

Supplementary Information

- 1. Supplementary Data Legends**
- 2. Supplementary Figures**
- 3. Supplementary Tables**
- 4. Supplementary File 1**
- 5. Supplementary File 2**

SUPPLEMENTARY DATA LEGENDS

Supplementary Figure 1: Pattern of DNA methylation within 5 kb up and downstream of the TSS of all lncRNA, genic lncRNA (lncRNA falling within 1kb upstream of TSS to 1kb downstream of TES of mRNA coding region) and lncRNA genes falling outside genic regions (non genic), in Brain germinal matrix tissue (A), Brain frontal cortex (B), H1 cell (C) and PBMCs (D).

Supplementary Figure 2: Association of average histone modification density around TSS with gene expression in H1 cells. (A), (B) represents the H3K4me3 density around TSS of highly / lowly expressed protein coding and lncRNA genes respectively. (C), (D) represents the H3K36me3 density around TSS of highly / lowly expressed protein coding and lncRNA genes respectively. (E), (F) represents the H3K27me3 density around TSS of highly / lowly expressed protein coding and lncRNA genes respectively. (G), (H) represents the H3K9me3 density around TSS of highly / lowly expressed protein coding and lncRNA genes respectively. Peak summit count in 100bp sliding window was normalized by dividing count with total number of genes in that category. The plots were further smoothed by taking a moving average of 5.

Supplementary Figure 3: Average density of H3K4me3 and H3k36me3 density around the TSS of various sub categories of lncRNA genes. (A) Represents the H3K4me3 density around TSS four different sub categories of lncRNA genes in different cell and tissue types. (B) Represents the H3K36me3 density around TSS four different sub categories of lncRNA genes in different cell and tissue types.

Supplementary Figure 4: Distribution of ORegAnno regulatory sites around 5kb upstream and downstream from the TSS of protein coding and lncRNA genes.

Supplementary Table 1: Showing the details of the raw and processed data.

Supplementary Table 2: Distribution of the H3K4me3, H3K9me3 and H3K27me3 histone modifications across the CGI present at the TSS of protein coding and lncRNA genes. Each value is the ratio of the count in each interval by total number of genes in that category. (A)

Count of histone modification in each category. (B) Normalised distribution of histone modification in each category.

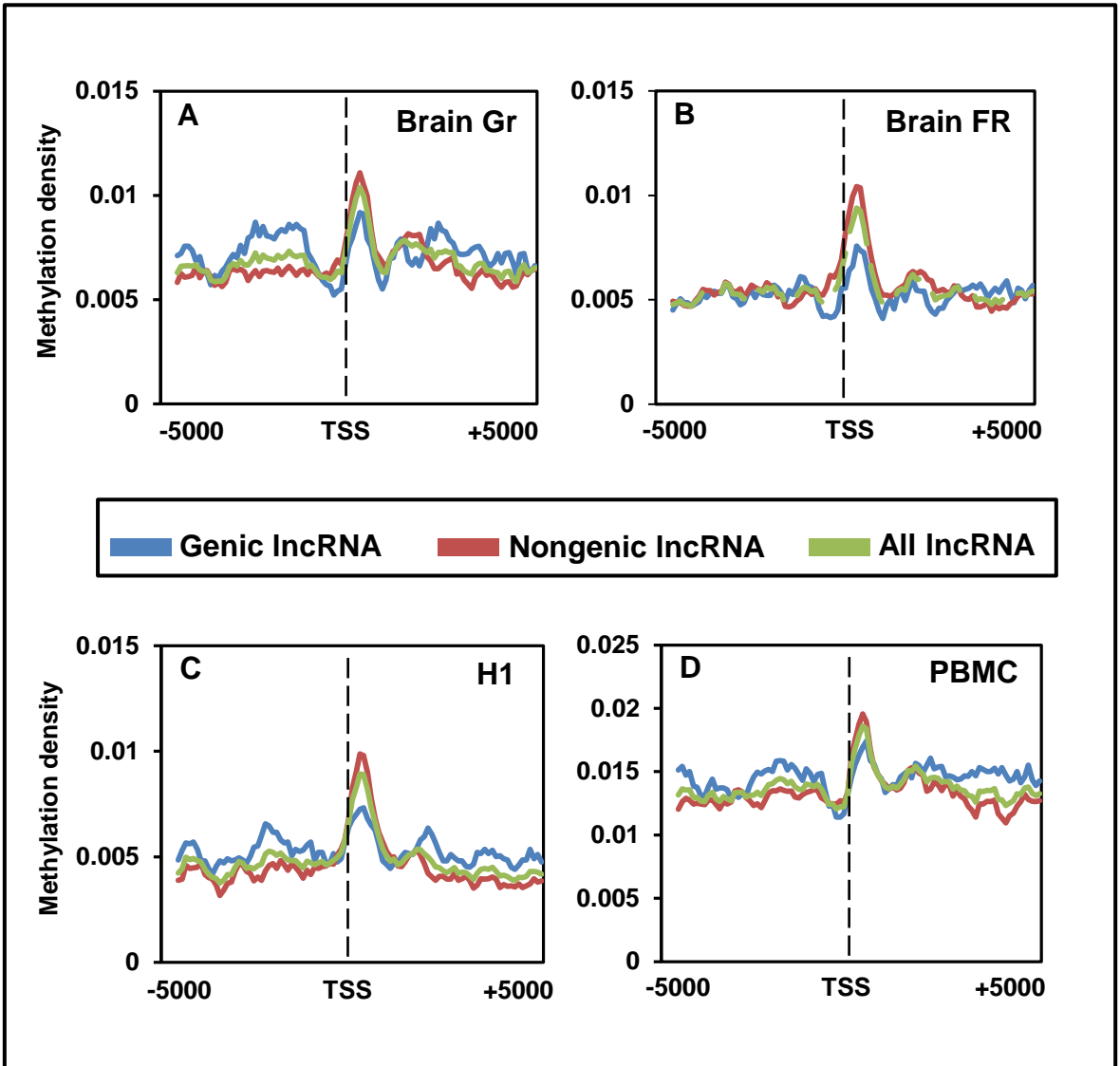
Supplementary Table 3: Global distribution of the H3K4me3, H3K9me3, H3K27me3, H3K36me3 histone modifications and DNA methylation around ± 2 Kb of the TSS of protein coding and lncRNA genes in all datasets analysed. Each value represents the ratio of the number of genes having the modification to the total number of genes. (A) Count of histone modification in each category. (B) Percentage distribution of histone modification in each category. (C) Number of genes having at least one epigenetic mark in three datasets.

Supplementary Table 4: Details of the state of the datasets present in the NIH Roadmap Epigenomics Project Data Listings downloaded at the time of data analysis.

Supplementary File 1: Venn diagram showing the coexistence of two or more of these histone modifications at 2kb up or down stream of TSS of lncRNA and mRNA coding regions in different cell and tissue types.

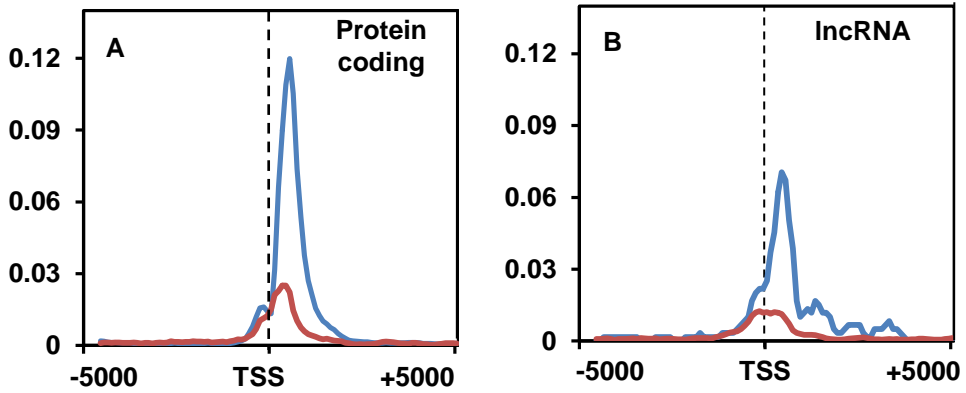
Supplementary File 2: File containing UCSC images showing mapping of the epigenetic marks around TSS of a few regions including housekeeping genes and some lncRNA genes.

Supplementary Figure 1



Supplementary Figure 2 A-D

H3K4me3

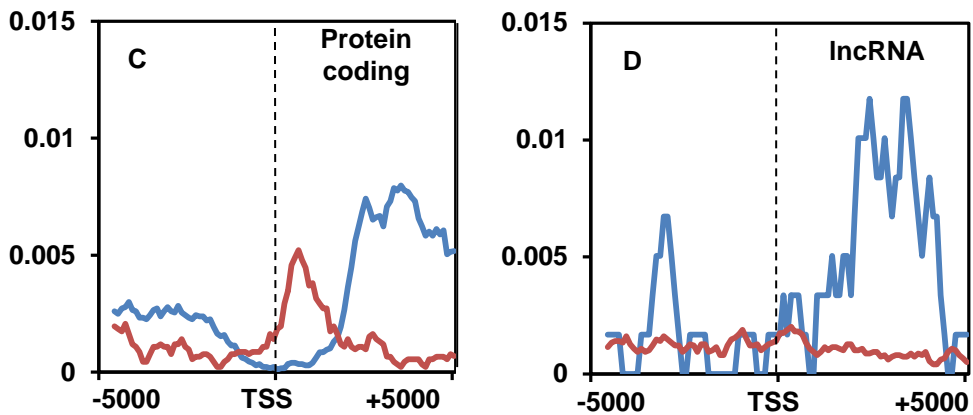


Highly Expressed



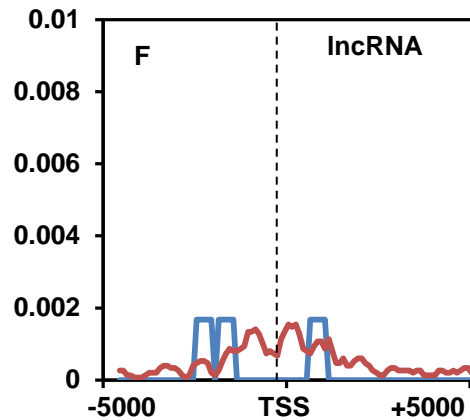
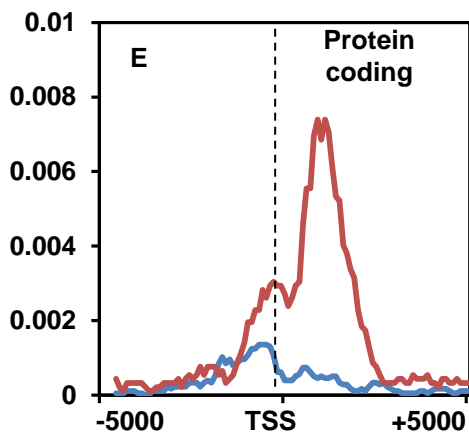
Lowly Expressed

H3K36me3



Supplementary Figure 2 E-G

H3K27me3

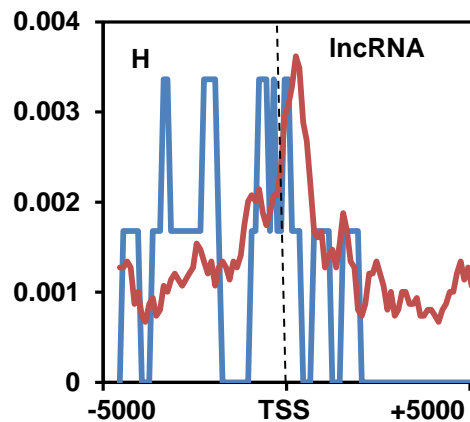
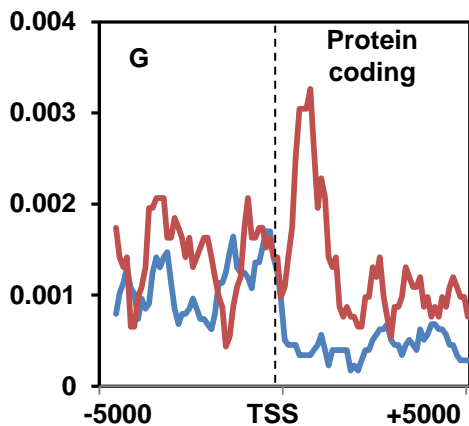


Highly Expressed



Lowly Expressed

H3K9me3

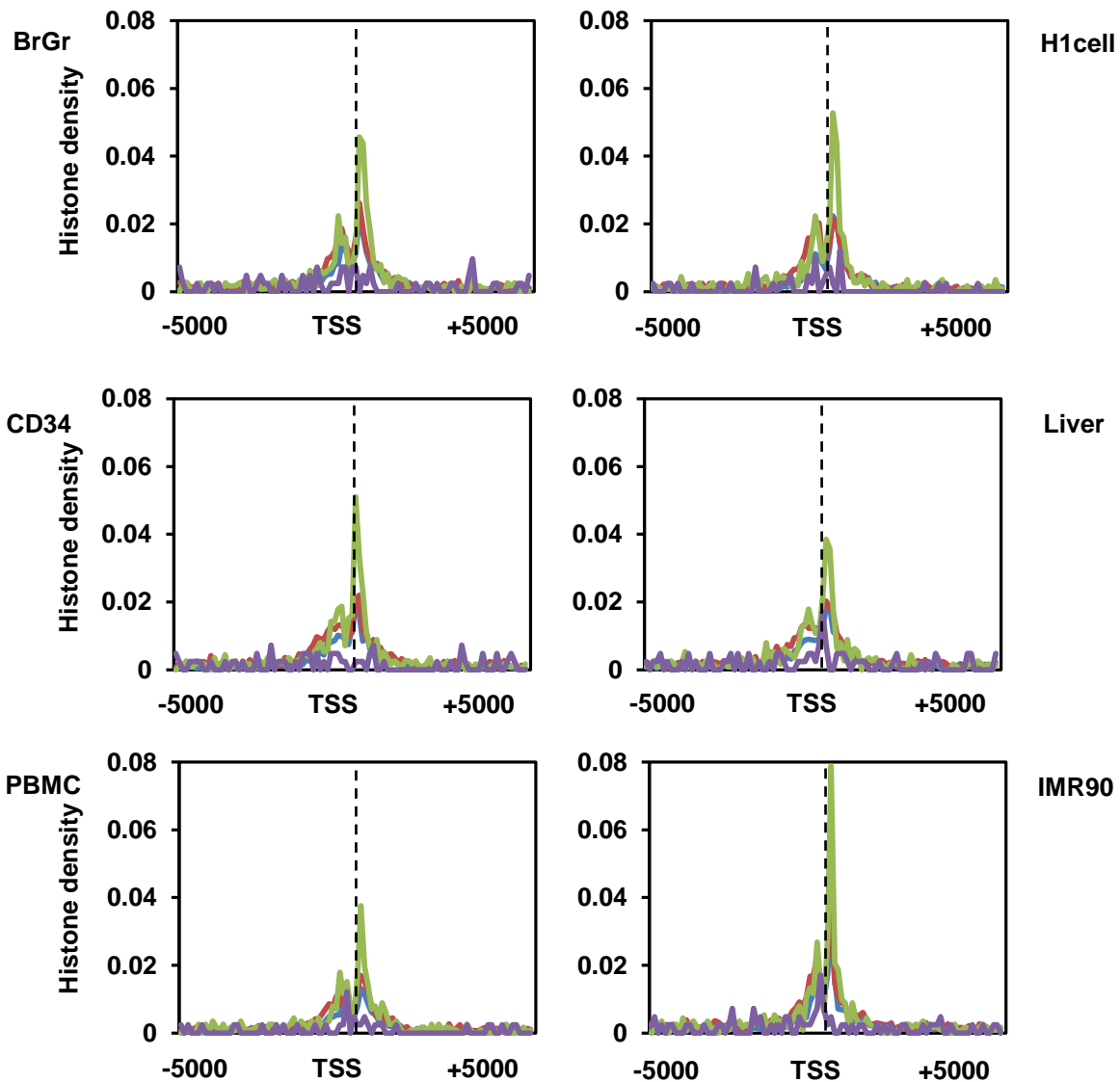


Supplementary Figure 3A

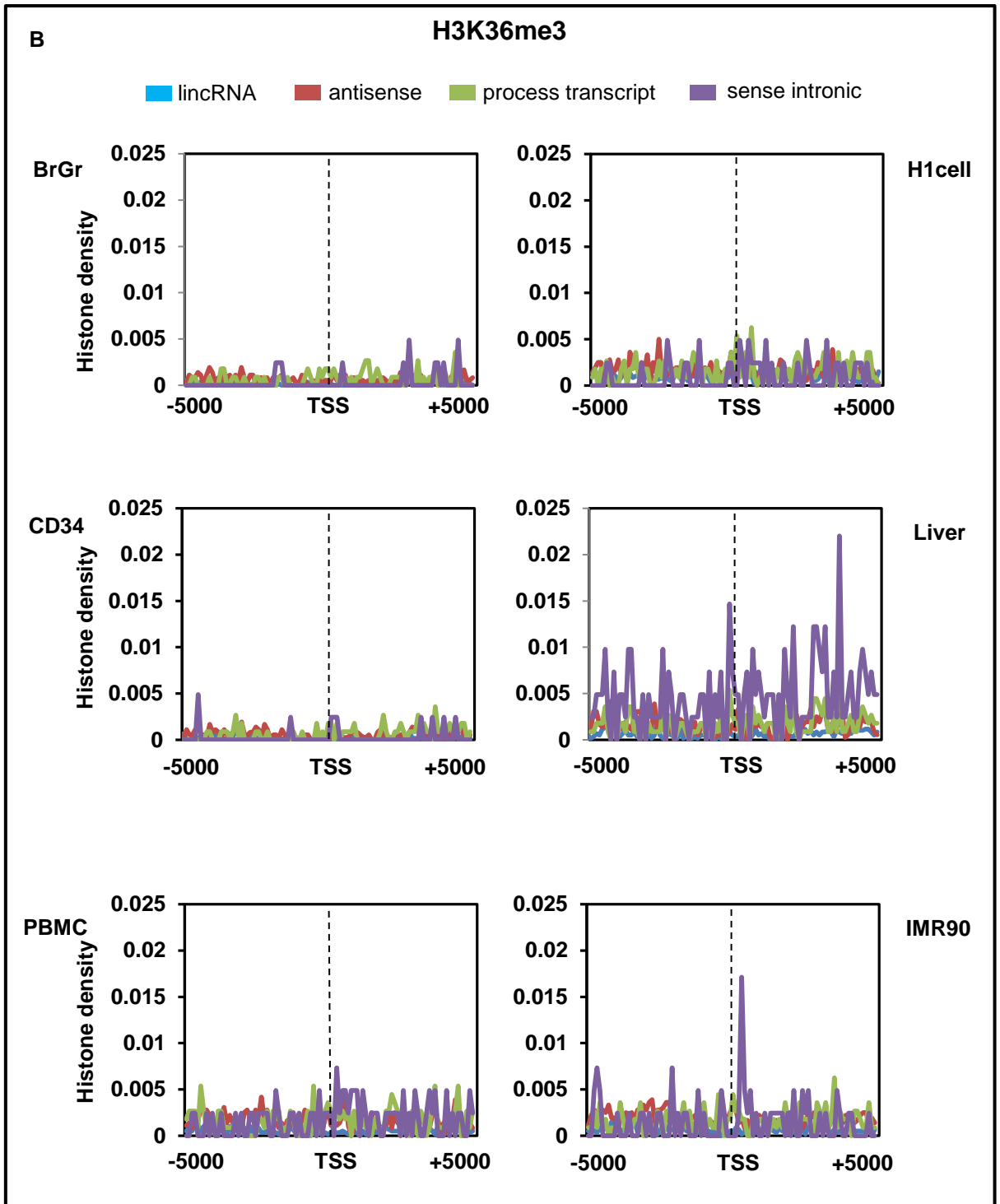
A

H3K4me3

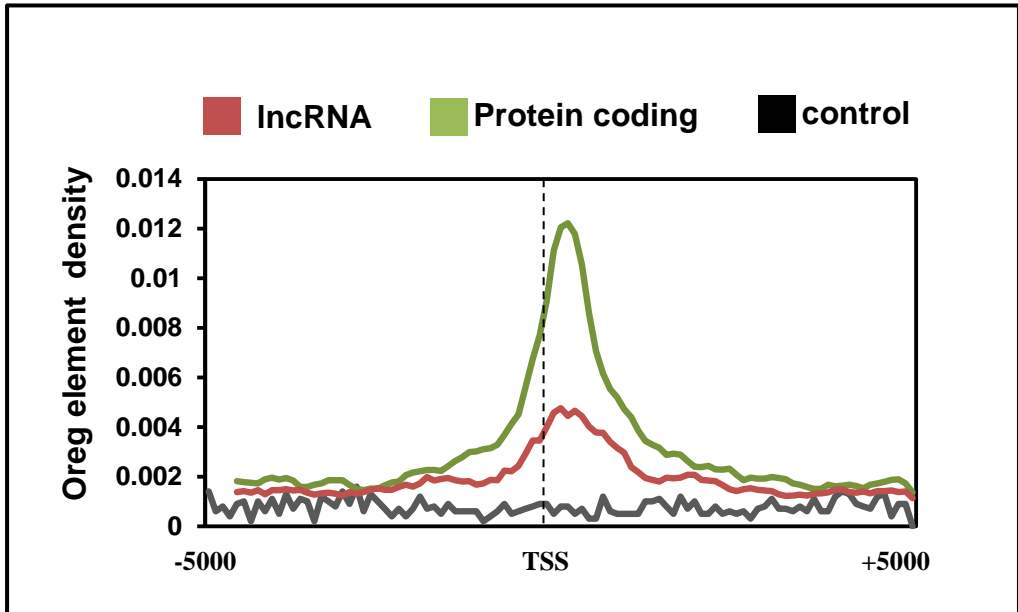
■ lincRNA ■ antisense ■ process transcript ■ sense intronic



Supplementary Figure 3B



Supplementary Figure 4



Supplementary Table 1.

Datasets	Number of entries
hg19_refseq_genes	40845
hg19_refseq_exons	394627
hg19_refseq_introns	353782
Human CpG islands	28691
OregAnno data points	23089
H1_expressed_gene_fpkms	51653
Brain_expressed_gene_fpkms	51652
gencode v9 lncRNA transcript count	18878
gencode v9 lncRNA exon	67935
gencode v9 lncRNA gene	11004
mRNA coding region count	20012
Histone_Br_H3K4me3	24888
Histone_H1_H3K4me3	21498
Histone_H1_H3K27me3	4788
Histone_H1_H3K9me3	43128
Histone_H1_H3K36me3	25996
Histone_IMR90_H3K4me3	33148
Histone_IMR90_H3K27me3	24929
Histone_IMR90_H3K9me3	64601
Histone_IMR90_H3K36me3	25488
Histone_BrGr_H3K4me3	25417
Histone_BrGr_H3K27me3	13468
Histone_BrGr_H3K9me3	33568
Histone_BrGr_H3K36me3	6810
Histone_CD34_H3K4me3	25363
Histone_CD34_H3K27me3	3871
Histone_CD34_H3K9me3	27573
Histone_CD34_H3K36me3	5240
Histone_Liver_H3K4me3	32781
Histone_Liver_H3K27me3	4218
Histone_Liver_H3K9me3	54372
Histone_Liver_H3K36me3	22449
Histone_PBMC_H3K4me1	18172
Histone_PBMC_H3K27me3	33094
Histone_PBMC_H3K9me3	27243
Histone_PBMC_H3K36me3	26657
Methylation_Brain_Frontal_Cortex	128607
Methylation_Brain_Germinal_Matrix	163597
Methylation_PBMC	354811
Methylation_H1	107372
H1_mRNA_coding_region (Mean + SD)	3532
H1_mRNA_coding_region (Mean - SD)	1839
H1_lncRNA_coding_region (Mean + SD)	119
H1_lncRNA_coding_region (Mean - SD)	2983
Br_mRNA_coding_region (Mean + SD)	4624
Br_mRNA_coding_region (Mean - SD)	1415
Br_lncRNA_coding_region (Mean + SD)	171
Br_lncRNA_coding_region (Mean - SD)	3665

Supplementary Table 2. (A)

	PROTEIN CODING GENES								LncRNA GENES							
	With CGI				Without CGI				With CGI				Without CGI			
	H3 K4 me3	H3 K9 me3	H3 K27 me3	H3 K36 me3	H3 K4 me3	H3 K9 me3	H3 K27 me3	H3 K36 me3	H3 K4 me3	H3 K9 me3	H3 K27 me3	H3 K36 me3	H3 K4 me3	H3 K9 me3	H3 K27 me3	H3 K36 me3
BRAIN GERMINAL	11963	168	8292	163	1202	124	527	75	2128	59	1386	52	582	227	154	98
CD34	11730	287	906	125	1464	139	146	50	2065	75	166	31	573	238	69	71
H1	12197	597	1738	959	1062	313	167	204	2156	174	345	252	402	480	84	323
IMR90	11724	1319	1640	665	1402	527	643	219	2059	333	334	148	770	694	430	329
LIVER	11097	384	313	249	1663	295	97	195	1907	94	61	74	719	427	290	60
PBMC	9301	217	2939	232	994	142	835	162	1537	69	58	623	360	217	330	488

Number of genes having the histone modifications.

(B)

	H3K4me3	H3K9me3	H3K27me3	H3K36me3	
BRAIN GERMINAL	3.5711	0.0382	4.6713	0.1816	Protein Coding with CGI
CD34	3.5090	0.0812	1.7758	0.1810	
H1	4.3364	0.1064	2.8428	0.2799	
IMR90	2.6835	0.1589	0.5043	0.1980	
LIVER	2.5902	0.0543	0.5632	0.0842	
PBMC	3.9627	0.0610	0.6738	0.0660	
BRAIN GERMINAL	0.6922	0.0541	0.5727	0.1612	Protein Coding without CGI
CD34	0.8449	0.0738	0.5521	0.1397	
H1	0.7231	0.1062	0.5105	0.1149	
IMR90	0.6191	0.1194	0.3775	0.1258	
LIVER	0.7425	0.0794	0.3366	0.1271	
PBMC	0.8006	0.0763	0.3693	0.0890	
BRAIN GERMINAL	3.1228	0.0656	3.8385	0.2848	LncRNA with CGI
CD34	3.0368	0.1015	1.5995	0.2207	
H1	3.7407	0.1505	2.6876	0.3616	
IMR90	2.3169	0.1923	0.4997	0.2166	
LIVER	2.1699	0.0645	0.5394	0.1230	
PBMC	3.1548	0.0945	0.0654	0.8717	
BRAIN GERMINAL	0.2751	0.0812	0.1374	0.1729	LncRNA without CGI
CD34	0.2714	0.1037	0.2142	0.1627	
H1	0.2247	0.1337	0.2108	0.1493	
IMR90	0.2791	0.1291	0.2072	0.1551	
LIVER	0.2635	0.0944	0.8261	0.0321	
PBMC	0.2380	0.0957	0.1198	0.2200	

Each value = (Number of entries in a category / Total number of entries) x 100000

Supplementary Table 3. (A)

	PROTIEN CODING GENES				LncRNA GENES			
	H3K4me3	H3K9me3	H3K27me3	H3K36me3	H3K4me3	H3K9me3	H3K27me3	H3K36me3
BRAIN GERMINAL	13165	292	8819	238	2710	286	1540	150
CD34	13194	426	1052	175	2639	313	235	102
H1	13259	910	1905	1163	2558	654	429	575
IMR90	13126	1846	2283	884	2829	1027	764	477
LIVER	12760	679	410	444	2626	521	121	364
PBMC	10295	359	3774	394	1897	286	1111	388

Number of genes having the histone modification

(B)

	PROTIEN CODING GENES				LncRNA GENES			
	H3K4me3	H3K9me3	H3K27me3	H3K36me3	H3K4me3	H3K9me3	H3K27me3	H3K36me3
BRAIN GERMINAL	65.79	1.46	44.07	1.19	24.63	2.60	13.99	1.36
CD34	65.93	2.13	5.26	0.87	23.98	2.84	2.14	0.93
H1	66.26	4.55	9.52	5.81	23.25	5.94	3.90	5.23
IMR90	65.59	9.22	11.41	4.42	25.71	9.33	6.94	4.33
LIVER	63.76	3.39	2.05	2.22	23.86	4.73	1.10	3.31
PBMC	51.44	1.79	18.86	1.97	17.24	2.60	10.10	3.53

Percentage = (Number of entries in a category / Total number of entries) x 100

(C)

	PROTIEN CODING GENES	LncRNA GENES
BRAIN GERMINAL	15634	5357
H1	15387	4771
PBMC	16382	6974

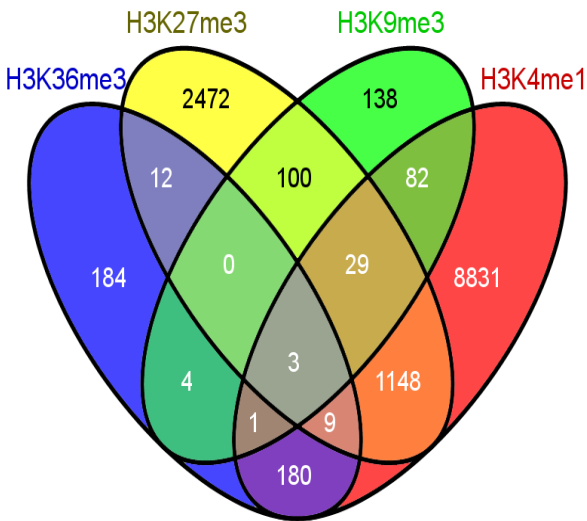
Number of genes having at least one epigenetic mark (DNA Methylation, H3K4me3, H3K9me3, H3K27me3 and H3K36me3) in \pm 2kb of the TSS

Supplementary Table 4.

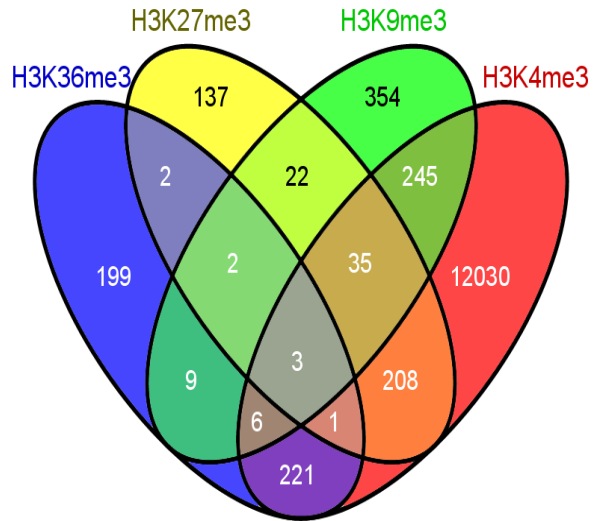
Dataset in NIH Roadmap Epigenomics Project Data Listings	Criteria based on which data was taken or left (based on Embargo date, presence of H3K4me3, H3K9me3, H3K27me3, H3K36me3 modification at the time of data analysis)
Embryonic stem cells	H1 cells taken as one of the representative data
CD34 primary cells	Taken
Adipose	Partial dataset available – Not taken
Adrenal gland	Partial dataset available – Not taken
Bone marrow	Derived cells - Not taken
Brain Tissue	One representative data set taken
Small Intestine regions	Partial dataset available – Not taken
Kidney	Partial dataset available – Not taken
Muscle cells	Partial dataset available – Not taken
PBMCs	Taken
Liver	Taken
Differentiated cells	IMR90 cells taken as one of the representative data

Supplementary File 1

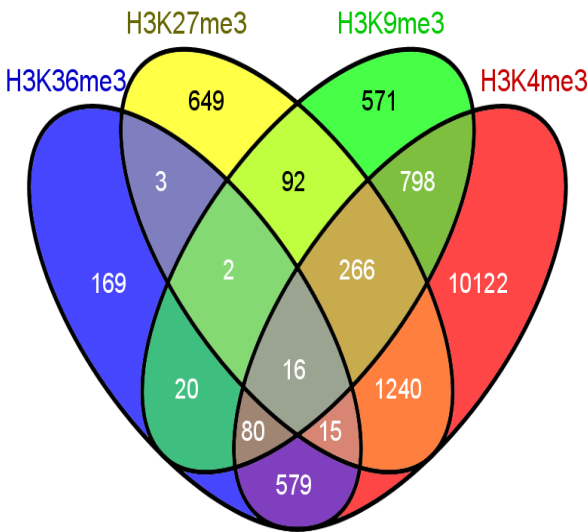
PBMC: co-occurrence of marks in protein coding gene



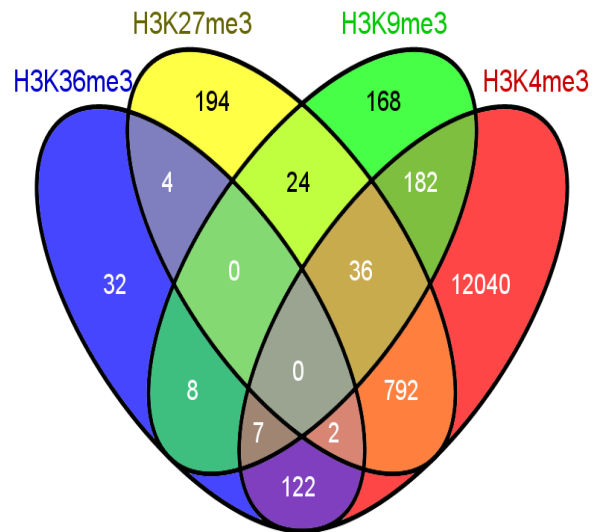
Liver: co-occurrence of marks in protein coding gene



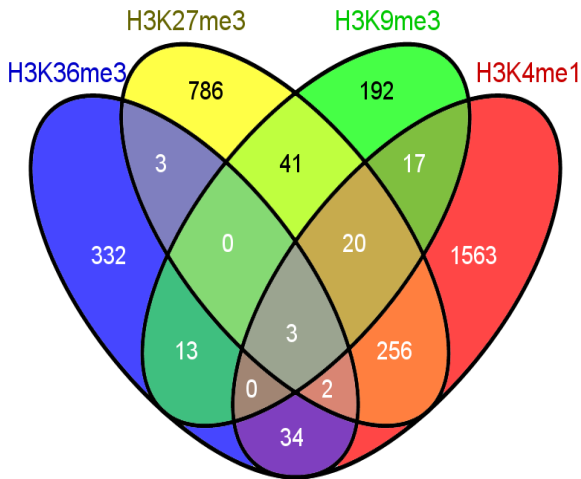
IMR90: co-occurrence of marks in protein coding gene



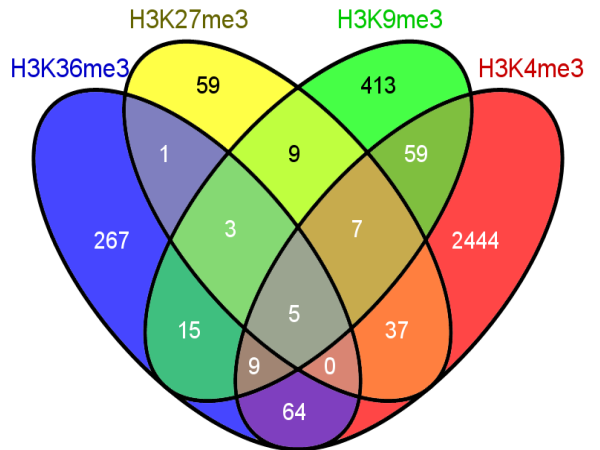
CD34: co-occurrence of marks in protein coding gene



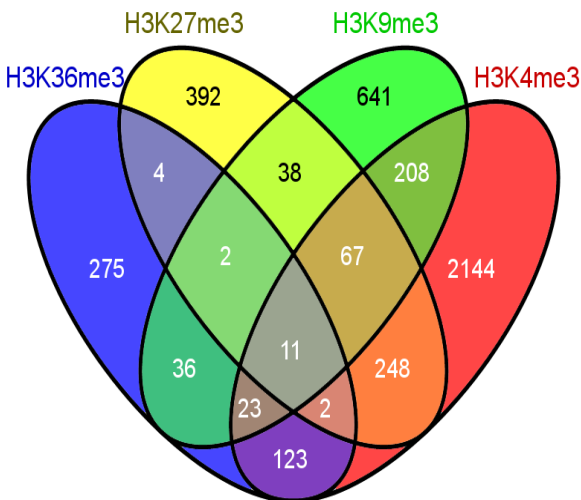
PBMC: co-occurrence of marks in lncRNA gene



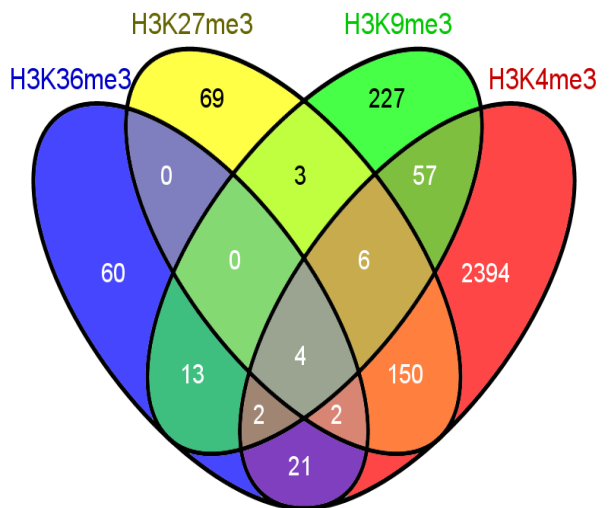
Liver: co-occurrence of marks in lncRNA gene



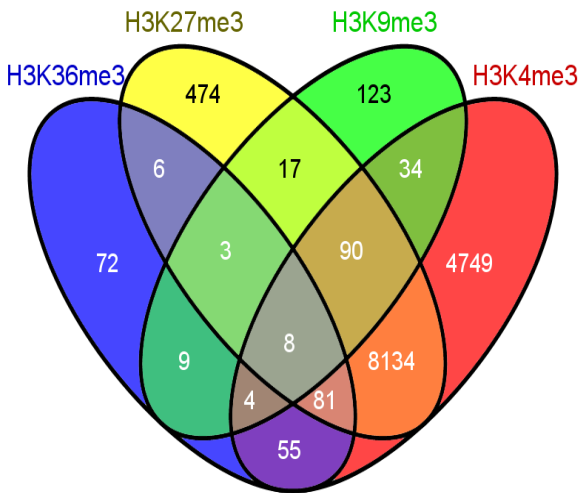
IMR90: co-occurrence of marks in lncRNA gene



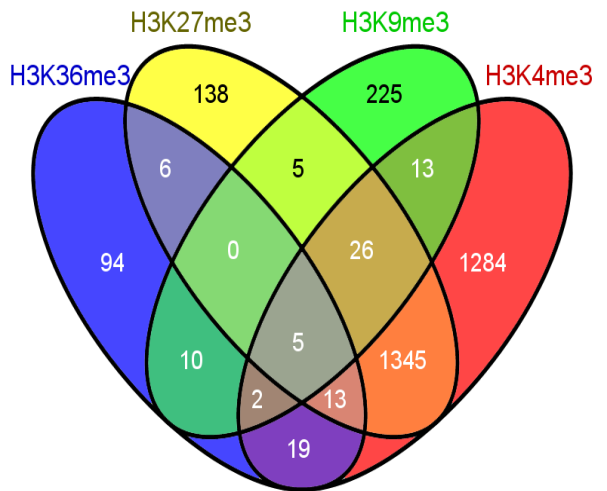
CD34: co-occurrence of marks in lncRNA gene



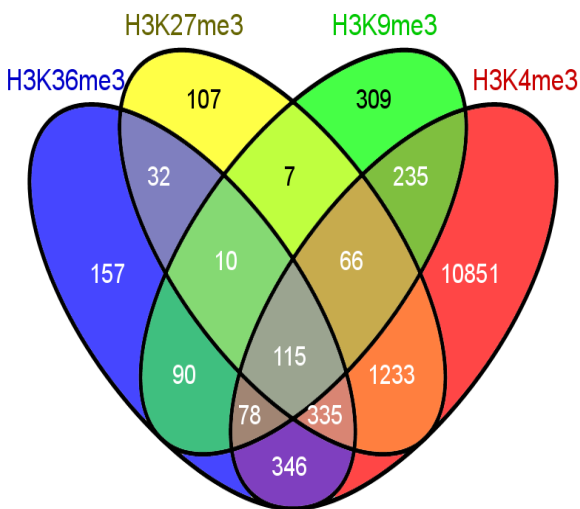
Brain germinal: co-occurrence of marks in protein coding gene



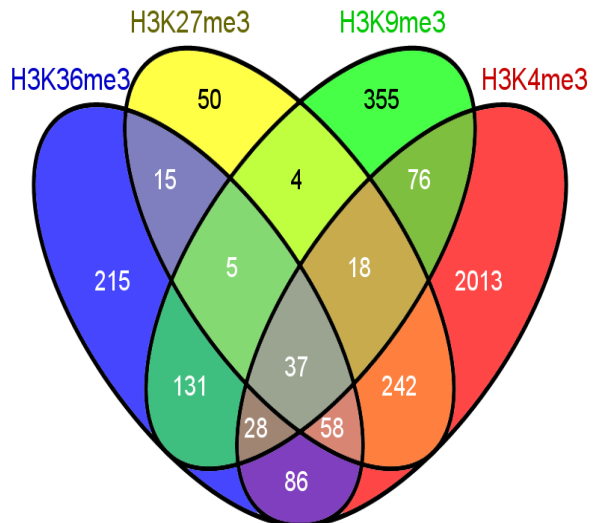
Brain germinal: co-occurrence of marks in lncRNA gene



H1 cell line: co-occurrence of marks in protein coding gene

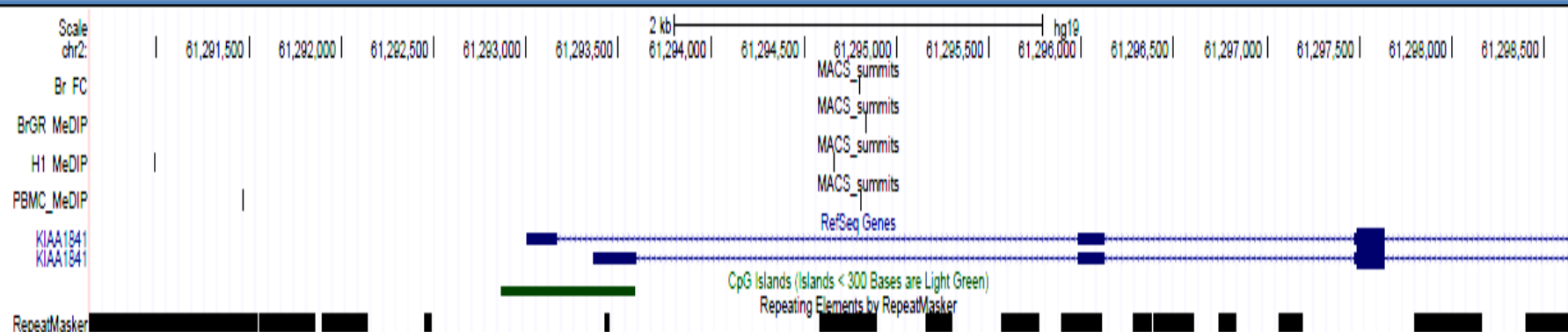
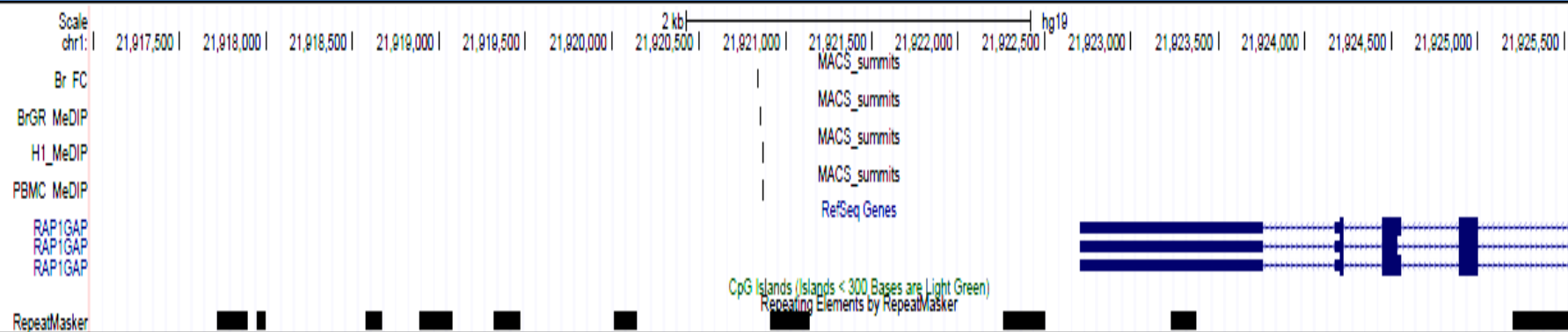
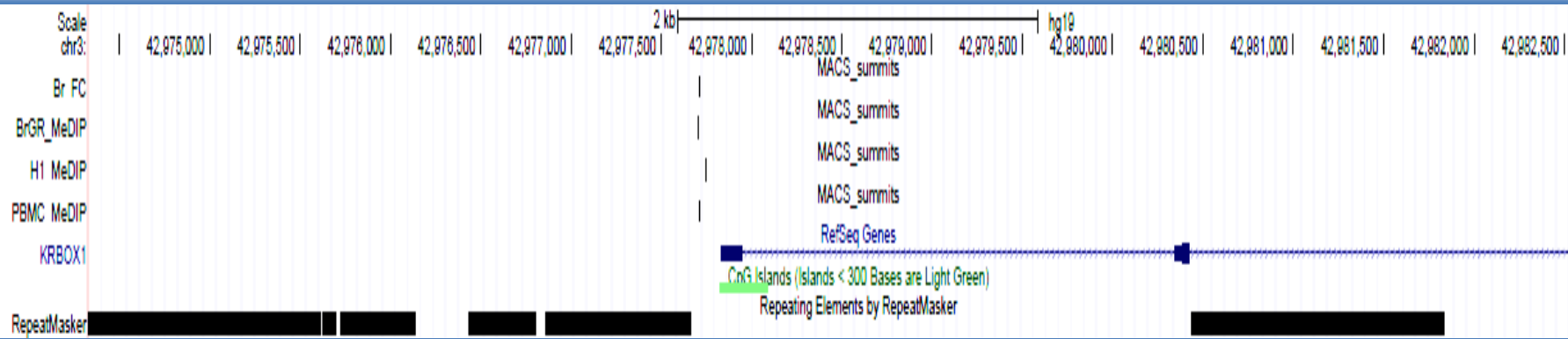


H1 cell line: co-occurrence of marks in lncRNA gene

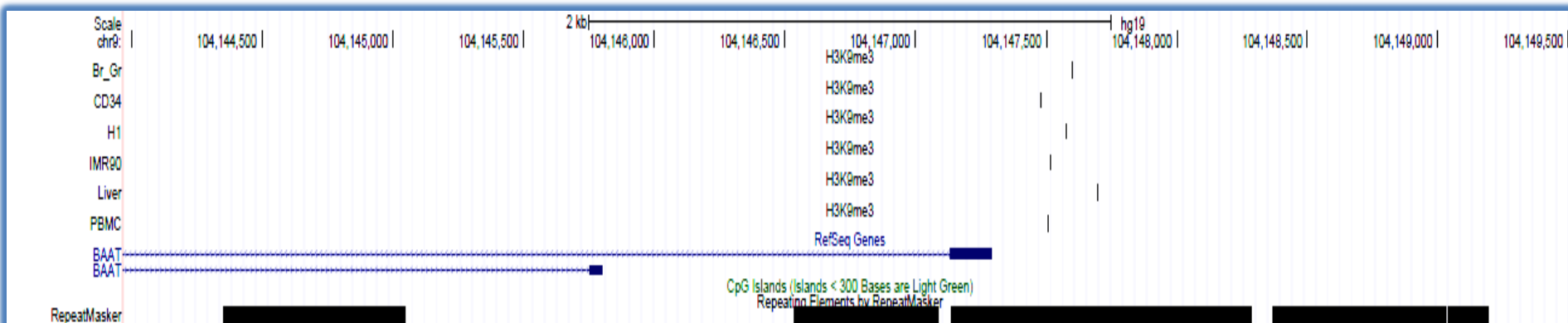
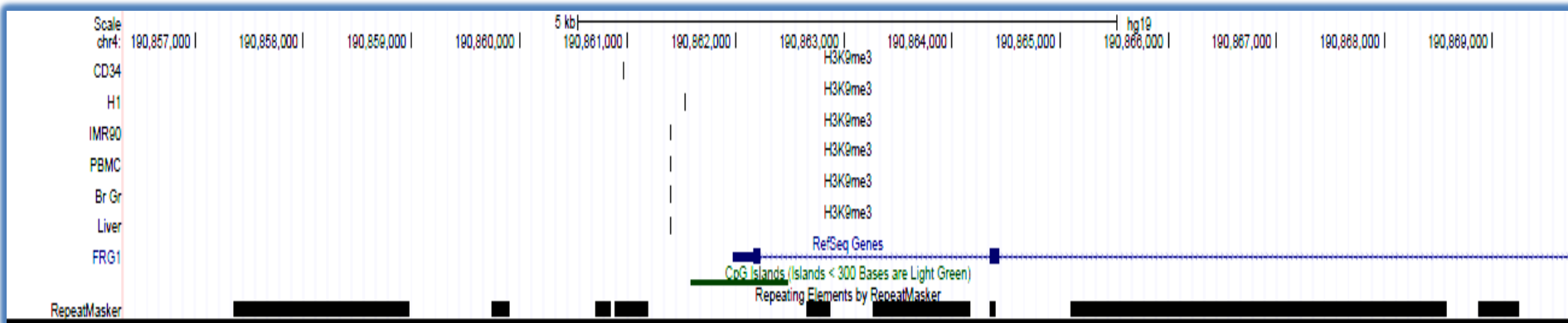
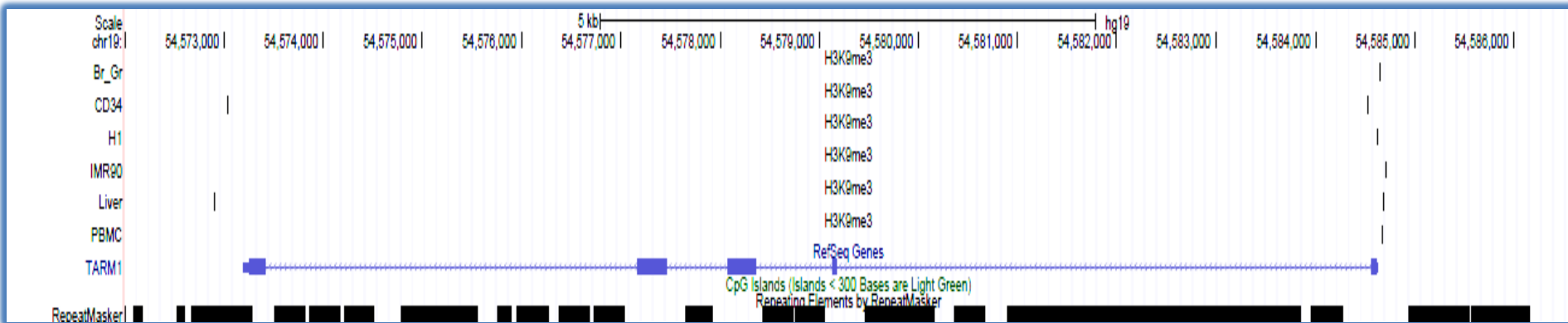


Supplementary File 2

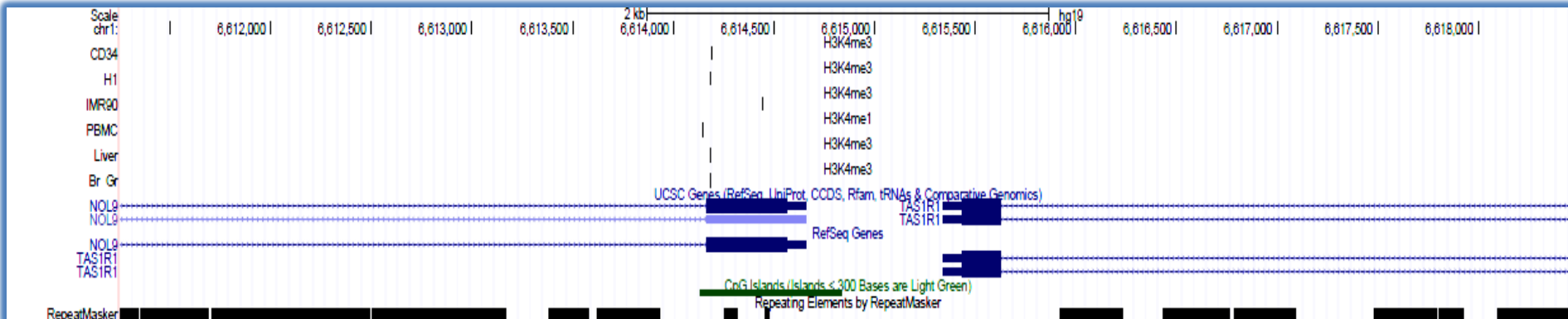
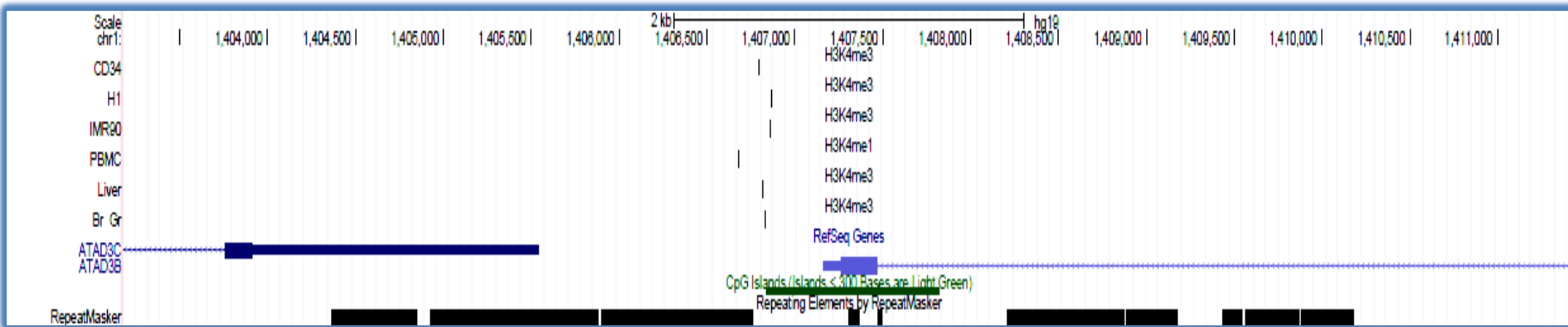
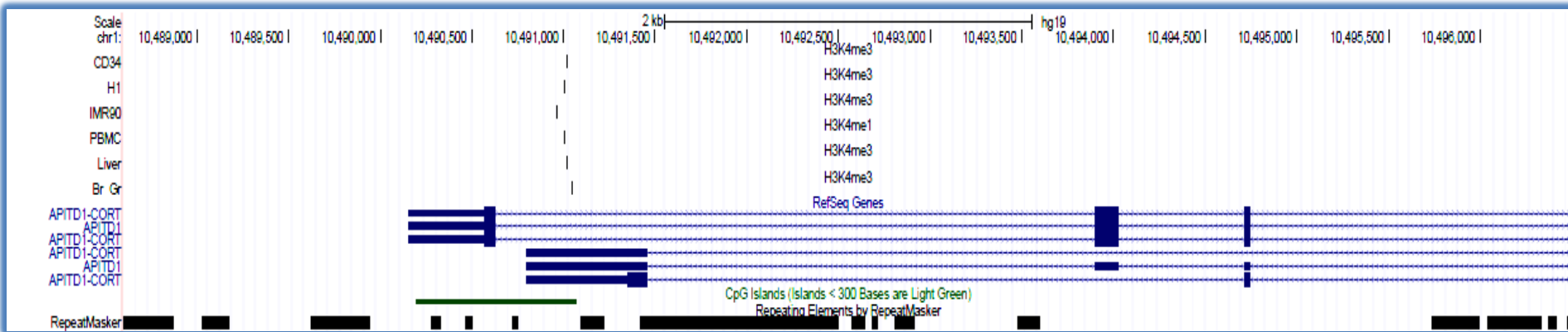
mRNA coding genes having similar methylation pattern across different cell/tissue types



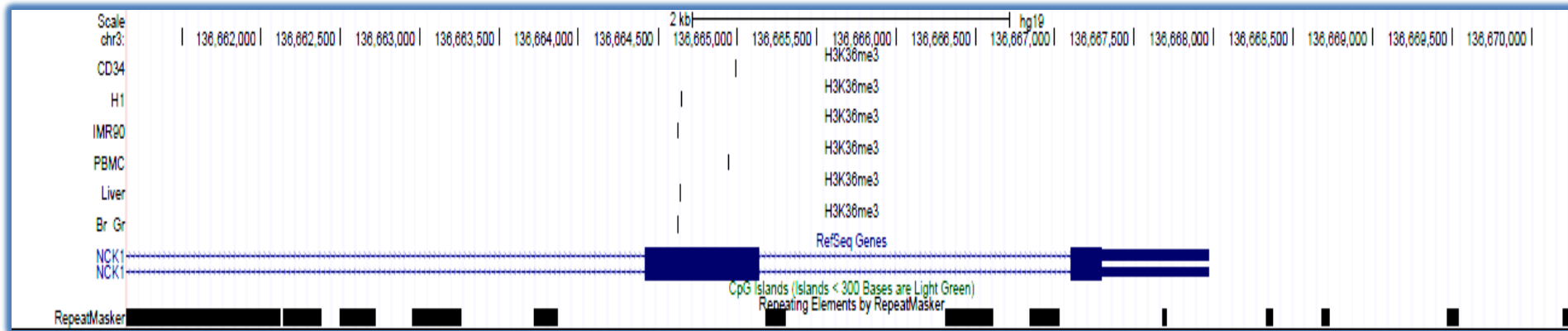
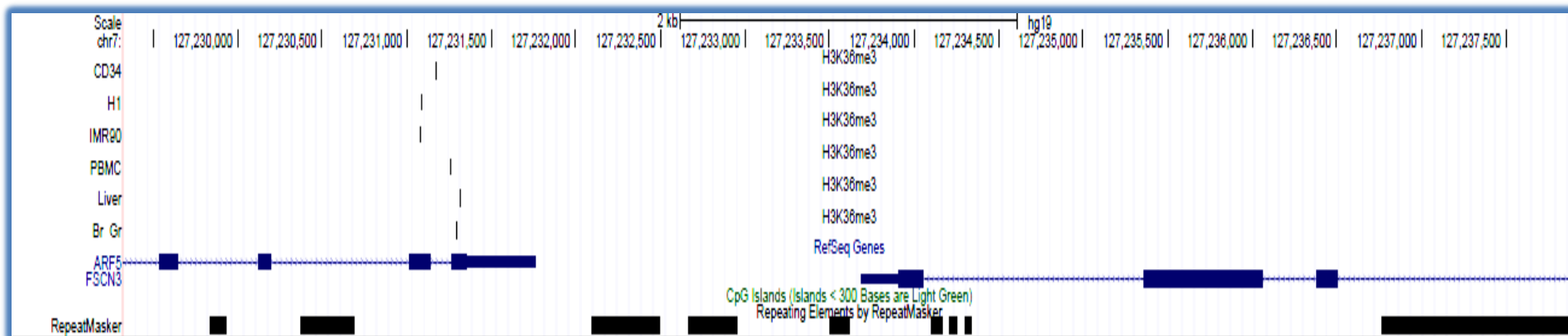
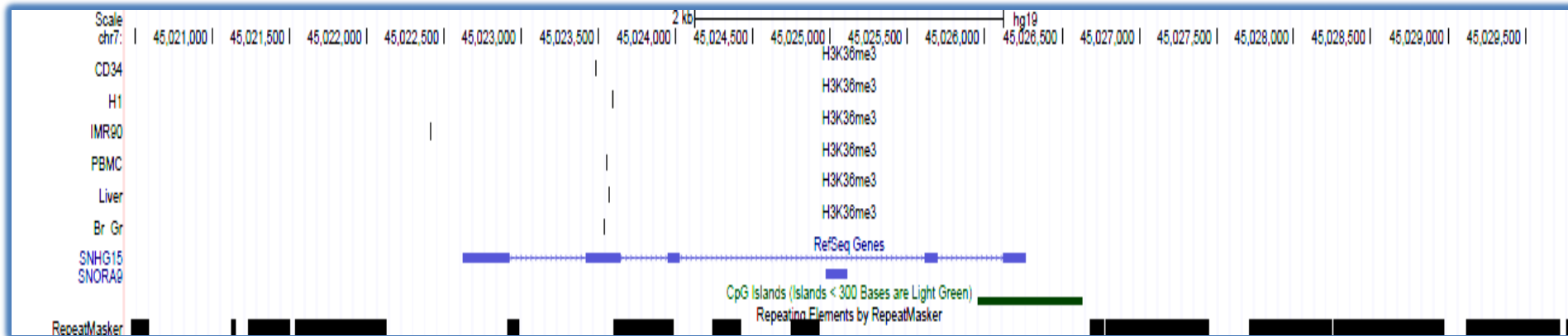
mRNA coding genes having similar H3K9me3 pattern across different cell/tissue types



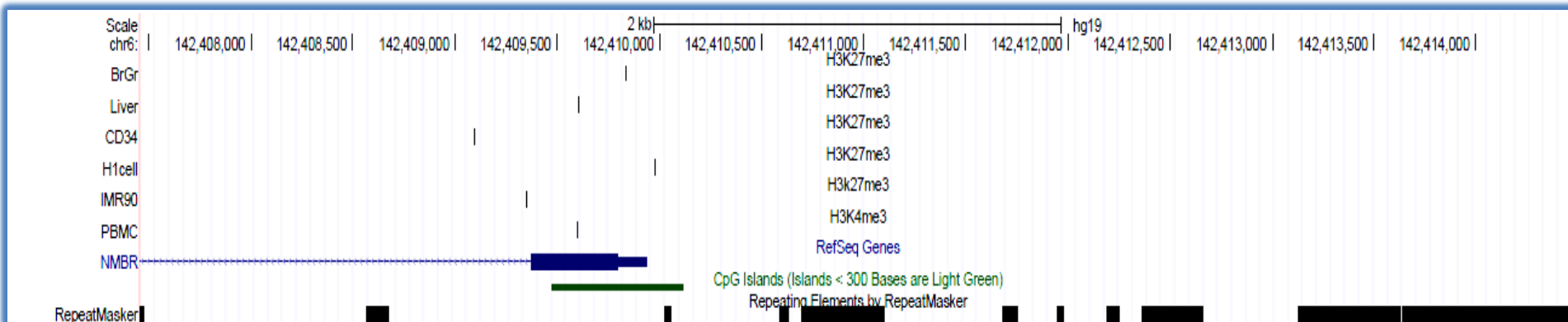
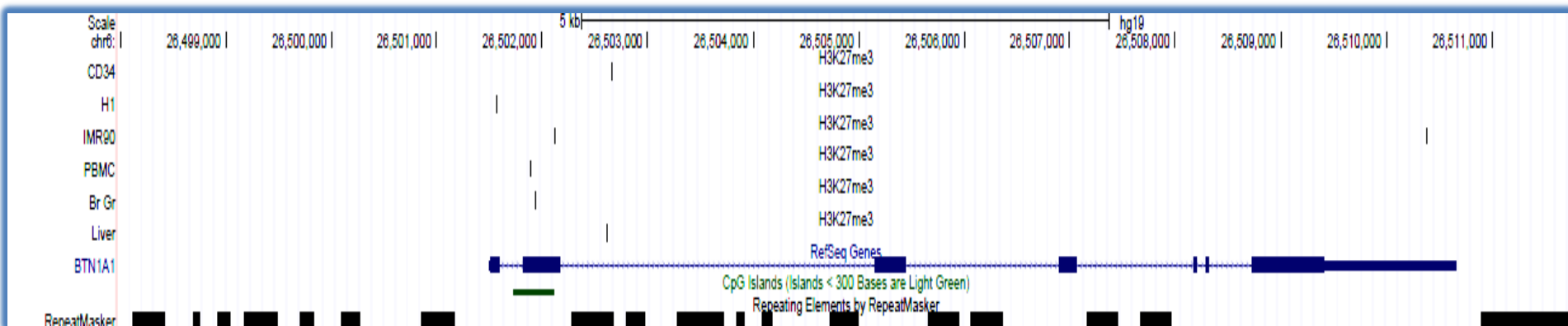
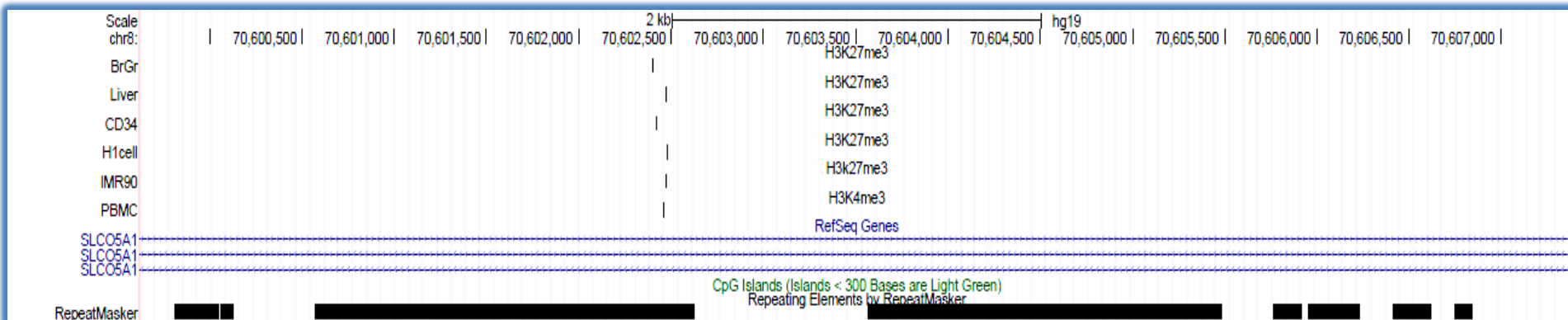
mRNA coding genes having similar H3K4me3 pattern across different cell/tissue types



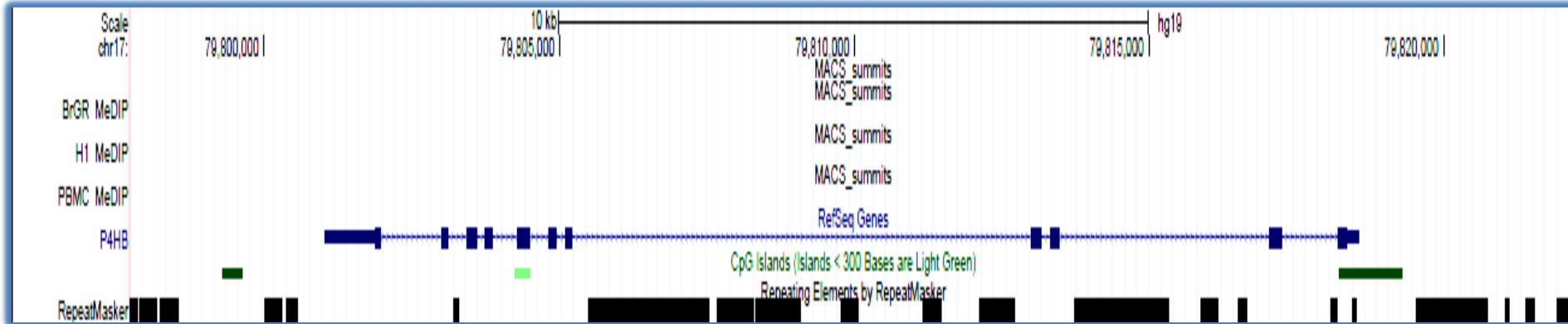
mRNA coding genes having similar H3K36me3 pattern across different cell/tissue types



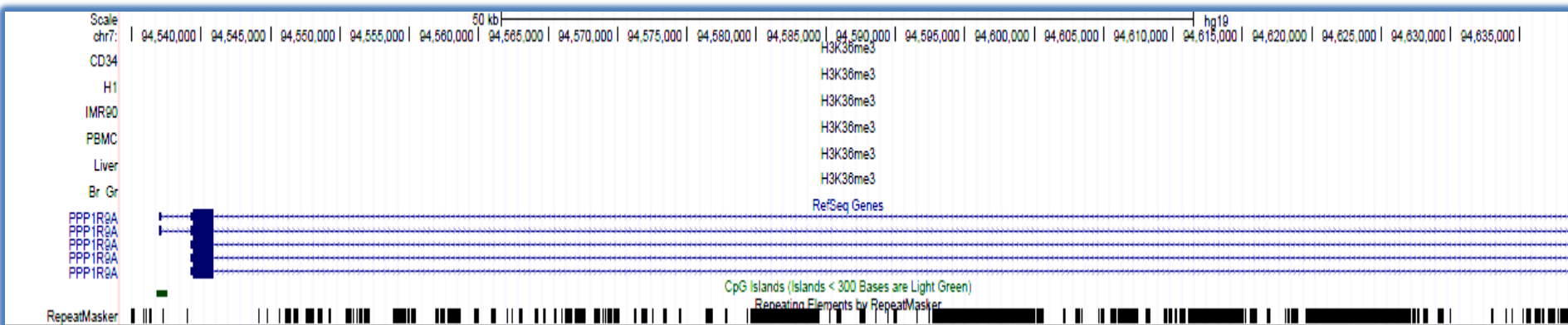
mRNA coding genes having similar H3K27me3 pattern across different cell/tissue types



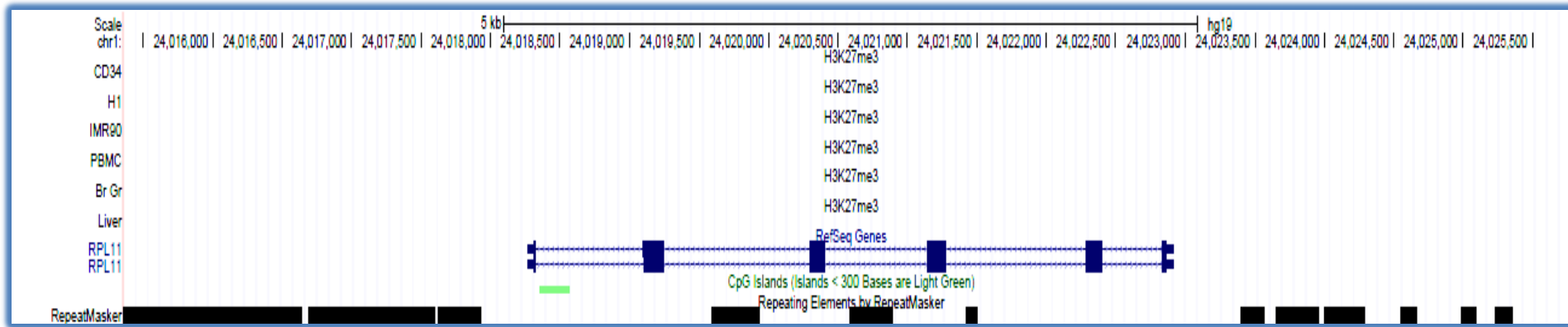
mRNA coding genes not having methylation marks across different cell/tissue types



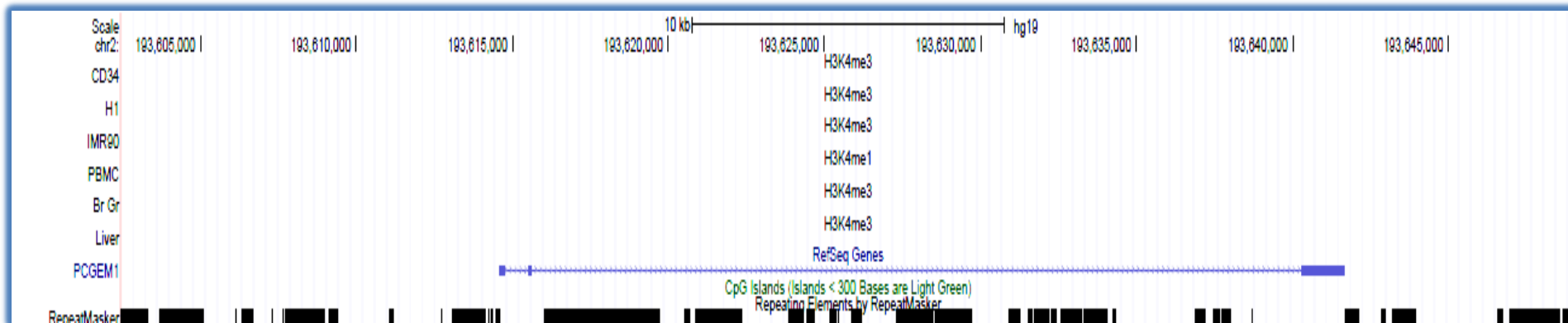
mRNA coding genes not having H3K36me3 marks across different cell/tissue types



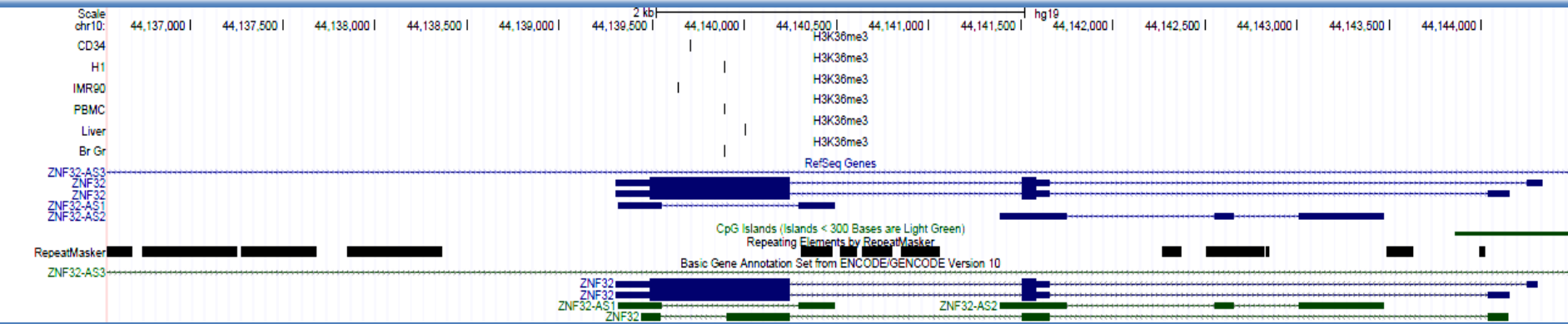
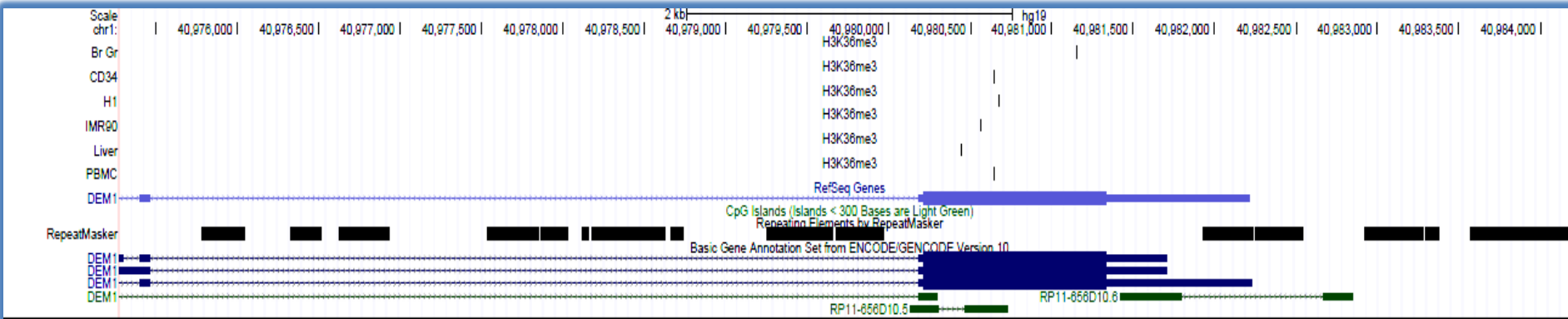
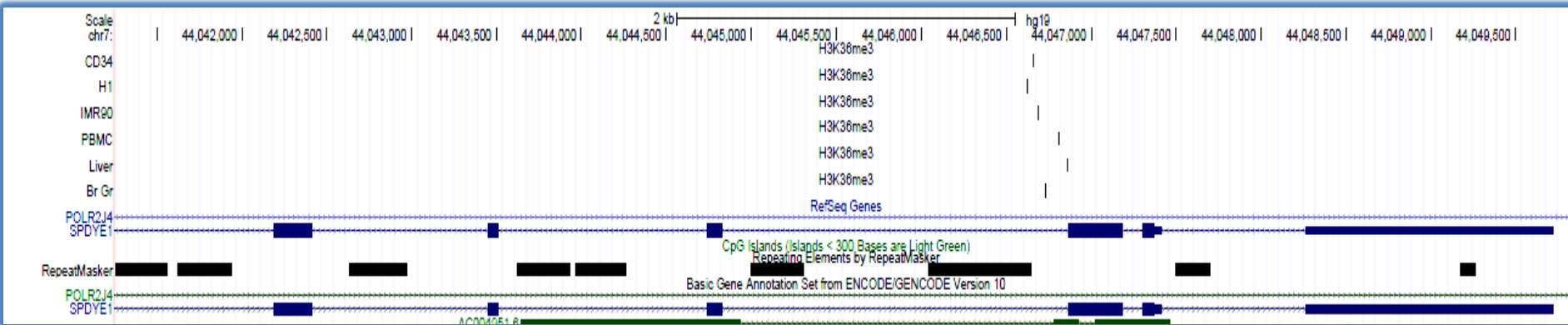
mRNA coding genes not having H3K27me3 marks across different cell/tissue types



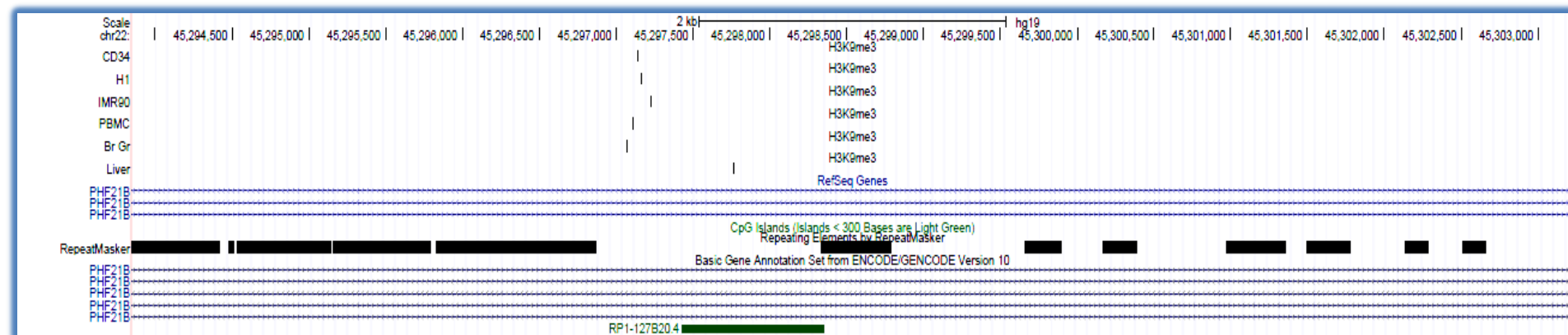
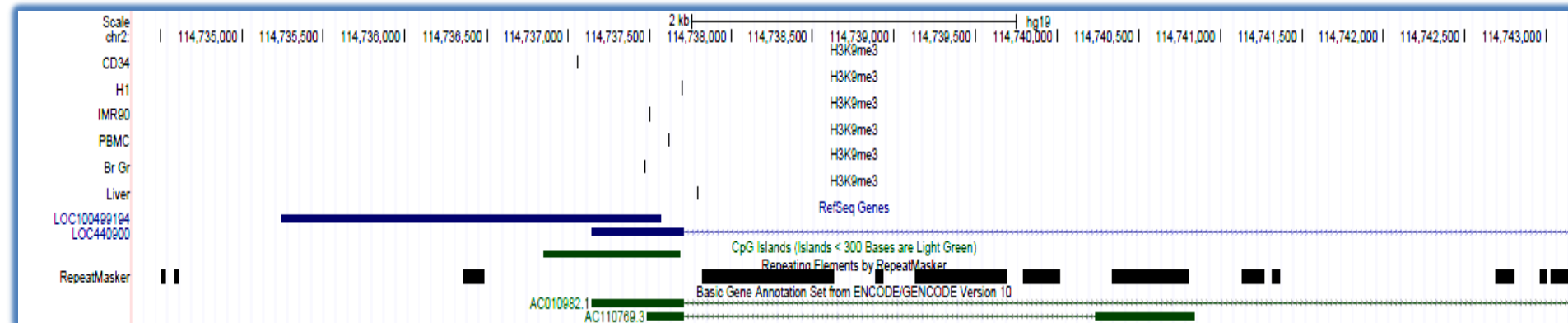
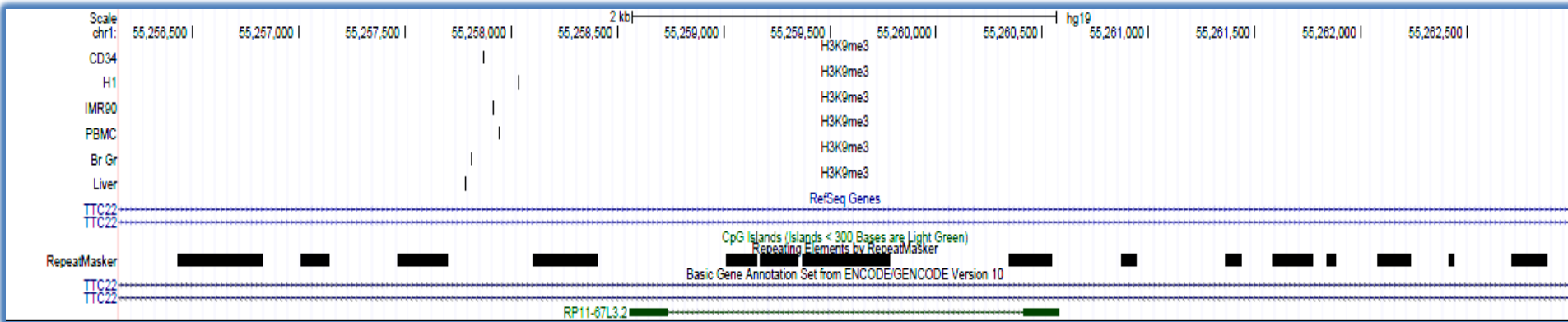
mRNA coding genes not having H3K4me3 marks across different cell/tissue types



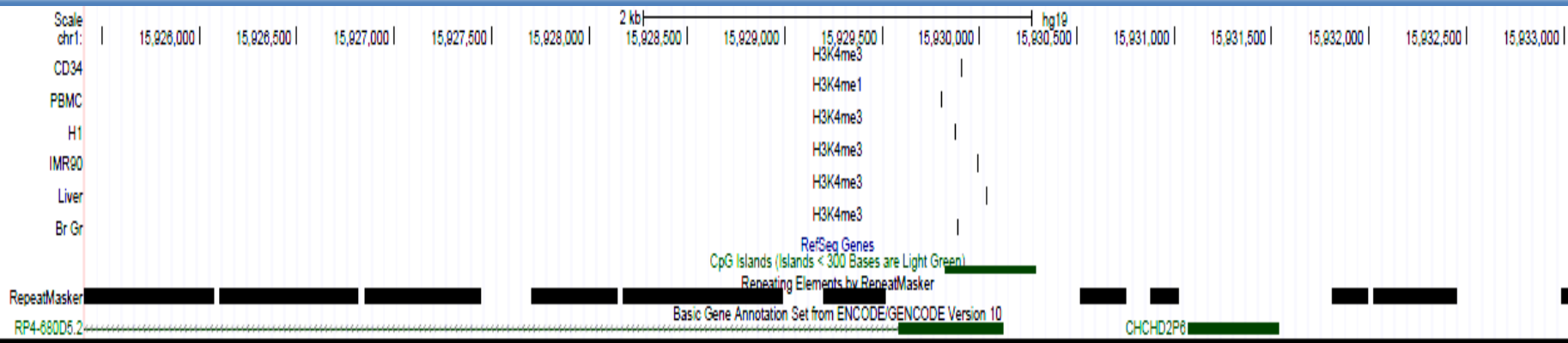
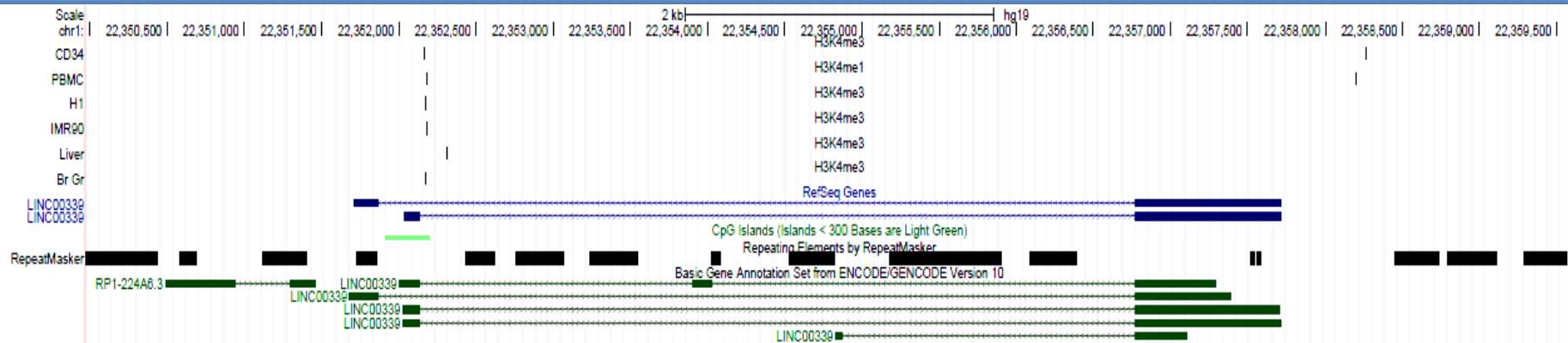
lncRNA coding regions having similar H3K36me3 pattern across different cell/tissue types



IncRNA coding regions having similar H3K9me3 pattern across different cell/tissue types



lncRNA coding regions having similar H3K4me3 pattern across different cell/tissue types



IncRNA coding regions having similar H3K4me3 pattern across different cell/tissue types

