-S3 H3K4m3
K562 H3K4m3
HepG2 H3K4m3
HUVEC H3K4m3
CD14+ H3K4m3
Dnd41 H3K4m3
HMEC H3K4m3
HSMM H3K4m3
HSMMtube H3K4m3
NH-A H3K4m3
NHDF-Ad H3K4m3
NHEK H3K4m3
NHLF H3K4m3
Osteobl H3K4m3
GM12878 H3K27m3
A549 DEX H3K27m3
A549 EtOH H3K27m3
HeLa-S3 H3K27m3
HepG2 H3K27m3
HUVEC H3K27m3
CD14+ H3K27m3
Dnd41 H3K27m3
HMEC H3K27m3
HSMM H3K27m3
HSMMtube H3K27m3
NH-A H3K27m3
NHDF-Ad H3K27m3
NHEK H3K27m3
NHLF H3K27m3
Osteobl H3K27m3
K562 H3K27m3
POU5F1
POU5F1
Adipose
Adrenal
Brain
Brain_R
Breast
Colon
Foreskin_R
Heart
hLF_r1
hLF_r2
Kidney
Liver
Lung
LymphNode
Ovary
Placenta_R
Prostate
SkeletalMuscle
Testes
Testes_R
Thyroid
WhiteBloodCell
DNase Clusters
ORegAnno
TFBS Conserved
NKI LADs (Tig3)
100 Vert. Cons
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(138)
RepeatMasker
H3K4me3 RawSignal
MRE CpG
MeDIP CpG
MeDIP RawSignal
RNA-seq RawSignal
Smart RawSignal
H1-hESC 1
K562 1
A549 1
HeLa-S3 1
HepG2 1
IMR90 1
MCF-7 1
SK-N-SH 1
Adrenal_BC 1
AG04449 1
AG04450 1
AG09309 1
AG09319 1
AG10803 1
AoSMC 1
BE2_C 1
BJ 1
Brain_BC 1
Breast_BC 1
Caco-2 1
CMK 1
ECC-1 1
Fibrobl 1
GM06990 1
GM12878-XiM 1
GM12891 1
GM12892 1
GM19239 1
GM19240 1
HAEpiC 1
HCF 1
HCM 1
HCPEpiC 1
HCT-116 1
HEEpiC 1
HEK293 St 1
HEK293 UW 1
Hepatocytes 1
HIPEpiC 1
HL-60 1
HMEC 1
HNPCEpiC 1
HPAEpiC 1
HRCEpiC 1
HRE 1
HRPEpiC 1
HSMM 1
HSMM_FSHD 1
HSMMtube 1
HSMMtube_FSHD 1
HTR8svn 1
Jurkat 1
Kidney_BC 1
Ventricle 1
Leukocyte_BC 1
Liver BC 1
LNCaP Andro 1
LNCaP Duke 1
LNCaP UW 1
Lung BC 1
MCF10A Tam 1
MCF10A 1
MCF-7 Cont 1
MCF-7 CTCF 1
Melano 1
Myometr 1
NB4 1
NH-A 1
NHBE 1
NHDF-neo 1
NT2-D1 1
Osteobl 1
Ovcar-3 1
PANC-1 HA 1
PANC-1 UW 1
Pancreas_BC 1
Panislets 1
Pericardium_BC 1
PFSK-1 1
Placenta_BC 1
PrEC 1
ProgFib 1
RPTEC 1
SAEC 1
Skeletal_02 1
Skeletal_7N 1
Skin BC 1
SKMC 1
SK-N-MC 1
SK-N-SH_RA 1
Stomach_BC 1
T-47D DMSO 1
T-47D Estra 1
Testis_BC 1
U87 1
UCH-1 1
Uterus_BC 1

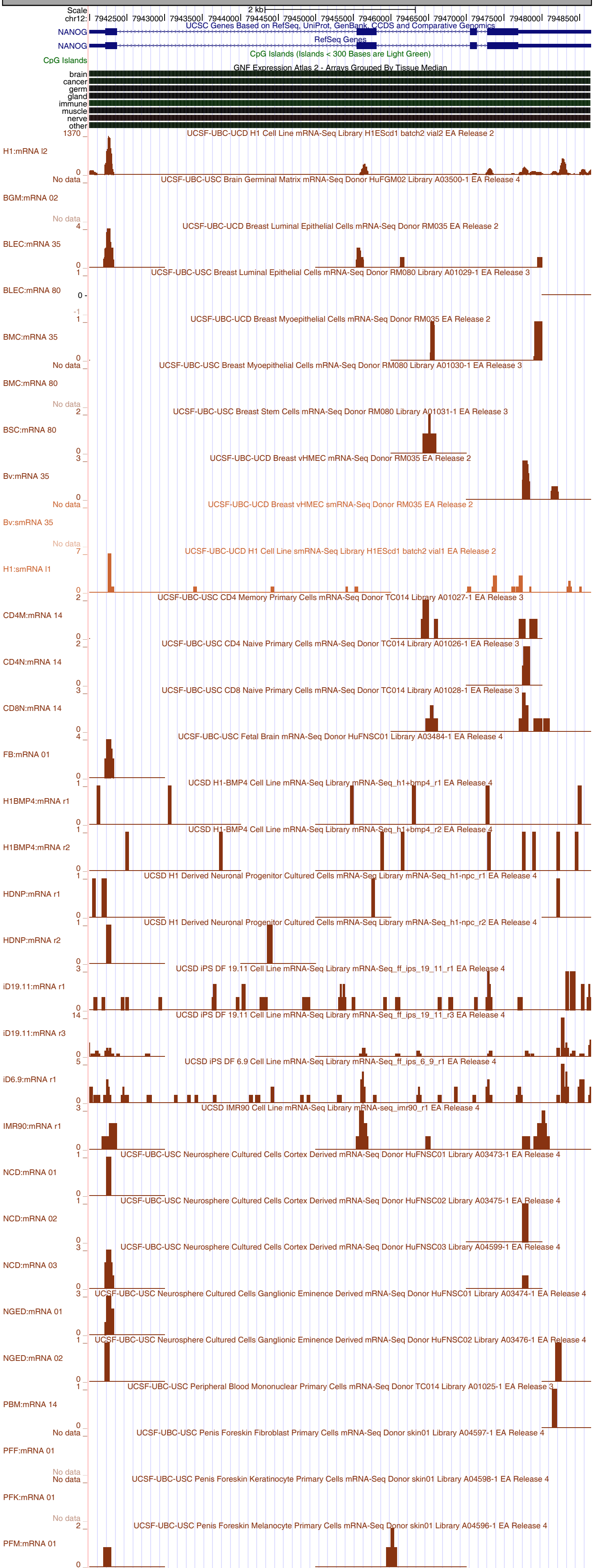
POU5F1
Expression at http://www.genboree.org/EpigenomeAtlasBrowser



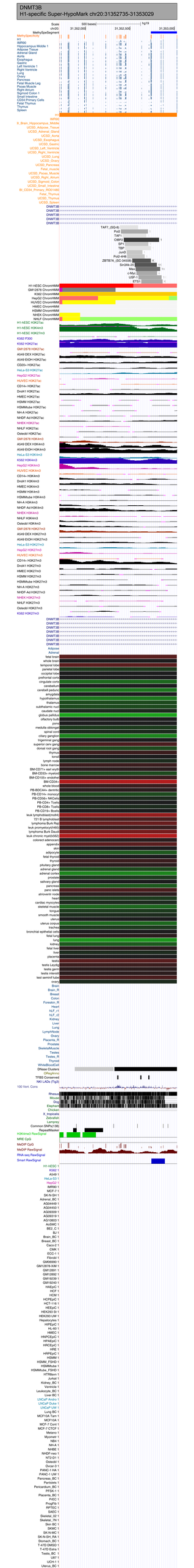
NANOG
H1-specific Super-HypoMark chr12:7943418-7943904



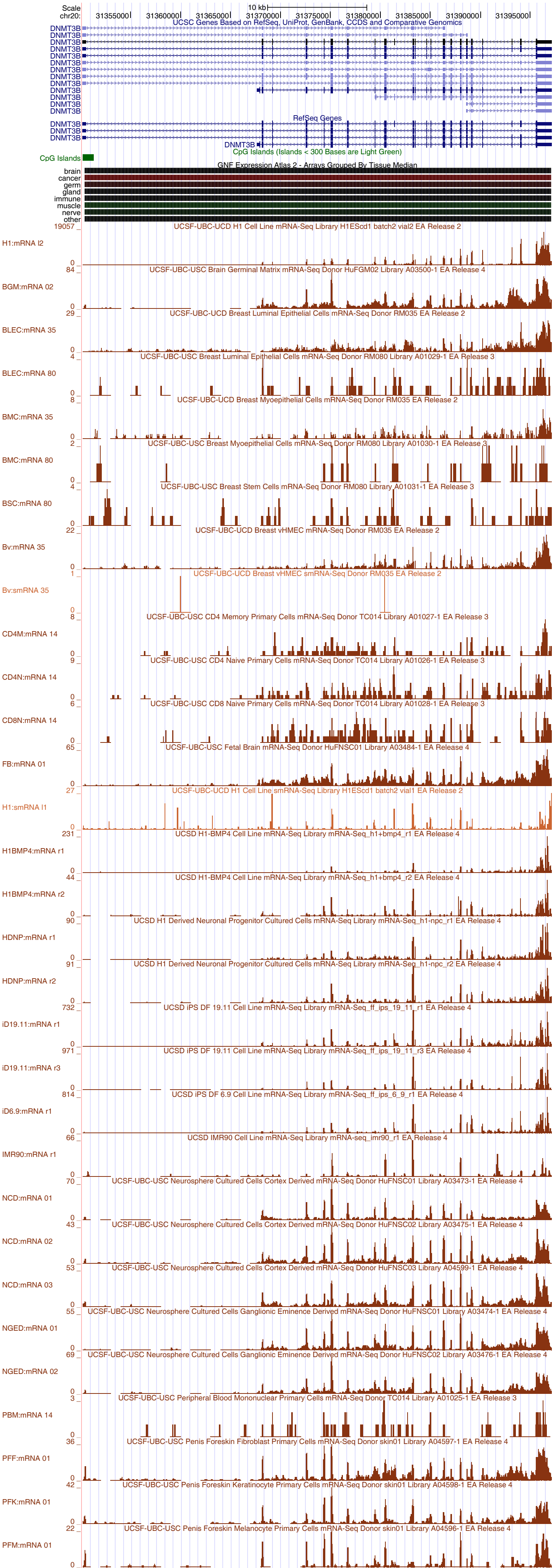
NANOG
Expression at http://www.genboree.org/EpigenomeAtlasBrowser



DNMT3B
H1-specific Super-HypoMark chr20:31352735-31353029

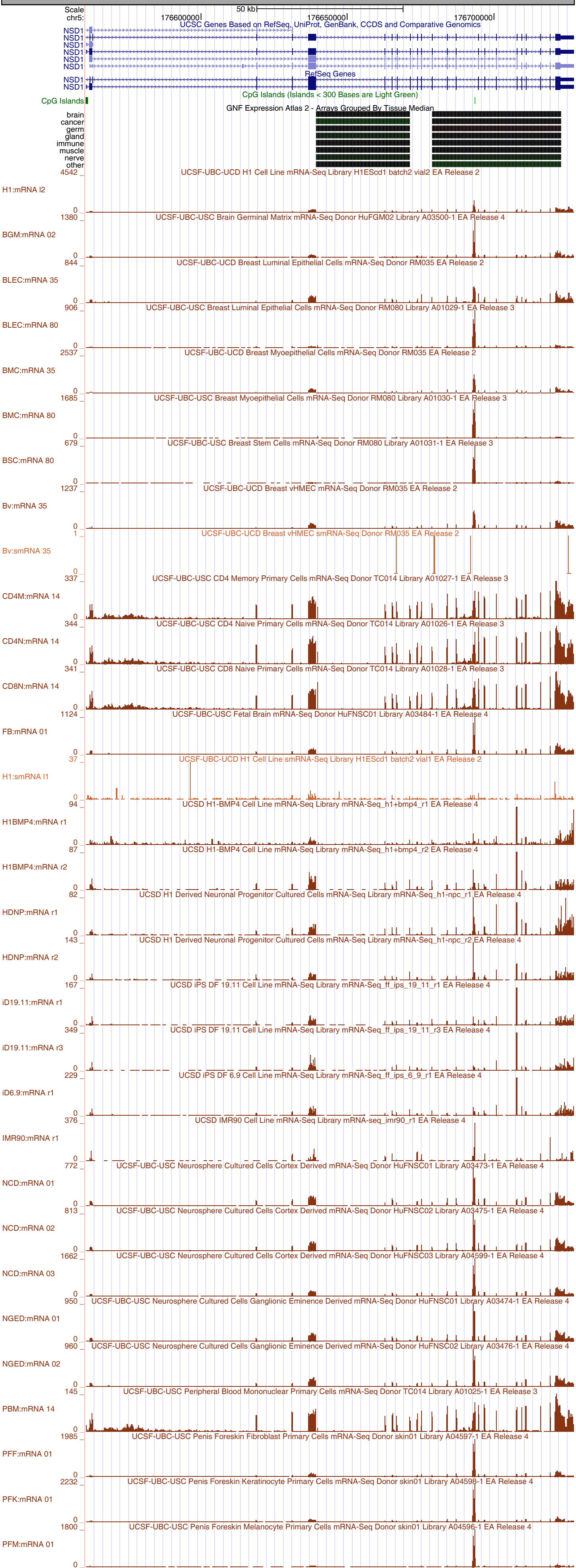


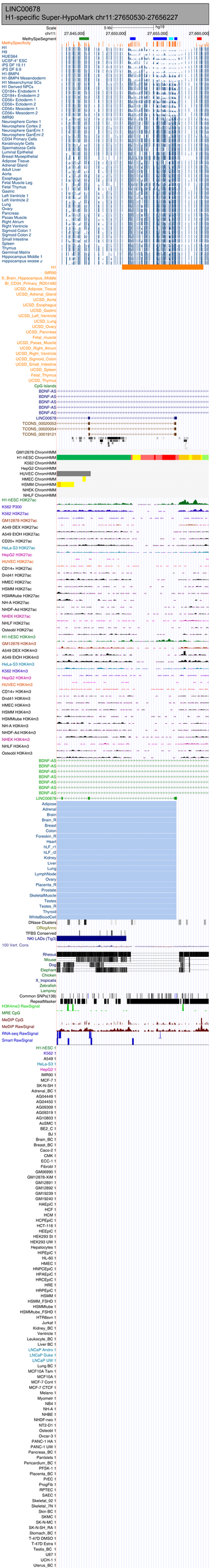
DNMT3B
Expression at http://www.genboree.org/EpigenomeAtlasBrowser




NSD1
H1-specific Super-HypoMark chr5:176558853-176558910
Scale
chr5:
200 bases
hg19
176,559,000
176,559,500
MethySpeSegment
MethySpecificity
H1
IMR90
Hippocampus Middle 1
Adipose Tissue
Adrenal Gland
Aorta
Esophagus
Gastric
Left Ventricle 1
Right Ventricle
Lung
Ovary
Pancreas
Fetal Muscle Leg
Psoas Muscle
Right Atrium
Sigmoid Colon 1
Small Intestine
CD34 Primary Cells
Fetal Thymus
Thymus
Spleen
H1
IMR90
Bl_Brain_Hippocampus_Middle
UCSD_Adipose_Tissue
UCSD_Adrenal_Gland
UCSD_Aorta
UCSD_Esophagus
UCSD_Gastric
UCSD_Left_Ventricle
UCSD_Right_Ventricle
UCSD_Lung
UCSD_Ovary
UCSD_Pancreas
Fetal_muscle
UCSD_Psoas_Muscle
UCSD_Right_Atrium
UCSD_Sigmoid_Colon
UCSD_Small_Intestine
Bl_CD34_Primary_RO01480
Fetal_Thymus
UCSD_Thymus
UCSD_Spleen
CpG: 263
Max
ATF3
USF2
USF-1
Pol2
Pol2-4H8
TAF1
Klf4
H1-hESC ChromHMM
GM12878 ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HUVEC ChromHMM
HMEC ChromHMM
HSMM ChromHMM
NHEK ChromHMM
NHLF ChromHMM
H1-hESC H3K27ac
H1-hESC H3K4m3
H1-hESC H3K27m3
K562 P300
K562 H3K27ac
GM12878 H3K27ac
A549 DEX H3K27ac
A549 EtOH H3K27ac
CD20+ H3K27ac
HeLa-S3 H3K27ac
HepG2 H3K27ac
HUVEC H3K27ac
CD14+ H3K27ac
Dnd41 H3K27ac
HMEC H3K27ac
HSMM H3K27ac
HSMMtube H3K27ac
NH-A H3K27ac
NHDF-Ad H3K27ac
NHEK H3K27ac
NHLF H3K27ac
Osteobl H3K27ac
GM12878 H3K4m3
A549 DEX H3K4m3
A549 EtOH H3K4m3
HeLa-S3 H3K4m3
K562 H3K4m3
HepG2 H3K4m3
HUVEC H3K4m3
CD14+ H3K4m3
Dnd41 H3K4m3
HMEC H3K4m3
HSMM H3K4m3
HSMMtube H3K4m3
NH-A H3K4m3
NHDF-Ad H3K4m3
NHEK H3K4m3
NHLF H3K4m3
Osteobl H3K4m3
GM12878 H3K27m3
A549 DEX H3K27m3
A549 EtOH H3K27m3
HeLa-S3 H3K27m3
HepG2 H3K27m3
HUVEC H3K27m3
CD14+ H3K27m3
Dnd41 H3K27m3
HMEC H3K27m3
HSMM H3K27m3
HSMMtube H3K27m3
NH-A H3K27m3
NHDF-Ad H3K27m3
NHEK H3K27m3
NHLF H3K27m3
Osteobl H3K27m3
K562 H3K27m3
Adipose
Adrenal
Brain
Brain_R
Breast
Colon
Foreskin_R
Heart
hLF_r1
hLF_r2
Kidney
Liver
Lung
LymphNode
Ovary
Placenta_R
Prostate
SkeletalMuscle
Testes
Testes_R
Thyroid
WhiteBloodCell
DNase Clusters
ORegAnno
TFBS Conserved
NKI LADs (Tig3)
100 Vert. Cons
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(138)
RepeatMasker
H3K4me3 RawSignal
MRE CpG
MeDIP CpG
MeDIP RawSignal
RNA-seq RawSignal
Smart RawSignal
H1-hESC 1
K562 1
A549 1
HeLa-S3 1
HepG2 1
IMR90 1
MCF-7 1
SK-N-SH 1
Adrenal_BC 1
AG04449 1
AG04450 1
AG09309 1
AG09319 1
AG10803 1
AoSMC 1
BE2_C 1
BJ 1
Brain_BC 1
Breast_BC 1
Caco-2 1
CMK 1
ECC-1 1
Fibrobl 1
GM06990 1
GM12878-XMI 1
GM12891 1
GM12892 1
GM19238 1
GM19240 1
HAEpiC 1
HCF 1
HCM 1
HCPEpiC 1
HCT-116 1
HEEpiC 1
HEK293 St 1
HEK293 UW 1
Hepatocytes 1
HIPEpiC 1
HL-60 1
HMEC 1
HNPCEpiC 1
HPAEpiC 1
HRCEpiC 1
HRE 1
HRPEpiC 1
HSMM 1
HSMM_FSHD 1
HSMMtube 1
HSMMtube_FSHD 1
HTR8svn 1
Jurkat 1
Kidney_BC 1
Ventricle 1
Leukocyte_BC 1
Liver BC 1
LNCaP Andro 1
LNCaP Duke 1
LNCaP UW 1
Lung BC 1
MCF10A Tam 1
MCF10A 1
MCF-7 Cont 1
MCF-7 CTCF 1
Melano 1
Myometr 1
NB4 1
NH-A 1
NHBE 1
NHDF-neo 1
NT2-D1 1
Osteobl 1
Ovcar-3 1
PANC-1 HA 1
PANC-1 UW 1
Pancreas_BC 1
Panislets 1
Pericardium_BC 1
PFSK-1 1
Placenta_BC 1
PrEC 1
ProgFib 1
RPTEC 1
SAEC 1
Skeletal_02 1
Skeletal_7N 1
Skin BC 1
SKMC 1
SK-N-MC 1
SK-N-SH_RA 1
Stomach_BC 1
T-47D DMSO 1
T-47D Estra 1
Testis_BC 1
U87 1
UCH-1 1
Uterus_BC 1

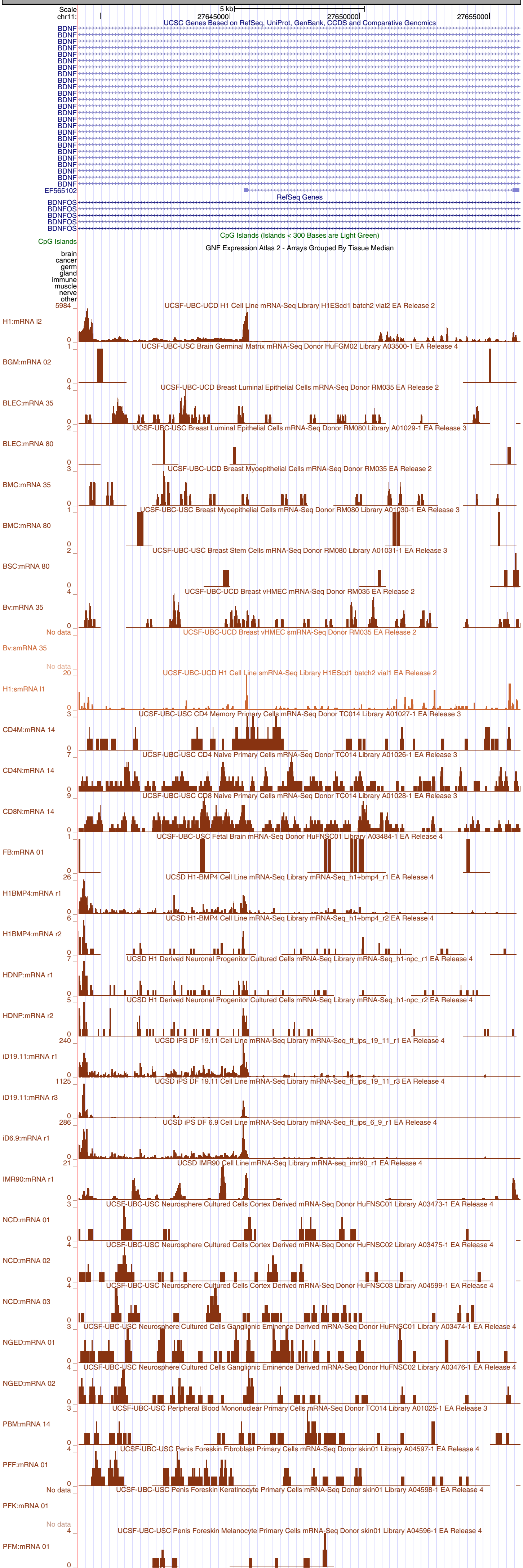
NSD1
Expression at http://www.genboree.org/EpigenomeAtlasBrowser

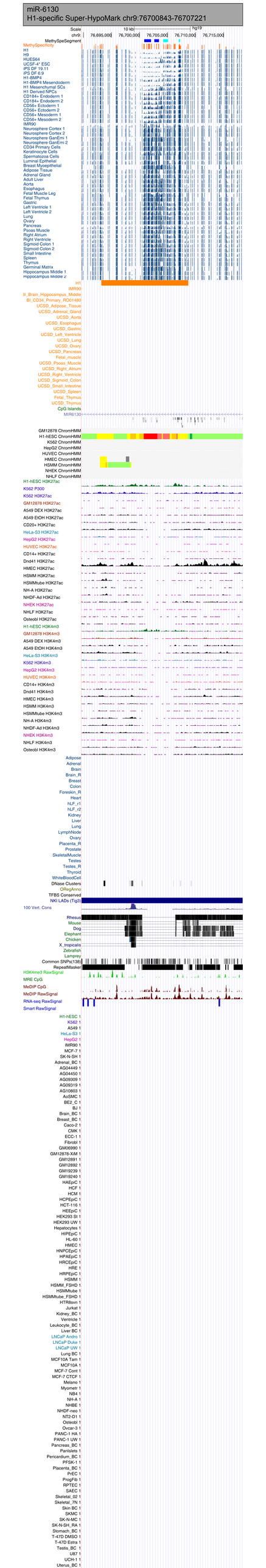



LINC00678
H1-specific Super-HypoMark chr11:27650530-27656227
Scale
chr11:
5 kb
hg19
27,645,000
27,650,000
27,655,000
27,660,000
MethySpecificity
MethySeSegment
H1
H9
HUES64
UCSF-4* ESC
iPS DF 19.11
iPS DF 6.9
H1-BMP4
H1-BMP4 Mesendoderm
H1 Mesenchymal SCs
H1 Derived NPCs
CD184+ Endoderm 1
CD184+ Endoderm 2
CD56+ Ectoderm 1
CD56+ Ectoderm 2
CD56+ Mesoderm 1
CD56+ Mesoderm 2
IMR90
Neurosphere Cortex 1
Neurosphere Cortex 2
Neurosphere GanEmt 1
Neurosphere GanEmt 2
CD34 Primary Cells
Keratinocyte Cells
Spermatozoa Cells
Luminal Epithelial
Breast Myoepithelial
Adipose Tissue
Adrenal Gland
Adult Liver
Aorta
Esophagus
Fetal Muscle Leg
Fetal Thymus
Gastric
Left Ventricle 1
Left Ventricle 2
Lung
Ovary
Pancreas
Psoas Muscle
Right Atrium
Right Ventricle
Sigmoid Colon 1
Sigmoid Colon 2
Small Intestine
Spleen
Thymus
Germinal Matrix
Hippocampus Middle 1
Hippocampus Middle 2
H1
IMR90
BI_Brain_Hippocampus_Middle
BI_CD34_Primary_RO01480
UCSD_Adipose_Tissue
UCSD_Adrenal_Gland
UCSD_Aorta
UCSD_Esophagus
UCSD_Gastric
UCSD_Left_Ventricle
UCSD_Lung
UCSD_Ovary
UCSD_Pancreas
Fetal_muscle
UCSD_Psoas_Muscle
UCSD_Right_Atrium
UCSD_Right_Ventricle
UCSD_Sigmoid_Colon
UCSD_Small_Intestine
UCSD_Spleen
Fetal_Thymus
UCSD_Thymus
CpG Islands
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
LINC00678
TCONS_00020053
TCONS_00020054
TCONS_00019121
GM12878 ChromHMM
H1-hESC ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HUVEC ChromHMM
HMEC ChromHMM
HSMM ChromHMM
NHEK ChromHMM
NHLF ChromHMM
H1-hESC H3K27ac
K562 P300
K562 H3K27ac
GM12878 H3K27ac
A549 DEX H3K27ac
A549 EtOH H3K27ac
CD20+ H3K27ac
HeLa-S3 H3K27ac
HepG2 H3K27ac
HUVEC H3K27ac
CD14+ H3K27ac
Dnd41 H3K27ac
HMEC H3K27ac
HSMM H3K27ac
HSMMtube H3K27ac
NH-A H3K27ac
NHDF-Ad H3K27ac
NHEK H3K27ac
NHLF H3K27ac
Osteobl H3K27ac
H1-hESC H3K4m3
GM12878 H3K4m3
A549 DEX H3K4m3
A549 EtOH H3K4m3
HeLa-S3 H3K4m3
K562 H3K4m3
HepG2 H3K4m3
HUVEC H3K4m3
CD14+ H3K4m3
Dnd41 H3K4m3
HMEC H3K4m3
HSMM H3K4m3
HSMMtube H3K4m3
NH-A H3K4m3
NHDF-Ad H3K4m3
NHEK H3K4m3
NHLF H3K4m3
Osteobl H3K4m3
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
BDNF-AS
LINC00678
Adipose
Adrenal
Brain
Brain_R
Breast
Colon
Foreskin_R
Heart
hLF_r1
hLF_r2
Kidney
Liver
Lung
LymphNode
Ovary
Placenta_R
Prostate
SkeletalMuscle
Testes
Testes_R
Thyroid
WhiteBloodCell
DNase Clusters
ORegAnno
TFBS Conserved
NKI LADs (Tig3)
100 Vert. Cons
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(138)
RepeatMasker
H3K4me3 RawSignal
MRE CpG
MeDIP CpG
MeDIP RawSignal
RNA-seq RawSignal
Smart RawSignal
H1-hESC 1
K562 1
A549 1
HeLa-S3 1
HepG2 1
IMR90 1
MCF-7 1
SK-N-SH 1
Adrenal_BC 1
AG04449 1
AG04450 1
AG09309 1
AG09319 1
AG10803 1
AoSMC 1
BE2_C 1
BJ 1
Brain_BC 1
Breast_BC 1
Caco-2 1
CMK 1
ECC-1 1
Fibrobl 1
GM06990 1
GM12878-XiM 1
GM12891 1
GM12892 1
GM19239 1
GM19240 1
HAEpiC 1
HCF 1
HCM 1
HCPEpiC 1
HCT-116 1
HEEpiC 1
HEK293 St 1
HEK293 UW 1
Hepatocytes 1
HIPEpiC 1
HL-60 1
HMEC 1
HNPCEpiC 1
HPAEpiC 1
HRCEpiC 1
HRE 1
HRPEpiC 1
HSMM 1
HSMM_FSHD 1
HSMMtube 1
HSMMtube_FSHD 1
HTRBsvn 1
Jurkat 1
Kidney_BC 1
Ventricle 1
Leukocyte_BC 1
Liver BC 1
LNCaP Andro 1
LNCaP Duke 1
LNCaP UW 1
Lung BC 1
MCF10A Tam 1
MCF10A 1
MCF-7 Cont 1
MCF-7 CTCF 1
Melano 1
Myometr 1
NB4 1
NH-A 1
NHBE 1
NHDF-neo 1
NT2-D1 1
Osteobl 1
Ovcar-3 1
PANC-1 HA 1
PANC-1 UW 1
Pancreas_BC 1
Panislets 1
Pericardium_BC 1
PFSK-1 1
Placenta_BC 1
PrEC 1
ProgFib 1
RPTEC 1
SAEC 1
Skeletal_02 1
Skeletal_7N 1
Skin BC 1
SKMC 1
SK-N-MC 1
SK-N-SH_RA 1
Stomach_BC 1
T-47D DMSO 1
T-47D Estra 1
Testis_BC 1
U87 1
UCH-1 1
Uterus_BC 1

LINC00678 NR_102708,chr11,-,27639172,27656174

Expression at http://www.genboree.org/EpigenomeAtlasBrowser




miR-6130
H1-specific Super-HypoMark chr9:76700843-76707221
Scale
chr9:
10 kb
hg19
76,695,000
76,700,000
76,705,000
76,710,000
76,715,000
MethySpeSegment
MethySpecificity
H1
H9
HUES64
UCSF-4* ESC
iPS DF 19.11
iPS DF 6.9
H1-BMP4
H1-BMP4 Mesendoderm
H1 Mesenchymal SCs
H1 Derived NPCs
CD184+ Endoderm 1
CD184+ Endoderm 2
CD56+ Ectoderm 1
CD56+ Ectoderm 2
CD56+ Mesoderm 1
CD56+ Mesoderm 2
IMR90
Neurosphere Cortex 1
Neurosphere Cortex 2
Neurosphere GanEm 1
Neurosphere GanEm 2
CD34 Primary Cells
Keratinocyte Cells
Spermatozoa Cells
Luminal Epithelial
Breast Myoepithelial
Adipose Tissue
Adrenal Gland
Adult Liver
Aorta
Esophagus
Fetal Muscle Leg
Fetal Thymus
Gastric
Left Ventricle 1
Left Ventricle 2
Lung
Ovary
Pancreas
Psoas Muscle
Right Atrium
Right Ventricle
Sigmoid Colon 1
Sigmoid Colon 2
Small Intestine
Spleen
Thymus
Germinal Matrix
Hippocampus Middle 1
Hippocampus Middle 2
H1
IMR90
BI_Brain_Hippocampus_Middle
BI_CD34_Primary_R001480
UCSD_Adipose_Tissue
UCSD_Adrenal_Gland
UCSD_Aorta
UCSD_Esophagus
UCSD_Gastric
UCSD_Left_Ventricle
UCSD_Lung
UCSD_Ovary
UCSD_Pancreas
Fetal_muscle
UCSD_Psoas_Muscle
UCSD_Right_Atrium
UCSD_Right_Ventricle
UCSD_Sigmoid_Colon
UCSD_Small_Intestine
UCSD_Spleen
Fetal_Thymus
UCSD_Thymus
CpG Islands
MIR6130
GM12878 ChromHMM
H1-hESC ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HUVEC ChromHMM
HMEC ChromHMM
HSMM ChromHMM
NHEK ChromHMM
NHLF ChromHMM
H1-hESC H3K27ac
K562 P300
K562 H3K27ac
GM12878 H3K27ac
A549 DEX H3K27ac
A549 EtOH H3K27ac
CD20+ H3K27ac
HeLa-S3 H3K27ac
HepG2 H3K27ac
HUVEC H3K27ac
CD14+ H3K27ac
Dnd41 H3K27ac
HMEC H3K27ac
HSMM H3K27ac
HSMMtube H3K27ac
NH-A H3K27ac
NHDF-Ad H3K27ac
NHEK H3K27ac
NHLF H3K27ac
Osteobl H3K27ac
H1-hESC H3K4m3
GM12878 H3K4m3
A549 DEX H3K4m3
A549 EtOH H3K4m3
HeLa-S3 H3K4m3
K562 H3K4m3
HepG2 H3K4m3
HUVEC H3K4m3
CD14+ H3K4m3
Dnd41 H3K4m3
HMEC H3K4m3
HSMM H3K4m3
HSMMtube H3K4m3
NH-A H3K4m3
NHDF-Ad H3K4m3
NHEK H3K4m3
NHLF H3K4m3
Osteobl H3K4m3
Adipose
Adrenal
Brain
Brain_R
Breast
Colon
Foreskin_R
Heart
hLF_r1
hLF_r2
Kidney
Liver
Lung
LymphNode
Ovary
Placenta_R
Prostate
SkeletalMuscle
Testes
Testes_R
Thyroid
WhiteBloodCell
DNase Clusters
ORegAnno
TFBS Conserved
NKI LADs (Tig3)
100 Vert. Cons
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(138)
RepeatMasker
H3K4me3 RawSignal
MRE CpG
MeDIP CpG
MeDIP RawSignal
RNA-seq RawSignal
Smart RawSignal
H1-hESC 1
K562 1
A549 1
HeLa-S3 1
HepG2 1
IMR90 1
MCF-7 1
SK-N-SH 1
Adrenal_BC 1
AG04449 1
AG04450 1
AG09309 1
AG09319 1
AG10803 1
AoSMC 1
BE2_C 1
BJ 1
Brain_BC 1
Breast_BC 1
Caco-2 1
CMK 1
ECC-1 1
Fibrobl 1
GM06990 1
GM12878-XiM 1
GM12891 1
GM12892 1
GM19239 1
GM19240 1
HAEpiC 1
HCF 1
HCM 1
HCPEpiC 1
HCT-116 1
HEEpiC 1
HEK293 St 1
HEK293 UW 1
Hepatocytes 1
HIPEpiC 1
HL-60 1
HMEC 1
HNPCEpiC 1
HPAEpiC 1
HRCEpiC 1
HRE 1
HRPEpiC 1
HSMM 1
HSMM_FSHD 1
HSMMtube 1
HSMMtube_FSHD 1
HTR8svn 1
Jurkat 1
Kidney_BC 1
Ventricle 1
Leukocyte_BC 1
Liver BC 1
LNCaP Andro 1
LNCaP Duke 1
LNCaP UW 1
Lung BC 1
MCF10A Tam 1
MCF10A 1
MCF-7 Cont 1
MCF-7 CTCF 1
Melano 1
Myometr 1
NB4 1
NH-A 1
NHBE 1
NHDF-neo 1
NT2-D1 1
Osteobl 1
Ovcar-3 1
PANC-1 HA 1
PANC-1 UW 1
Pancreas_BC 1
Panislets 1
Pericardium_BC 1
PFSK-1 1
Placenta_BC 1
PrEC 1
ProgFib 1
RPTEC 1
SAEC 1
Skeletal_02 1
Skeletal_7N 1
Skin BC 1
SKMC 1
SK-N-MC 1
SK-N-SH_RA 1
Stomach_BC 1
T-47D DMSO 1
T-47D Estra 1
Testis_BC 1
UB7 1
UCH-1 1
Uterus_BC 1

miR-6130 NR_106746,chr9,-,76368039,77204569

Expression at http://www.genboree.org/EpigenomeAtlasBrowser