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. (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 smoothened 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 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)

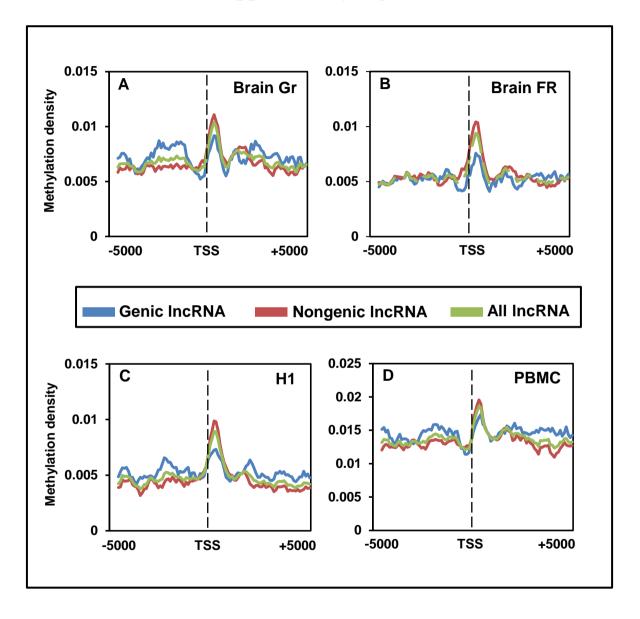
Count of histone modification in each category. (B) Normalised distribution of histone modification in each category.

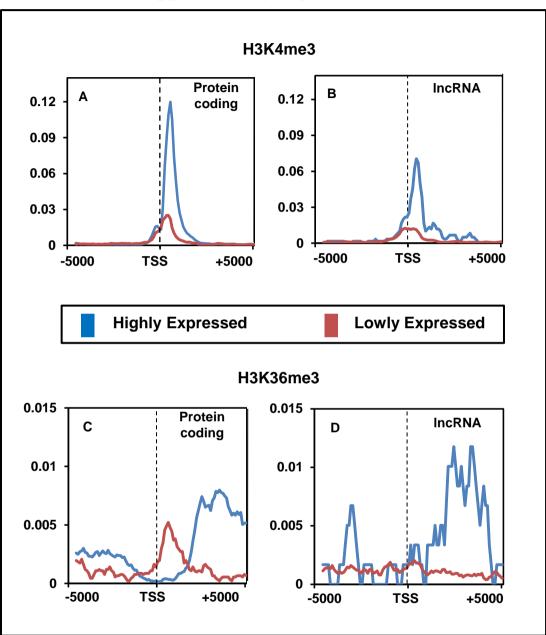
Supplementary Table 3: Global distribution of the H3K4me3, H3K9me3, H3K27me3, H3K36me3 histone modifications and DNA methylation around ±2Kb of the TSS of protein coding and lncRNA genes in all datasets analysed. Each value represents the ratio of the number of genes having the modification to the total 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.

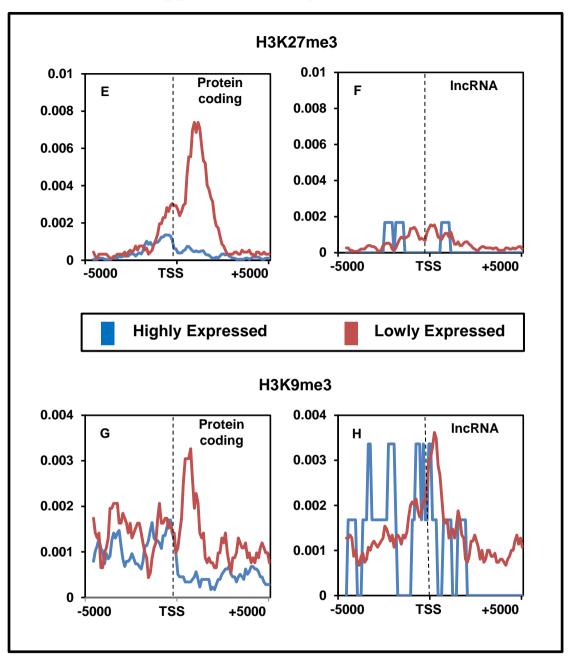
Supplementary File 1: Venn diagram showing the coexistence of two or more of these histone modifications at 2kb up or down stream of TSS of lncRNA and mRNA coding regions in different cell and tissue types.

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.

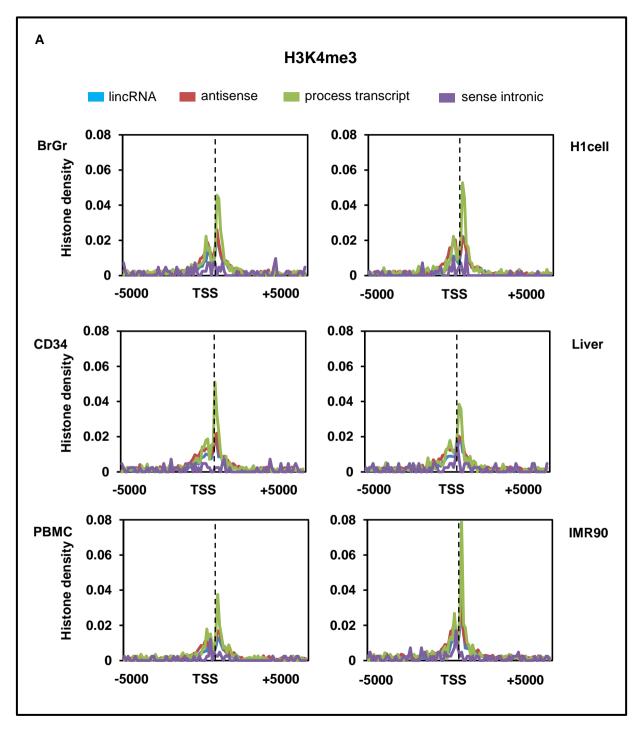




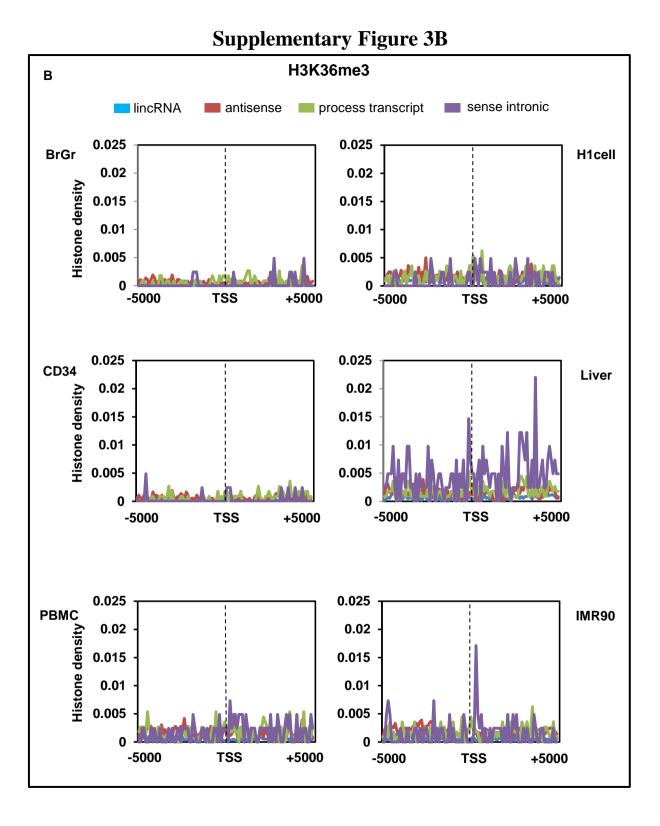
Supplementary Figure 2 A-D



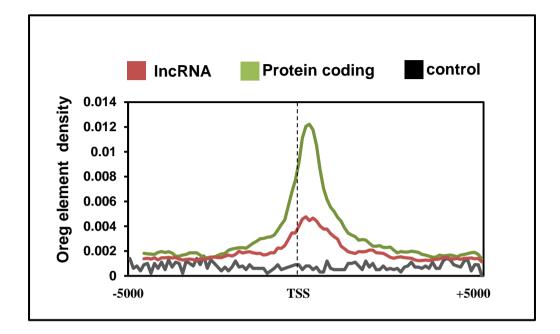
Supplementary Figure 2 E-G



Supplementary Figure 3A



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 fpkm	51653
Brain expressed gene fpkm	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) $H1$ 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) Br mRNA coding region (Mean - SD)	1415
Br lncRNA coding region (Mean + SD)	1415
Br lncRNA coding region (Mean - SD) Br lncRNA coding region (Mean - SD)	3665
DI_INCKINA_counig_region (Wean - 5D)	5005

Supplementary Table 2. (A)

		PF	ROTEI	N COI	DING	GENE	S		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						195	1907	94	61	74	719	427	290	60	
PBMC	9301 217 2939 232 994 142 835 162					162	1537	69	58	623	360	217	330	488		

Number of genes having the histone modifications.

(B)

	H3K4me3	H3K9me3	H3K27me3	H3K36me3	
BRAIN	3.5711	0.0382	4.6713	0.1816	Protein
GERMINAL					Coding with
CD34	3.5090	0.0812	1.7758	0.1810	CGI
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	0.6922	0.0541	0.5727	0.1612	Protein
GERMINAL					Coding
CD34	0.8449	0.0738	0.5521	0.1397	without CGI
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	3.1228	0.0656	3.8385	0.2848	LncRNA
GERMINAL					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	0.2751	0.0812	0.1374	0.1729	LncRNA
GERMINAL					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 C	ODING GENE	S	LncRNA GENES							
	H3K4me3 H3K9me3 H3		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				
РВМС	10295	359	3774	394	1897	286	1111	388				

Number of genes having the histone modification

(B)

		PROTIEN C	ODING GENE	S	LncRNA GENES							
	H3K4me3 H3K9me3 H3K2		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				
РВМС	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)

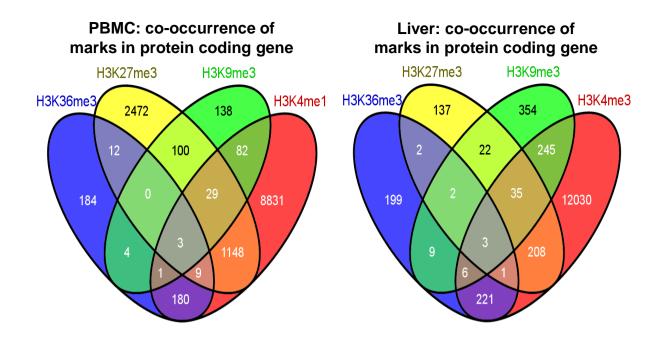
	PROTIEN CODING GENES	LncRNA GENES
BRAIN GERMINAL	15634	5357
H1	15387	4771
РВМС	16382	6974

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

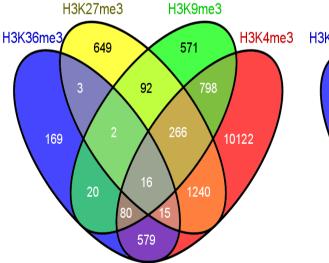
Supplementary Table 4.

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

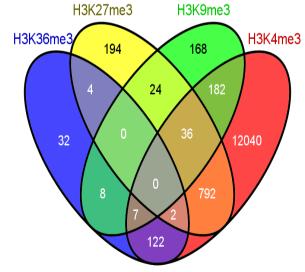
Supplementary File 1

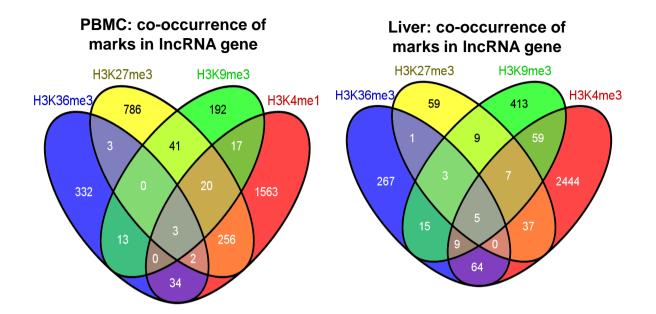


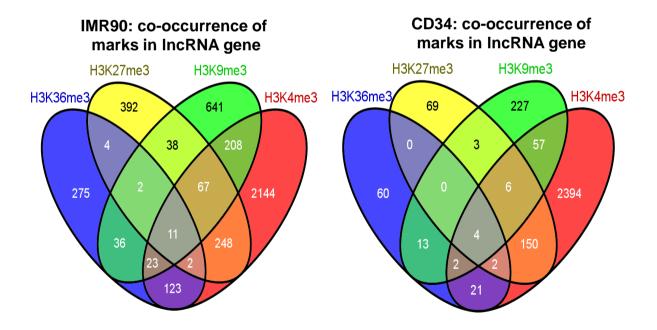
IMR90: co-occurrence of marks in protein coding gene

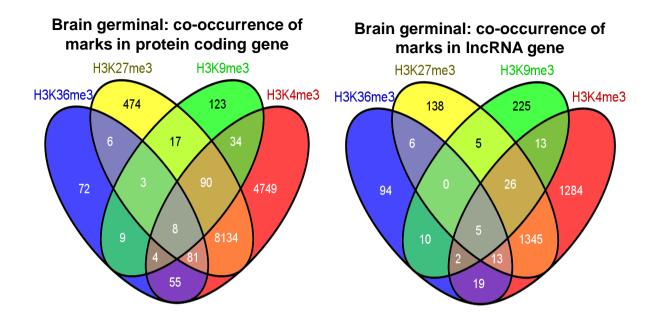


CD34: co-occurrence of marks in protein coding gene



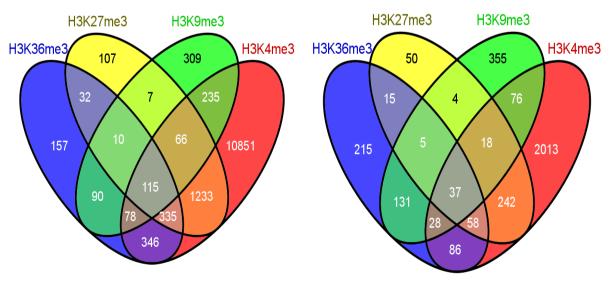






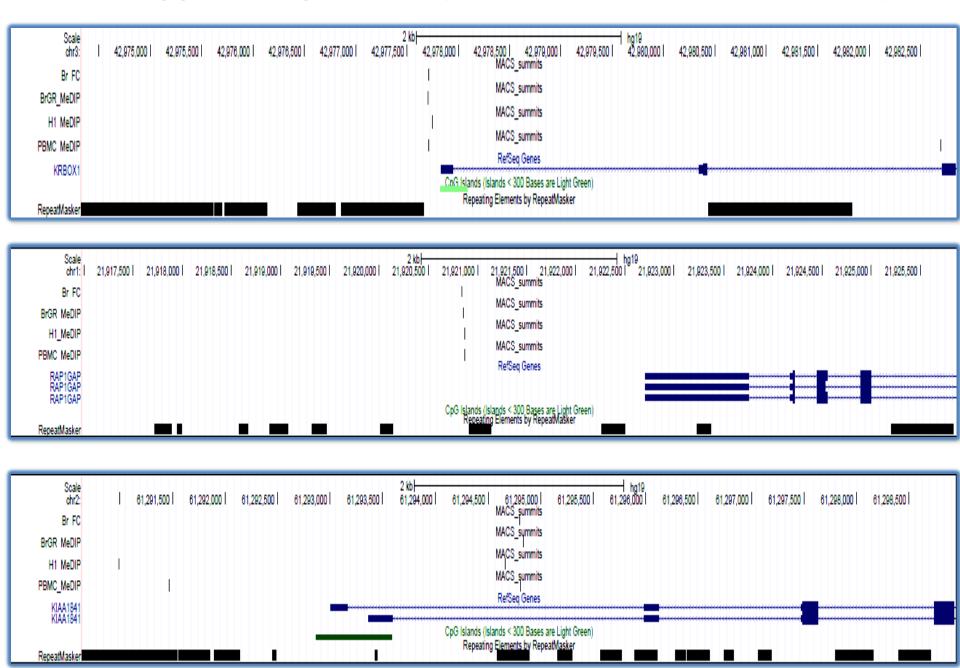
H1 cell line: co-occurrence of marks in protein coding gene

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

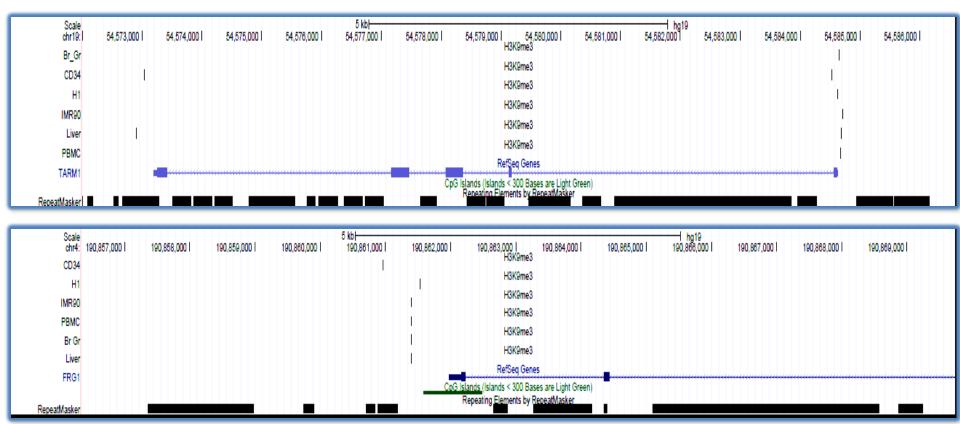


Supplementary File 2

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



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

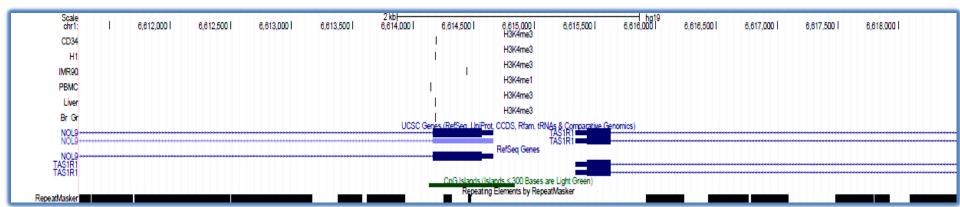


Scale chr9:	104,144,500 İ	104, 145,000 l	104, 145,500 l	2 kb 104,146,000	104,146,500	104,147,000 H3K9me3	104,147,500 l		104,148,500 l	104,149,000 l	104,149,500 l
Br_Gr							1				
CD34						H3K9me3	I				
H1						H3K9me3	I				
IMR90						H3K9me3	1				
Liver						H3K9me3		1			
PBMC						H3K9me3	1				
BAAT				****		RefSeq Genes					
BAAT	***********************	*******	********		CpG Islands (Isl	ands < 300 Bases are Light G	ireen)				
RepeatMasker					Repeating	ands < 300 Bases are Light G Elements by ReneatMasker					

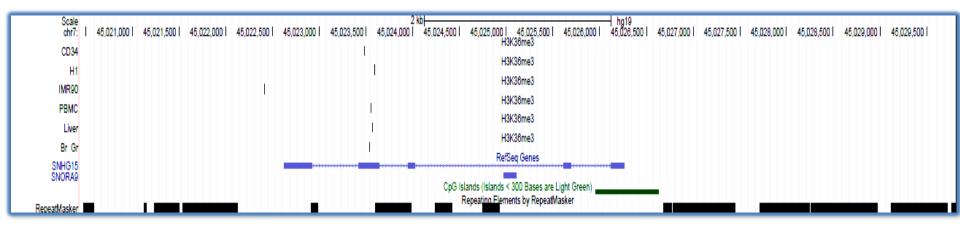
mRNA coding genes having similar H3K4me3 pattern across different cell/tissue types

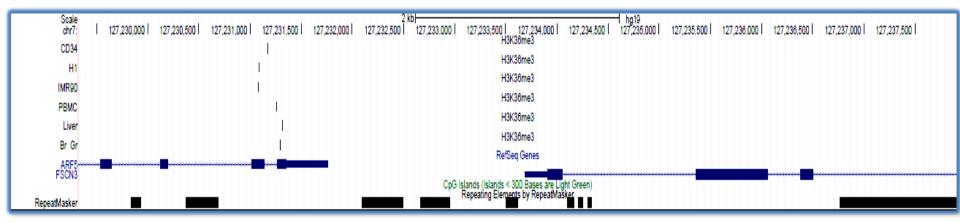
Scale chr1: CD34 H1 IMR90 PBMC Liver Br Gr APITD1-CORT	10,489,000 l 10,489,500 l	10,490,000 l 10,490,50) 10,491,000 10,49 	2 kb	10,492,500 10,493,000 H3K4me3 H3K4me3 H3K4me3 H3K4me1 H3K4me3 H3K4me3 RefSeq Genes	hg19 10,493,500 l 10,494,000 l	10,494,500 l	10,495,000 10,495	500 İ 10,496,000 İ
APITD1-CORT APITD1-CORT APITD1-CORT APITD1-CORT APITD1 APITD1-CORT RepeatMasker				CoG Isl	ands (Islands < 300 Bases are Light Gr epeating Elements by RepeatMasker	een)			

Г	Scale chr1:	I	1,404,000	1,404,500	1,405,000	1,405,500	1,408,000 l	2 kb 1,406,500 l	1,407,000	1,407,500 H3K4me3	1,408,000 l	hg19 1,408,500 l	1,409,000 l	1,409,500 l	1,410,000 l	1,410,500	1,411,000 l
	CD34								1								
	H1								1	H3K4me3							
	IMR90								1	H3K4me3							
	PBMC								1	H3K4me1							
	Liver								i I	H3K4me3							
	Br Gr								i i	H3K4me3							
	ATAD3C									RefSeq Genes							
	ATAD3C ATAD3B								CoG Islands (Isl	ands < 300 Bases a		*****	*****			****	******
	RepeatMasker									g Elements by Repea							



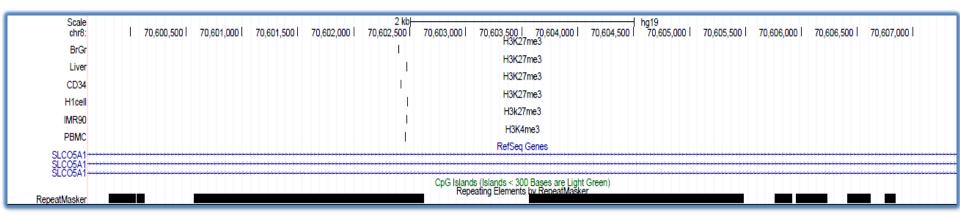
mRNA coding genes having similar H3K36me3 pattern across different cell/tissue types





	Scale chr3:	I	136,66	2,000 l	136,662,500	136,663,000	136,663,500	136,664,000	136,664,500	2 kb 138,665,000	136,665,500	136,666,000 H3K36me3	136,666,500	hg19 136,667,000	136,667,500 l	136,668,000	136,668,500 l	136,669,000 l	136,669,500 l	138,670,000 l
	CD34									1										
	H1									1		H3K36me3 H2K28me2								
	IMR90									1		H3K36me3								
	PBMC									I		H3K36me3								
	Liver									I.		H3K36me3								
	Br Gr									1		H3K36me3 RefSeq Genes								
	NON!											Reideq Gelles			÷					
Repe	atMasker									Ср	G Islands (Islan Repeating B	ids < 300 Bases a Elements by Repe	re Light Green) atMasker							

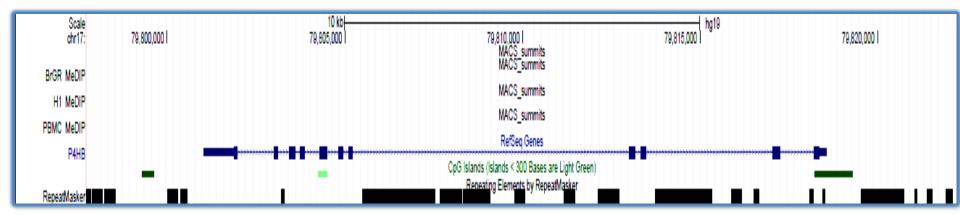
mRNA coding genes having similar H3K27me3 pattern across different cell/tissue types



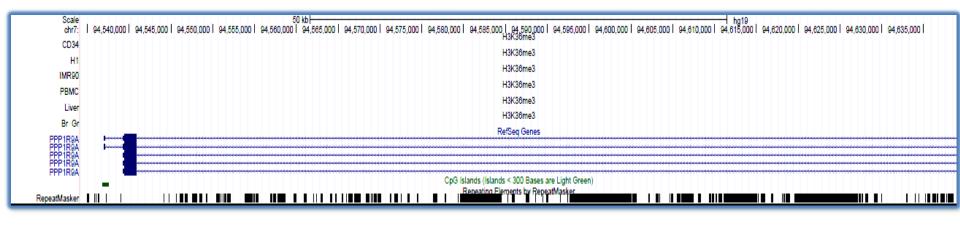
Scale chr6: I	26,499,000 l	26,500,000 l	26,501,000 l	26,502,000 l	5 kb 28,503,000 l	26,504,000 l	26,505,000 H3K27me3	26,506,000 l	28,507,000 l	 26,509,000 l	26,510,000 l	26,511,000 l
CD34					1							
H1				1			H3K27me3					
IMR90				1			H3K27me3					1
PBMC				I			H3K27me3					
Br Gr				1			H3K27me3					
Liver					1		H3K27me3					
BTN1A1							RefSeg Genes			 		_
					_		s (Islands < 300 Bases eating Elements by Repe					

Scale chr6:	142,408,000	142,408,500 l	142,409,000 l	142,409,500 l	2 kb	142,410,500 l	142,411,000 142,411,500 H3K27me3	hg19 142,412,000 142,412,500	142,413,000 l	142,413,500 l	142,414,000 l
BrGr					1						
Liver							H3K27me3 H3K27me3				
CD34				1							
H1cell					I		H3K27me3				
IMR90				1			H3k27me3				
PBMC					1		H3K4me3				
NMBR		*****	*****			0.01.1	RefSeq Genes				
				-			ls (Islands < 300 Bases are Light Gr e <u>ating Elements by R</u> epeatMasker	een)			
RepeatMasker											

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



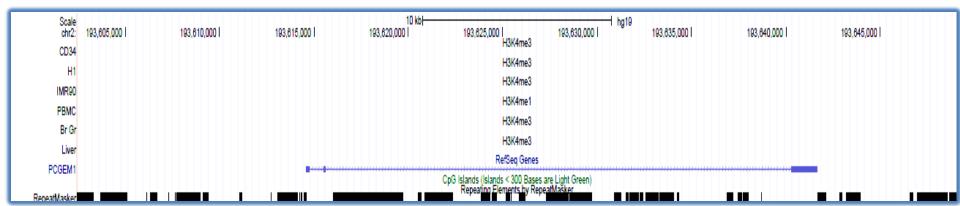
mRNA coding genes not having H3K36me3 marks across different cell/tissue types



mRNA coding genes not having H3K27me3 marksacross different cell/tissue types

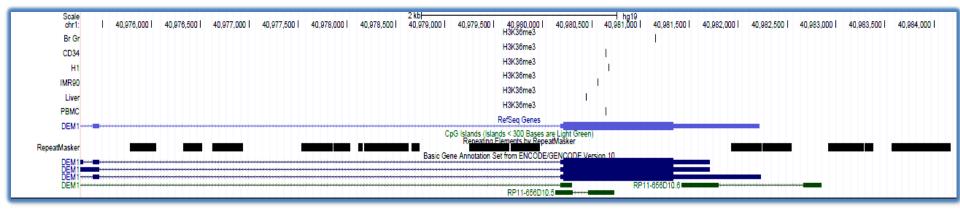


mRNA coding genes not having H3K4me3 marks across different cell/tissue types



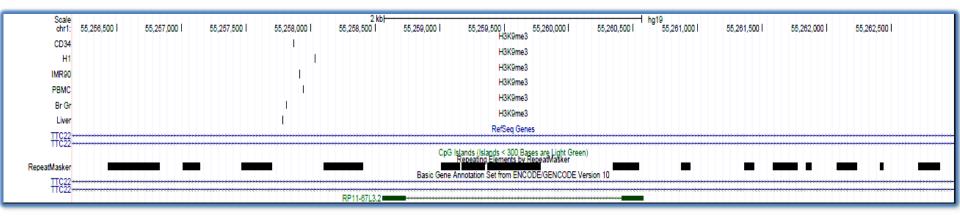
IncRNA coding regions having similar H3K36me3 pattern across different cell/tissue types







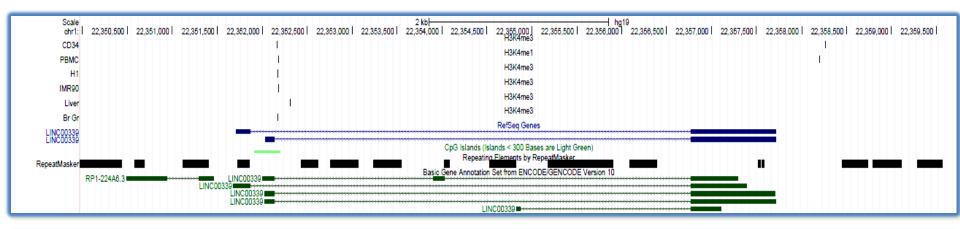
IncRNA coding regions having similar H3K9me3 pattern across different cell/tissue types



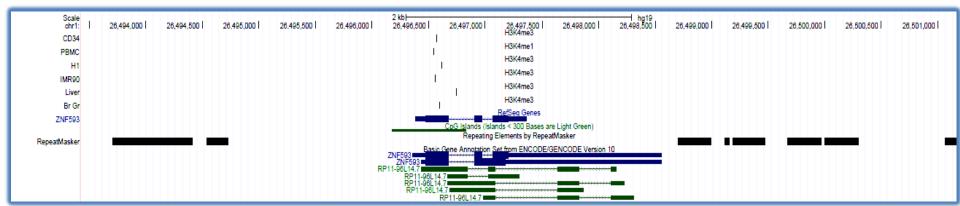
Scale chr2:	l 114,735,0	000 İ 114,735,500 İ	114,736,000 l	114,736,500 İ	114,737,000 l	114,737,500 l	2 kb 114,738,000	114,738,500 l	114,739,000 l H3K9me3	114,739,500 l	hg19 114,740,000 l	114,740,500 l	114,741,000 l	114,741,500 l	114,742,000 l	114,742,500 l	114,743,000 l
CD34 H1 IMR90 PBMC Br Gr Liver						 	1		H3K9me3 H3K9me3 H3K9me3 H3K9me3 H3K9me3								
LOC100499194 LOC440900 RepeatMasker		_		-	AC01098	2.1 AC110769.3		pG Islands (Islan Repeating F	RefSeq Genes ds < 300 Bases a Tements by Repe et from ENCODE/	atMasker	ion 10	*****	****			*****	



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



Scale chr1: I 15,9 CD34 PBMC H1 IMR90 Liver	26,000 15,926,500	15,927,000 l	15,927,500 l	15,928,000 l	2 kb	15,929,000 İ	15.929.500 H3K4me3 H3K4me1 H3K4me3 H3K4me3 H3K4me3	15,930,000 	<mark> h</mark> g19 15,930,500	15,931,000	15,931,500	15,932,000	15,932,500 İ	15,933,000
Br Gr RepeatMasker RP4-880D5.2					Basic	Repeating E	H3K4me3 RefSeg Genes nds < 300 Bases are Elements by Repeat Set from ENCODE/G	Masker	-	CHCHD2P6				-



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

70,920,500 l	70,921,000 l	70,921,500 l	70,922,000 l	2 kb 70,922,500 l	70,923,000 l	70,923,500 l	70,924,000 l	hg19 70,924,500 l	70,925,000 l	70,925,500 l	70,926,000 l	70,926,500 l
					1							
			1		1							
			1		1							
					1							
					Í	H3K27me3						
						H3K27me3						
					'	RefSeq Genes						
			-									
			RP11-5	61023.7 =++++++++	Sasic Gene Annotation	n Set from ENCODE/	GENCODE Version 1	0				
	70,920,500	70,920,500 İ 70,921,000 İ	70,920,500 70,921,000 70,921,500		70,920,500 I 70,921,000 I 70,921,500 I 70,922,000 I 70,922,500 I I I	CpG Islands (t Repeatir Basic Gene Annotatio	70,920,500 1 70,921,000 1 70,921,500 1 70,922,000 1 70,922,500 1 70,923,000 1 70,923,500 1 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 RefSeq Genes CpG Islands (Islands < 300 Bases a Repeating Elements by Repei Basis Gene Annotation Set from ENCODE/I	70,920,500 1 70,921,000 1 70,921,500 1 70,922,000 1 70,922,500 1 70,923,500 1 70,923,500 1 70,924,000 1 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 RefSeq Genes CpG Islands (Islands < 300 Bases are Light Green) Repeating Elements by RepeatMasker Basic Gene Annotation Set from ENCODE/GENCODE Version 1	70,920,500 70,921,000 70,921,500 70,922,500 70,923,500 70,924,500 70,924,500 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3 H3K27me3	70,920,500 70,921,000 70,922,000 70,922,500 70,923,000 70,923,500 70,924,500 70,924,500 70,924,500 70,924,500 70,924,500 70,925,000 10,925,000 10,925,000 10,925,000 10,925,000 10,925,000 10,925,000 10,925,000 10,924,500 70,924,500 10,924,500 <th>70,920,500 70,921,000 70,922,000 70,922,500 70,923,500 70,924,000 70,924,500 70,925,000 70,925,500 H3K27me3<!--</th--><th>70,920,500 70,921,500 70,922,500 70,922,500 70,923,500 70,924,600 70,924,600 70,925,500 70,925,500 70,925,500 70,925,500 70,925,500 70,926,000 H3K27me3 H3K27me3 H3</th></th>	70,920,500 70,921,000 70,922,000 70,922,500 70,923,500 70,924,000 70,924,500 70,925,000 70,925,500 H3K27me3 </th <th>70,920,500 70,921,500 70,922,500 70,922,500 70,923,500 70,924,600 70,924,600 70,925,500 70,925,500 70,925,500 70,925,500 70,925,500 70,926,000 H3K27me3 H3K27me3 H3</th>	70,920,500 70,921,500 70,922,500 70,922,500 70,923,500 70,924,600 70,924,600 70,925,500 70,925,500 70,925,500 70,925,500 70,925,500 70,926,000 H3K27me3 H3K27me3 H3



