Biomedical Sciences
2015; 1(3): 18-33
http://www.sciencepublishinggroup.com/j/bs
doi: 10.11648/j.bs.20150103.11

Epigenetic Transfiguration of H3K4me2 to H3K4me3 During Differentiation of Embryonic Stem Cell into Non-embryonic Cells

Smarajit Das[1, *], Pijush Das[2], Sanga Mitra[3], Medhanjali Dasgupta[4], Jayprokas Chakrabarti[3, 5], Eric Larsson[6]

[1]Department of Genetics, University of Georgia, Athens GA, USA
[2]Cancer Biology & Inflammatory Disorder Division, Indian Institute of Chemical Biology, Kolkata, India
[3]Computational Biology Group, Indian Association for the Cultivation of Science, Kolkata, India
[4]Department of Chemical Engineering (Bioprocess Engineering), Jadavpur University, Kolkata, India
[5]Gyanxet, Salt Lake, Kolkata, India
[6]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Email address:
smarajit@uga.edu (S. Das)

To cite this article:
Smarajit Das, Pijush Das, Sanga Mitra, Medhanjali Dasgupta, Jayprokas Chakrabarti, Eric Larsson. Epigenetic Transfiguration of H3K4me2 to H3K4me3 During Differentiation of Embryonic Stem Cell into Non-embryonic Cells. *Biomedical Sciences.* Vol. 1, No. 3, 2015, pp. 18-33.doi: 10.11648/j.bs.20150103.11

Abstract: Chromatin immune precipitation followed by high-throughput sequencing (Chip-Seq), investigate the genome-wide distribution of all histone modifications. Lysine residues within histones di or tri-methylated in *Saccharomyces cerevisiae have been studied earlier.* Tri-methylation of Lys 4 of histone H3K4me3 correlates with transcriptional activity, but little is known about this methylation state in human. It was also previously proved that deposition of H3K4me2 modification at TSS is associated with gene repression in the yeast cell. Overlapping non-coding RNA (ncRNA) transcript assumes a crucial role in this repression. Here, we examine the H3K4me2 and H3K4me3 methylation dynamics at the TSS region of human genes across the ENCODE (https://www. encode project. org/) Consortium 8 cell lines GM12878, H1-hESC, HeLa-S3, HepG2, HSMM, HUVEC, K562 and NHEK, we identified clear divergence of histone modification profiles in H1-hESC with respect to others. While, H3K4me2 modifications were found to be associated with the vast majority of genes in the H1-hESC with a significantly decreased amount in other differentiated cell lines, H3K4me3 modification showed completely reverse trends. By the process of differentiation, a distinct set of genes lose H3K4me2 in H1-hESCand gain H3K4me3 in differentiated cell, thereby, enhancing the expression level of the corresponding genes. On the level of gene ontology molecular function classification, these genes are mostly associated with protein binding, nucleotide binding, DNA binding and ATP binding. Other than that, signaling and receptor activity, metal ion binding and phosphorylation-dephosphorylating action can be correlated with these genes. We expect a crosstalk between the change of methylation status and gene functionality, as all these functions can be allied to transcriptional regulation and gene activation, which once again is linked to H3K4me3 mark.

Keywords: Epigenetic, H3K4me2, H3K4me3, RNA-Seq, Chip-Seq, UCSC, Methylation Dynamics

1. Introduction

A nucleosome consists of two copies of four core histones, namely H2A, H2B, H3, and H4, wrapped by 147bp of DNA [1]. The N-terminal tails of these histones are processed with different types of post-translational modifications such as acetylation, methylation, phosphorylation, ubiquitination, glycosylation, and sumoylation [2]. These modifications correlate with the transcriptional efficiency of the gene ie., expression or repression. The dynamic integration and dissociation of these modifications have been known to change the chromatin structure that provides binding sites for proteins and thereby regulate cellular processes such as transcription, repair, replication, and genome stability [3-6].

Genome-wide approaches of profiling histone modifications, initialized by tiling array analysis (chip-chip) and later followed by new generation sequencing technique

chromatin immune-precipitation followed by high-throughput sequencing (Chip-Seq), have revealed the characteristic genomic distribution and the association of gene functions and activities in various model organisms [7-10]. It has emerged from the analyses that there are six classes of histone H3 modifications that are subjected to epigenome profiling by International Human Epigenome Consortium (http://ihec-epigenomes. org/). In general, the particular histone methylation H3K4me3 and many other histone acetylations usually enrich the transcription start site (TSS) and positively correlate with genes expression [11]. In fact, active enhancers can be identified by the enrichment of both H3K4me1 and H3K27ac modifications. However, there are a lot of silent gene promoters with active markers, those can be found in Embryonic Stem cells (H1-hESC or simply ESC) and T-cells and active transcription can be addressed by an additional modification, H3K36me3, over transcribed gene body [12-14]. Gene repression can be mediated through two distinct mechanisms that involve tri-methylated H3K9 (H3K9me3) and tri-methylated H3K27 (H3K27me3). Interestingly H3K4me2 plays multiple roles; sometimes it is associated with activation, sometimes with repression and sometimes a combination of both [15-17]. Moreover, H3K4me2 marking precedes and persists transcription.

The availability of numerous histone modification data from different cell lines of the human genome facilitates the discovery of functionally significant sequence stretches via comparative genomics approach. Besides evolutionary conserved sequences, many novel elements can be found by examining chromatin accessibility and histone modification or DNA methylation patterns [18–22]. A representative international project aiming to find all the functional elements in the human genome, called the Encyclopedia of DNA Elements (ENCODE) pilot project, has examined human genomic sequences using a number of existing techniques [23]. Many functional elements examined by the ENCODE project are likely unconstrained across mammalian evolution and comprise a large reservoir of functionally conserved but non-orthologous elements between species as well as lineage-specific elements. The histone modification mapping could provide highly informative signatures to the estimation of presence and activity of gene promoters and distal regulatory sites [24].

The main objective of our Chip-Seq analysis project is to examine histone modification patterns and their dynamics as Embryonic stem cell differentiates into other normal and cancer cell lines and their correlation with gene expression level at these specific cell lines. It was previously proved that H3K4me2 modification is associated with gene repression in the yeast cell. Overlapping non-coding RNA transcript assumes a crucial role in this repression [25]. Our aim is to study the Chip-Seq data to investigate the genome-wide distribution of di and tri-methylation of H3K4 (H3K4me2 and H3K4me3) in different normal, embryonic and cancer cell lines in human. By analyzing published Chip-Seq data from UCSC genome browser (http://genome. ucsc. edu/cgi-bin/hg Gateway) this study validated that there is a clean divergence of H3K4me2 distribution in embryonic verses differentiated cell lines. By the process of differentiation, a distinct set of genes lose H3K4me2 in H1-hESCand gain H3K4me3 in differentiated cells, thereby, enhancing the expression level of the corresponding genes. To define this change of histone mark along with expression change from stem cell line to differentiated cell lines, we use the term transfiguration. The histone marks that appear mainly in generic regions were studied around the transcription start sites (TSSs) of the genes. Thus, this analysis demonstrates that H3K4me2depositions around TSS (s) are associated with gene repression in human H1-hESC.

2. Results

2.1. The Epigenetic Landscape of H3K4me2 and H3K4me3 Modifications

Epigenetic mechanism is emerging as one of the major factors of the dynamics of gene expression in different human cells. To elucidate the role of chromatin remodeling in transcriptional regulation associated with gene expression, we mapped the spatial pattern of chromosomal association with histone H3 modifications using Chip-Seq. Here, we concentrated on the epigenetic map of two histone modifications, namely, H3K4me2 and H3K4me3 of the protein coding genes for 8 different cell lines, namely, H1-hESC, GM12878, HeLa, HepG2, HUVEC, HSMM, NHEK and K562 [Fig. 1].

While, H3K4me2 modifications were found to be associated with the vast majority of genes in the H1-hESC with a significantly decreased amount in other differentiated cell lines, H3K4me3 modification showed completely reverse trends. Analyzing a 2*2 Fisher's Exact Test revealed, the association between groups (H1-hESC and other cell lines) and outcomes (me2 exclusive and me3 exclusive) is considered to be very statistically significant. The two-tailed P values are less than 0.01 for all groups with one exception between H1-hESC and HepG2.This data display the dynamic distribution, which is the focus of our study [Table 1]. More importantly, H3k4me2 and H3K4me3 modifications, both displaying tight correlations with transcript levels, show differential affinity to distinct genomic regions while occupying predominantly the transcription start site (TSS). These promoter occupancies of H3K4me2 at different loci indicate the repression of specific DNA elements in H1-hESC, which is ultimately nullified by the loss of H3K4me2 in differentiated cell lines. The repressed genes became hyperactive with the introduction of H3K4me3 exclusively instead of H3k4me2. In addition, we brought to light the effect of the presence of multivalent domains, focusing on the importance of combinatorial effects on transcription. H3K4me2 and H3K4me3 mixed TSS have an intermediate effect of H3k4me2 and H3K4me3. Overall, our work portrays a substantial association between the chromosomal locations of these two epigenetic markers, transcriptional activity and cell type specific transitions in the epigenome.

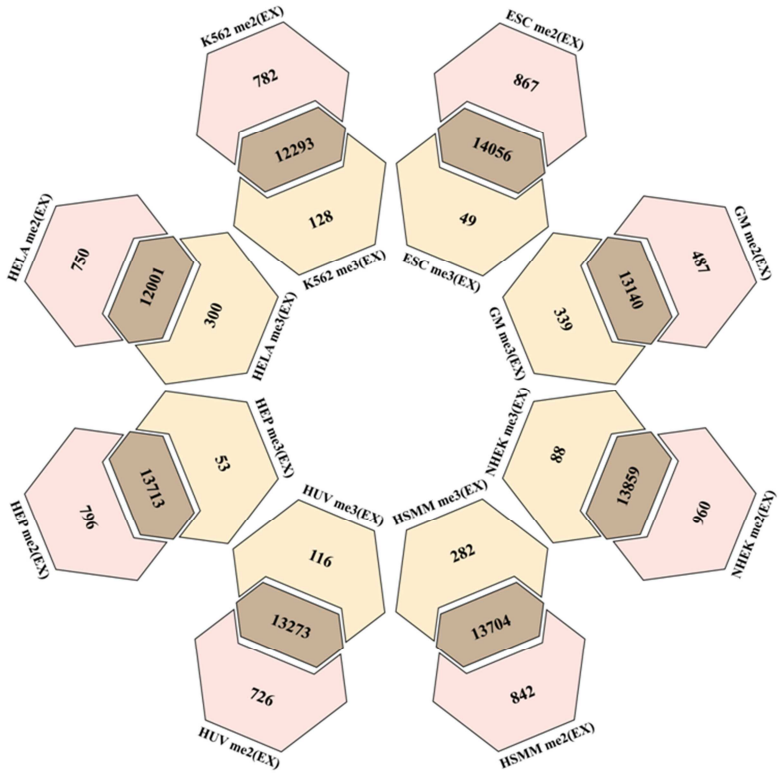


Figure 1. *Exclusive Methylation (H3K4me 2 and H3K4me 3) Data.*

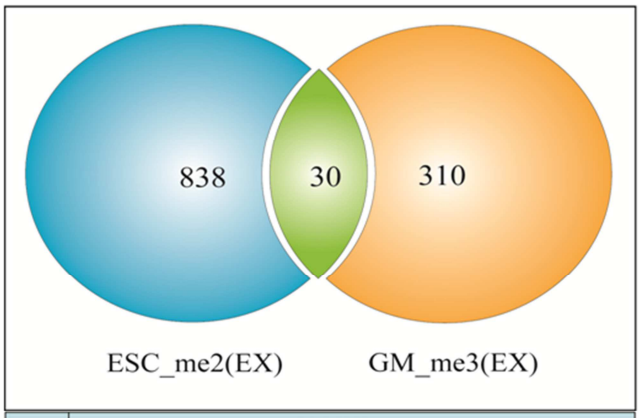
Here Exclusive H3K4me2 [me2 (EX)] and H3K4me3 [me3 (EX)] methylations in the 8 cell lines chosen for this study have been shown. It has been observed that out of 19999 protein-coding genes there are 867 genes, which are exclusively me2 modified in the ESC line. Only 49 of the protein coding genes are exclusively me3 modified and the rest 14066 genes have both me2 and me3 modification at the TSS within the range of +/- 2Kb. Nearly 5,000 genes have no me2 or me3 modification. These studies have been similarly performed for other cell lines (GM, HeLa, HuVec, HepG2, HSMM, K562 and NHEK) and the corresponding Venn diagrams have been represented in this Fig1.

Table 1. *Epigenetic landscape of H3K4me2 and H3K4me3for 8 different cell lines.*

Cell Lines	me2_Exclusive	me3_Exclusive	me2_me3Common	me2/me3
H1-hESC	867	49	14066	17. 6938
GM12878	487	339	13140	1. 4365
K 562	782	128	12293	6. 10
HeLa S3	750	300	12001	2. 5000
HepG2	796	53	13713	15. 0188
HuVEC	726	116	13273	6. 2586
HSMM	842	282	13704	2. 9858
NHEK	960	88	13859	10. 9090

2.2. Methylation Dynamics from Embryonic to Differentiated Cell

We studied gene methylation profiles using Chip-Seq to identify pair wise histone dynamics [Fig. 2]. E.g., 867 genes expressed in H1-hESC were exclusively H3K4me2 modified, where as 339 genes were found in GM cell having H3K4me3 exclusively. This analysis indicates that 30 genes lose their H3K4me2 modification and gain H3K4me3 modification exclusively as the ESC differentiates into GM.



	GM [30]				
ESC	ATPAF1	S100A5	FMO1	GPR61	ZRANB1
	ANO3	KDM2A	RFC5	C14orf105	PPP4R4
	PPP4R4	ACSBG1	CALML4	ANXA2	FAM227B
	TFEC	GH2	RFX3	ODF4	HCN2
	CRNKL1	SRMS	SRXN1	ADORA2A-AS1	C4orf45
	FAM198B	DOK7	PRDM13	CTD-2054N24.2	ELP3

Figure 2. *ESC_ me2 Converted to GM me_3 and CDF Plot.*

A total of 30 genes with H3K4 me2methylations in the ESC

line have been converted into H3K4me3 methylation when theH1-hESCdifferentiates intoaGM12878cell line.

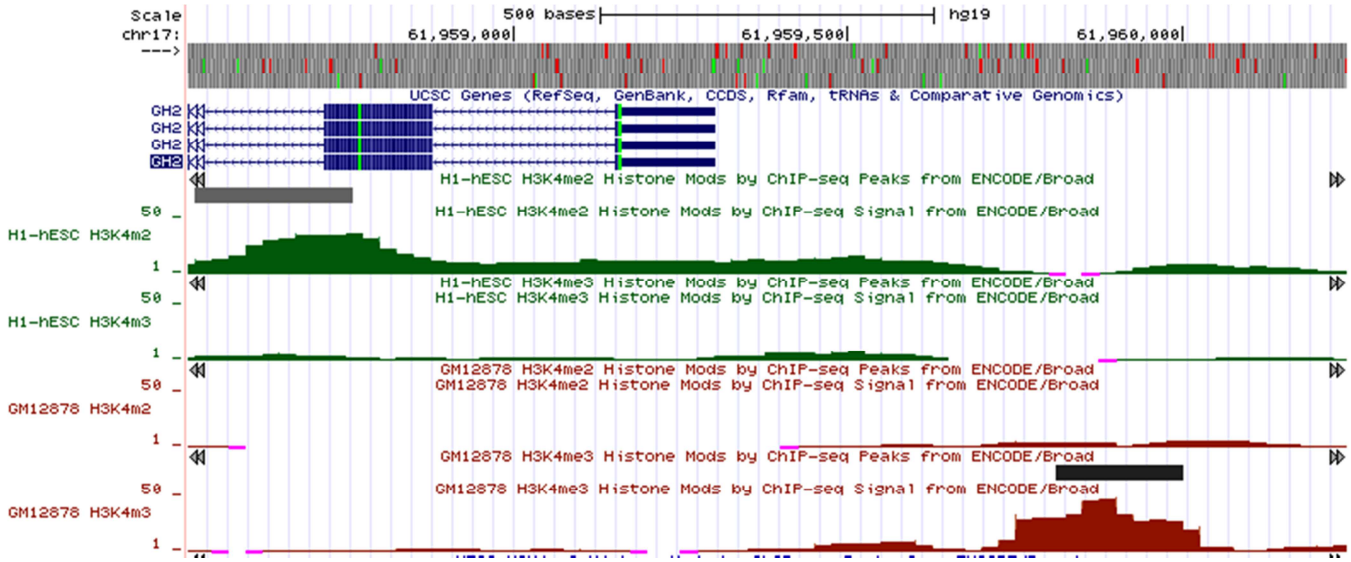


Figure 3. *UCSC Image showing loss of H3K4me2 methylation in Embryonic cell line and gain of H3K4me3 methylation in GM cell line post differentiation for the GH2 gene.*

Out of the 30 genes, one of them, namely GH2, has been represented in [Fig. 3]. The UCSC Genome image of GH2 distinctly shows the presence of H3K4me2 modification in the +/- 2 KB region of the TSS in the H1-hESC while no such peak or signal is observed in case of H3K4me3 modification. Again, for the GM cell line, a clear presence of H3K4me3 methylation is observed in the +/- 2 KB region of the TSS while no such peak or signal is noted in case of H3K4me2 methylation. This clearly shows the loss of H3K4me2 methylation in GH2 in the embryonic stem cell line and incorporation of H3K4me3 as it differentiates into the GM cell line. We recorded this differentiated dynamics and calculated the epigenetic dissociation of H3K4me2 for the other 6 pair of cell lines. [Table S1 and Fig. S1].

Table S1. *ESC_ me 2 converted to non-ESC_ me 3.*

	ESC							
GM [30]	ATPAF1	S1000A5	FMO1	GPR61	ZRANB1	ANO3	KDM2A	RFC5
	C14orf105	PPP4R4	ODF4	HCN2	TEFC	ELP3	CT2054N24.2	RFX3
	GH2	FAM198B	DOK7	PRDM13	ANXA2	SRXN1	ADORA2A-AS1	
	C4orf45	PPP4R	ACSBG1	CALML4	SRMS	CRNKL1	FAM227B	
HELA [27]	SLC16A12	RP11	DUPD1	DUSP13	DNMT3LNR5A1	BESST1	FAM166A	
	F10	CNFN	IL1RL2	SRXN1	PLA2G16	AHRR	BC032117	
	GCK	CPA5	SLURP1	SLC39A4	PWP2	MYOZ3	SLC6A12	
	FAM27A	CFP	HMGB3	DLG5		C9orf62		
HepG2 [8]	CTD-2054N24.2	CLUAP1	ALKBH7	SEMA3C	GCK	FAM27A	PRDX4	AMELY
HUVEC [9]	TMEM82	PLCB2	CNFN	ASPDH	SEMA3G	GTF2IP1	D-2054N2	GSN
	F8							
HSMM	MYBPH	CEND1	REFC5	NPIPL1	GH2	KCNA7	TMEM88B	CFD
[18]	ASPDH	SPPL2B	GTF2IP1	SRMS	AREG	GLRA1	DJ031154	
	IL17REL	USP9Y					L3MBTL2	
NHEK [13]	CAHLM1	ERCC6	INSC	PVRL1	FAM195B	CRIM1	C10orf129	SRXN1
	RHOH	WDR41	GCK	GBX1	MAGEB10			
K562 [17]	HAPLN2	C10orf55	ERCC6	ANXA2	ABCC12	POU2F2	KLRK1	NYX
	USHBP1	TANK	CRNKL1	PPBP	RNF14	BAAT	FOXP3	FGF13
	RC4							

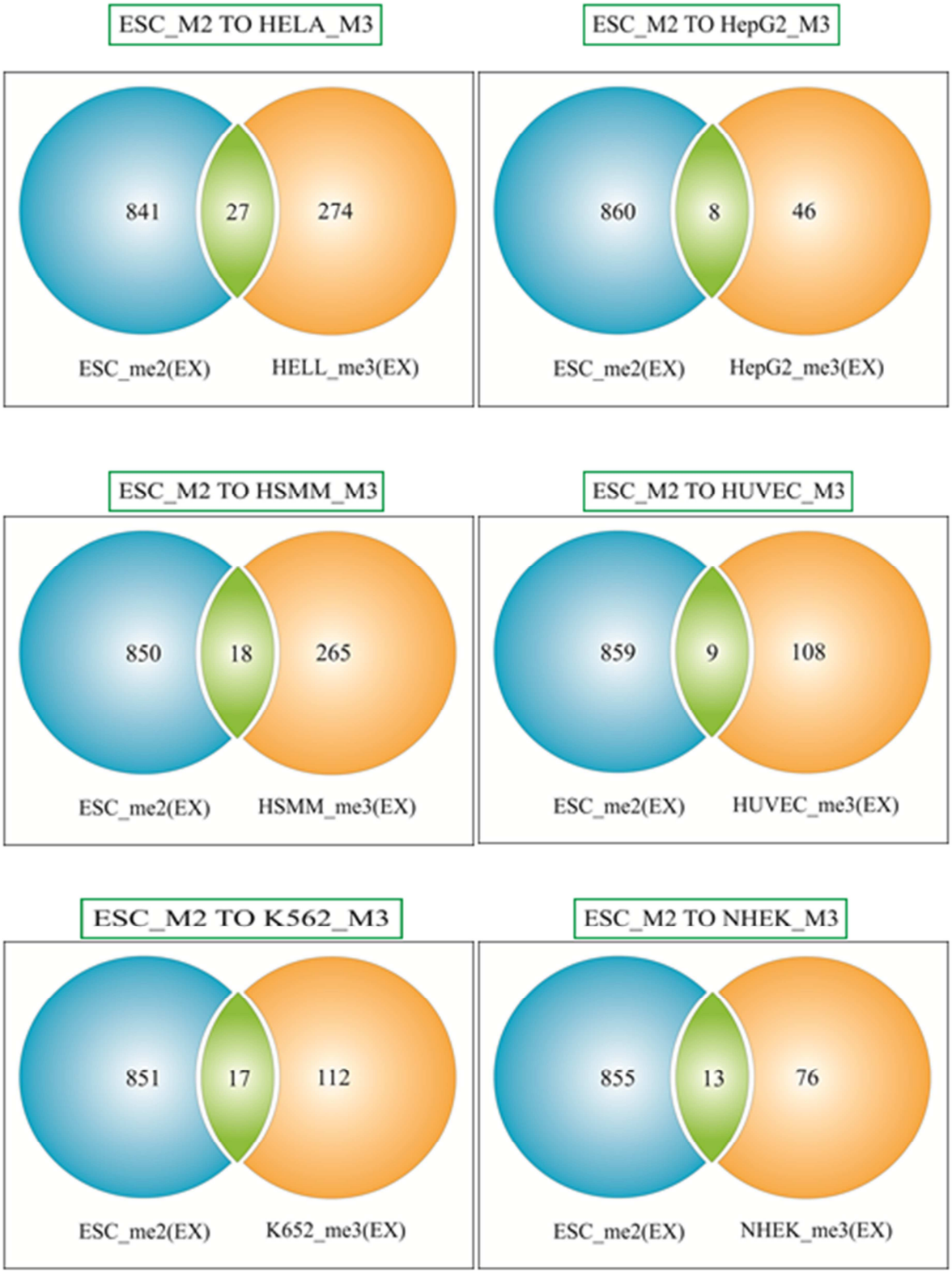


Figure S1. *ESC_me 2 converted to non-ESC_me 3.*

The UCSC Genome image of GH2 shows a clear peak and signal of H3K4me2 methylation in the +/-2 KB region of the TSS in the embryonic cell line. As the ES cell differentiates into GM cell line, a clear peak and signal of H3K4me3 methylation is noticed in the +/- 2KB region of the TSS while no such peak or signal of H3K4me2 methylation is noticed in the GM cell line.

After identifying the dynamic genes, i.e. those genes that

have H3K4me2 in only H1-hESC, we intersected the consequence of these genes when they dissolve the H3K4me2 and incorporateH3K4me3 in other differentiated normal/ cancer cell type.

Analysis of whether these genes are repressed or highly expressed in differentiated cell was done by pair-wise comparisons. We calculated gene expression profiles using RNA-Seq to identify pair wise differential expression. In addition, we showed that the methylation of both modifications in common domains have important combinatorial effects on transcription. Here, using CDF Plots, we have showed that H3K4me2 + H3K4me3 mixed TSS has an intermediate effect of H3k4me2 and H3K4me3 [Fig. 4 and Fig. S2]. Indeed, their overall expression was significantly higher in non-embryonic cell type [Fig. S3]. The gene lists per cell line where mixed me2-me3 modifications were converted into exclusive me3have been tabulated [Table S2]. CDF plot analysis depicts four conclusive results in human cell lines. Presence of me2 modification, especially in H1-hESC, makes the characteristics of most gene sets repressive. On the other hand, switch over of me2 modification to me3 increases the expression of the sets of genes. Conversion of me2modification to me3 modification up regulates the gene expression [Blue line in CDF plot] to a relatively greater extent compared to me2+3 transfigurations in to me3 modifications [Green line in CDF plot].

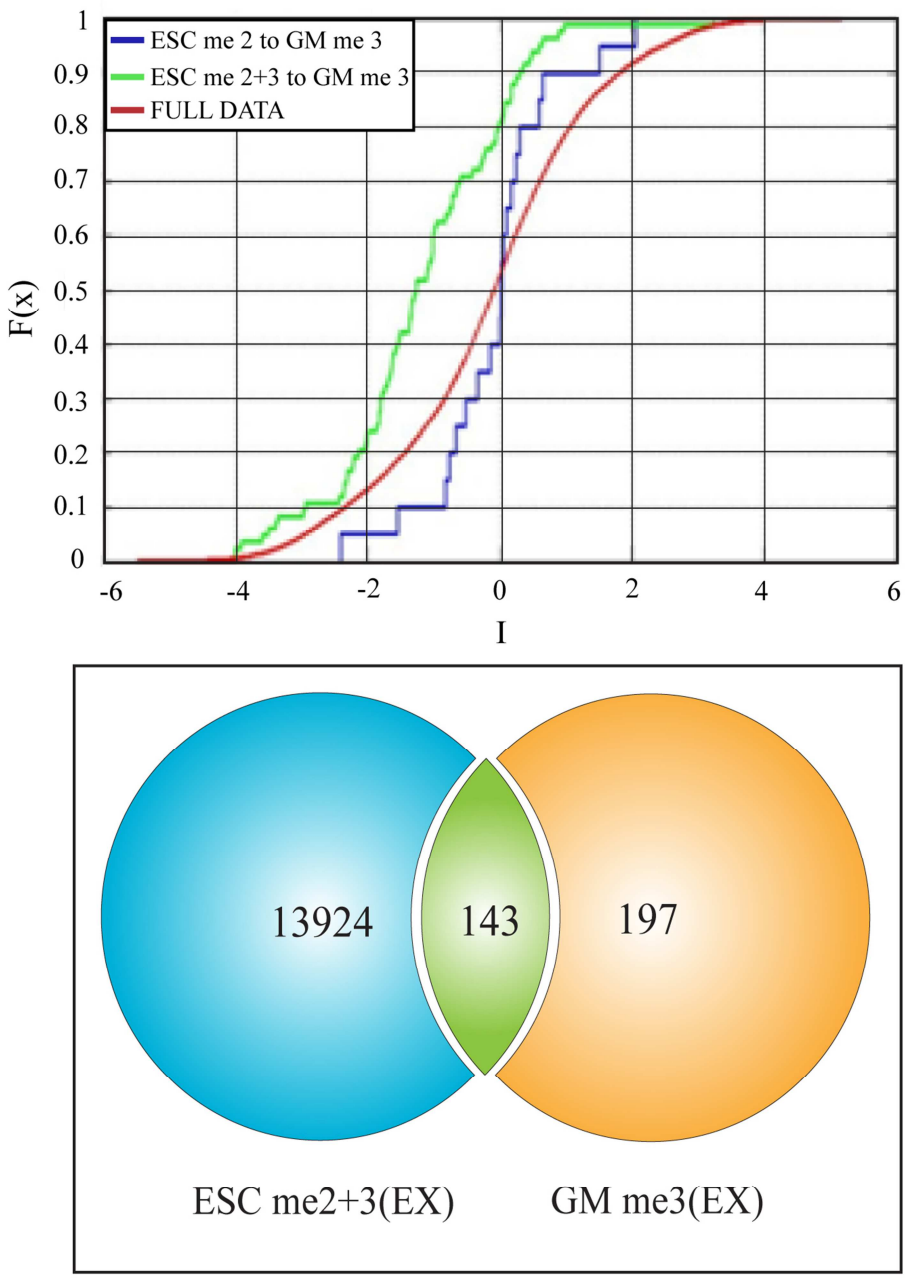


Figure. 4. *ESC_ me 2+3 converted to GM_me3 and cdf plot.*

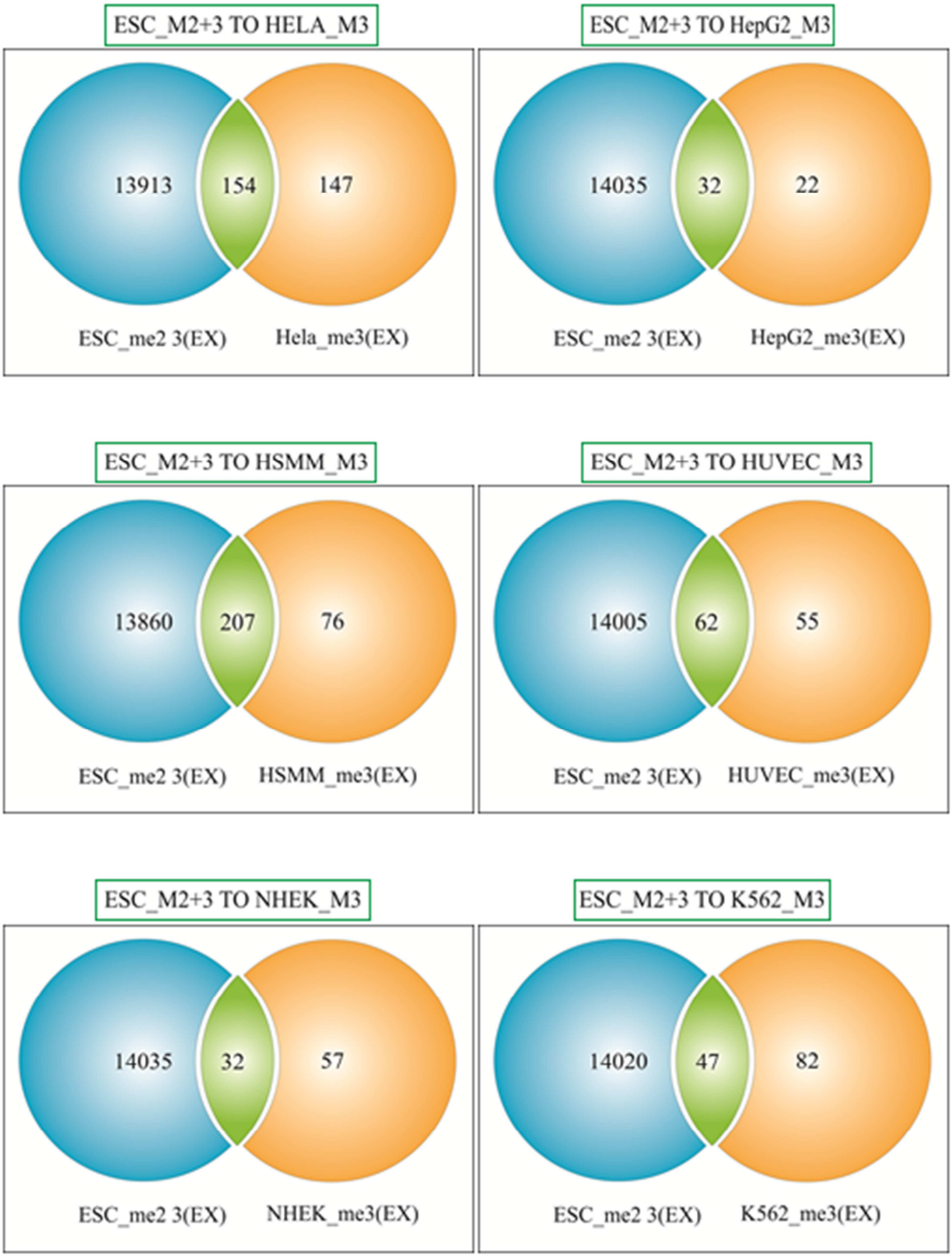
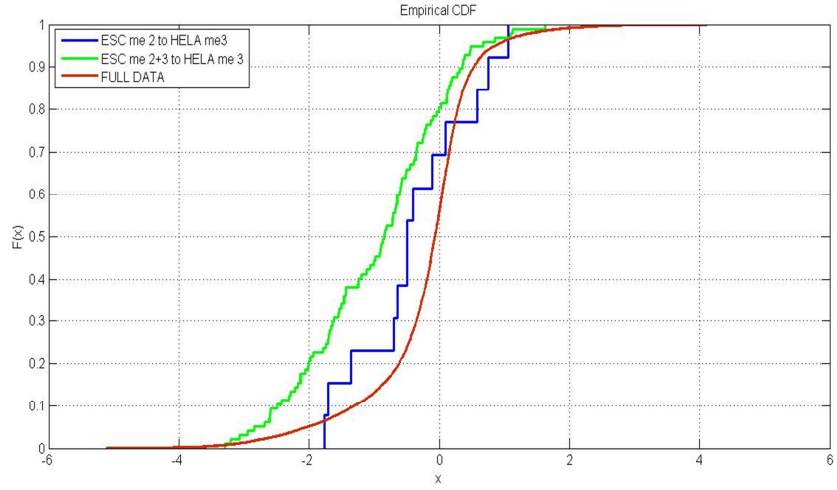
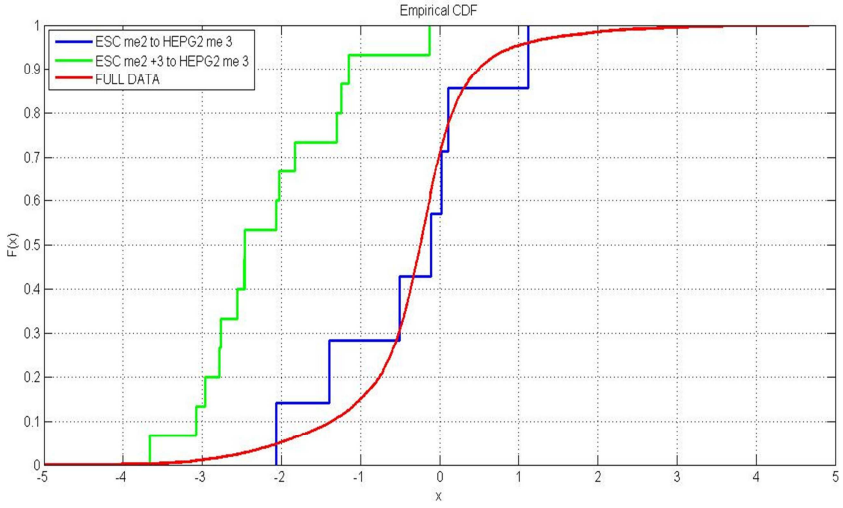
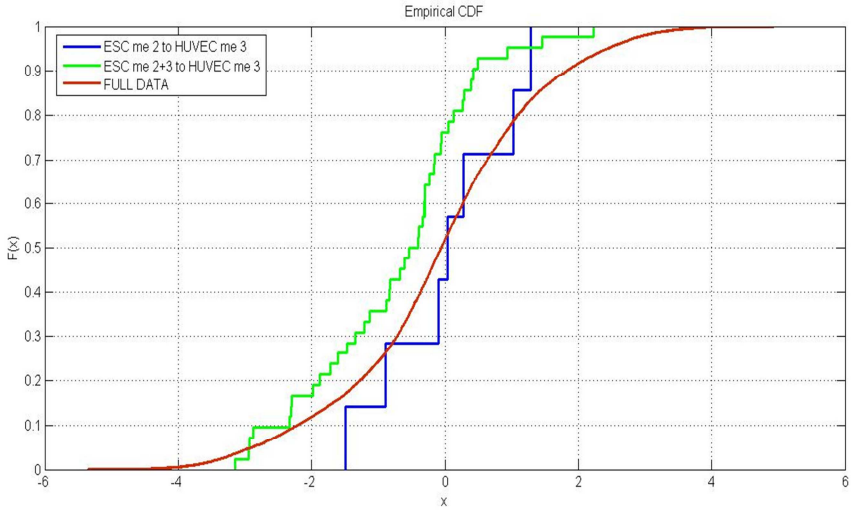


Figure S2. *_ ESC_ me 2 +3 converted to non-ESC_ me 3.*


Empirical CDF
ESC me 2 to HELA me3
ESC me 2+3 to HELA me 3
FULL DATA
F(x)
x


Empirical CDF
ESC me2 to HEPG2 me 3
ESC me2 +3 to HEPG2 me 3
FULL DATA
F(x)
x


Empirical CDF
ESC me 2 to HUVEC me 3
ESC me 2+3 to HUVEC me 3
FULL DATA
F(x)
x

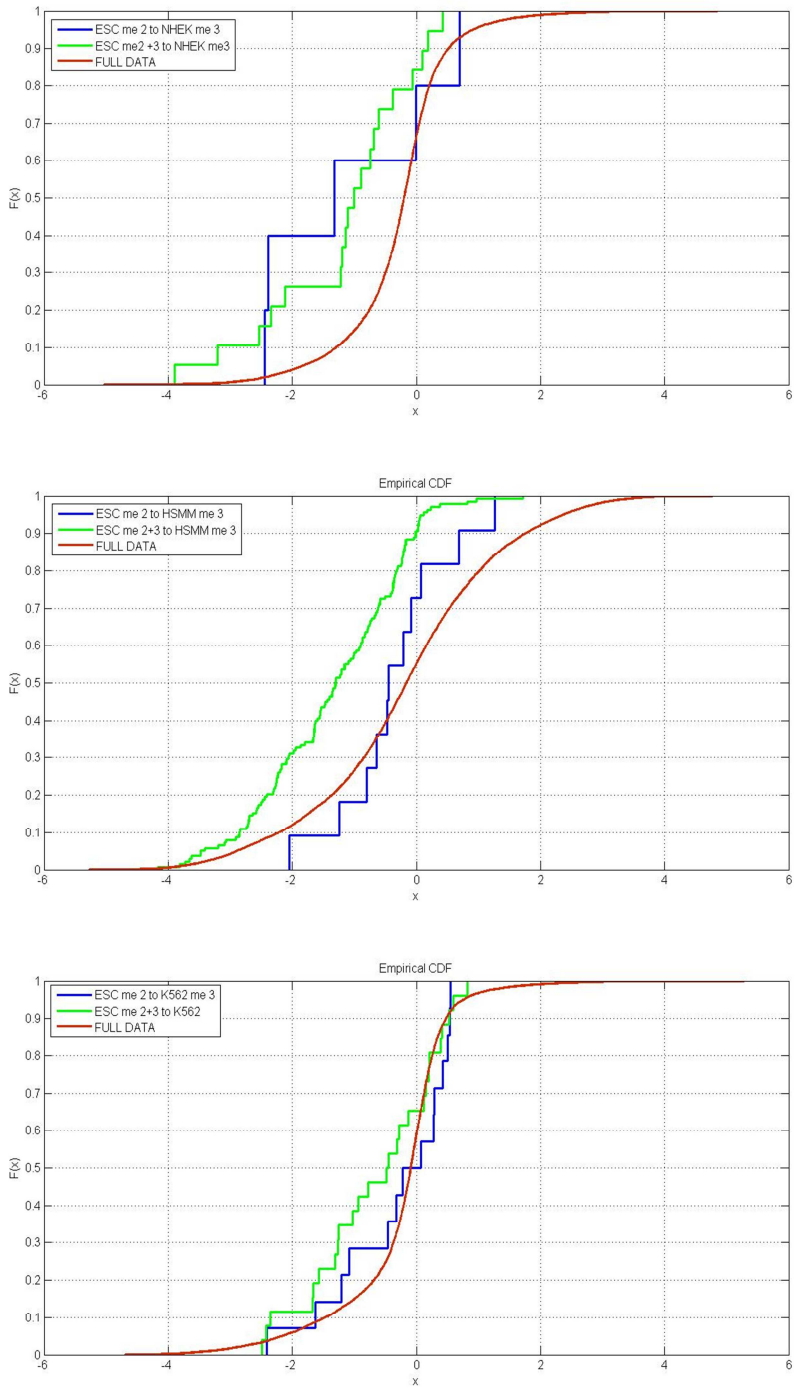


Figure S3. *Cdf plot of ESC_me 2+3 to non-ESC_me 3 vs ESC_me2 to non-ESC_me3.*

Table S2. *Esc_ me 2 + 3 converted to non - ESC _ me 3.*

	GM	HELA	HEPG2	HUVEC	HSMM	NHEK	K562
ESC	**143**	**154**	**32**	**62**	**207**	**32**	**47**
	SDC3	GPLT1A	RGS16	TTC39A	FGR	FAM72A	LEMD1-AS1
	FAM5C	FAM89A	KCNA3	KCND3	MEGF9	VENTX	CDK18
	CDK11B	POU3F1	PIK3CD	SCL35E2	COL8A2	KNDC1	KCNIP2
	LINC00862	OLFML3	CR1	GJA4	KCNA3	FFAR4	C10orf107
	TRABD2B	NBPF3	ATP12A	SPAG6	LEMD1-AS1	TIMM23	ANXA8
	TMEM240	CACHD1	EFS	SLC18A2	CDK11B	INPPL1	GUCY1A2
	PRRX1		GJD2	KNDC1	TMEM240	FAIM2	PRRG4
	BCAN	DNALI1	PRKCB	TIMM23	SLC35E2B	CRIP1	RACGAP1

	GM	HELA	HEPG2	HUVEC	HSMM	NHEK	K562
ESC	**143**	**154**	**32**	**62**	**207**	**32**	**47**
	TIMM23	FOXD3	CIDEA	DRD2	FAM229A	SKOR1	MMP14
	NKX6-2	RP11-312O7.2	OLFM2	KCNQ1	ANKRD65	JPH3	CA12
	PARG	NEBL	ZNF486	IGF2	PRKCZ	RAB11FIP4	GABARAPL3
	CNNM1	TLL2	PTPRN	MSI1	PRKCZ	EIF5A	FBXL16
	B3GAT1	NEURL	KCNK15	PARP16	WNT3A	DMPK	ELMO3
	LOC100499223	SFRP5	PMEPA1	DUOXA1	HTR6	DIRAS1	DNASE1L2
	GAL3ST3	PRAP1	HSPA12B	DUOXA2	RBP7		SKAP1
	BDNF-AS	CH25H	SEZ6L	RP11-89K11.1	FAM229A	DLGAP4	ARHGAP27
	SF3B2	TIMM23	WNT7A	DUOXA1	ZNF326	CRYAA	SGCA
	SF3B2	NPFFR1	PROK2	DUOXA1	TMEM183A		OR4D1
	PPP1CA	DYDC1	ZNF717	SKOR1	MFSD4	CRMP1	SLC16A3
	PACS1	GRID1	CASR		GABRD	PHOX2B	SLC16A3
	AX747517	CHAT		GSG1L	RP11-31207.2	PITX2	MADCAM1
	B4GALNT4	H2AFY2	EYA4	JPH3	SORCS1	OTP	ZNF880
	B4GALNT4	PALD1	FABP7	CA7	TUBGCP2	CLDN3	SPC25
	USP44		CALCR	WNT3	NKX6-2	NPTX2	EVX2
	GALNT9	NEURL	CARD11	ZNF232	TCERG1L	FAM71F1	SLC4A10
	BC042855	NKX2-3	ORPK1	C17orf102	RASGEF1A	MOB3B	ATP9A
	SRRM4	CNNM1	FAM86B1		ALOX5	LHX2	SDCBP2
	TMEM132C	RET	CHMP7	UNC45B	CHAT	FAM69B	UBE2C
	FAM155A	EMX2OS	BARX1	TMEM132E	FAM21A	RP11-262H14.4	RIPPLY3
	SALL2	SH2D4B	GALNT12	DMPK	C10orf11	ARX	SLC7A4
	REM2	CHAT	PDZD4	FAM83E	CHAT	GYG2	TOP3B
	PTGDR	HMX2	LRCH2	MAMSTR	TIMM23	HTR2C	IFT27
	FOXG1	TIMM23		BEST2	GPRIN2		CLDN5
	FAM189A1			APC2	ABCC8		AIFM3
	GJD2	LRP4-AS1		TFPI	B3GAT1		LOC400891
		MYBPC3			MTL5		ATOH1
	C15orf59	PHOX2A		AVP	PHOX2A		GNB2L1
	SMAD3	TP53I11		CHKB-CPT1B	BC038382		AC004041.2
	RP11-60L3.1	ASCL2		SULT4A1	ASCL2		FOXC1
	SYNM	SIGIRR		AC008103.1	MOB2		RSPH9
		SCL37A2		MAPK8IP2	KCNQ1		HIST1H4F
	TCRBV20S1	GRIK4		HIC2	GAL		
	BOLA2	NLRX1		TMEM40	CCKBR		HOXA-AS3
		GRIK4		RNF212	ART1		TRAF1
		LHX5		RP11-542P2.1	ST14		RALGDS
	CMTM2	BCL7A		LRAT	NELL1		RALGDS
	JPH3	TRIM9		NAT8L	LDHC		PAX5
	CACNAH1	OCA2		RAD1	OVOL1		
	SLX1B-SULT1A4	SEMA7A		MOCS2	RHOD		
	PIPTNM3	LINGO1		RAD1	VDR		
				PCDHB14	ACRBP		
	CDRT4	RNF111		NKD2			
	CDRT4			EGFLAM	ZNF385A		
	UNC45B	CCDC64B		DNAJC21	SLC5A8		
	AQP4-AS1	BOLA2		SIMC1	PLEKHG6		
		GNAO1		PTPRN2	DTX1		
	RAB27B			LHFPL3	SRRM4		
	ONECUT2	C16orf45		ANK1	TMEM132C		
	GALR1	SLX1B-SULT1A4		NKX6-3	ATP12A		
	DOK6			HMCN2	SOX1		
	ZFP112	BAIAP2-AS1		PGM5	RIPK3		
	IL11	SLC47A1		CYorf17	MDGA2		
	NOVA2	HIC1			ADCY4		
	FAM83E	ALOX15B			ABHD12B		
	SHC2	NXPH3			DLK1		
	ZNF577	MYO15B			HCN4		
	SHANK1	CELF4			RP11-89K11.1		
	MEX3D				SYNM		
	PALM3	FAM69C			RGS11		
	SHISA7	CLUL1			TOX3		
	ZNF626	GRP			HS3ST6		
	ZNF682	ADCYAP1					
	ZNF677	APCDD1			GSG1L		
	CNN1	CIDEA			NDRG4		

	GM	HELA	HEPG2	HUVEC	HSMM	NHEK	K562
ESC	**143**	**154**	**32**	**62**	**207**	**32**	**47**
	IGLON5	HMSD			NECAB2		
	ZNF90	PTPRS			HS3ST2		
	ZNF486	ZNF43					
	IGFBP5				C16orf11		
	OSR1	GIPC3			PRKCB		
	EPCAM	ZNF253			SCNN1G		
	MOGAT1	FN1			CBFA2T3		
	HOXD11	PAIP2B			SLC13A5		
	SH3BP4	AX747372			GRIN2C		
	MYCNOS	ARL4C			CYGB		
	GALNT13	CYS1			PDK2		
	KCNK3	SIX3-AS1			BAIAP2		
	SOX11	FBLN7			TBX21		
	TET3	TCF7L1			KCNH6		
		PCDP1			DSC3		
	MACROD2	KCNK3			AQP4-AS1		
	TCF15	HOXD8					
	SRXN1	LBH					
	SCRT2	BCAS1			ST8SIA3		
	BLCAP	RIMS4					
	SLC24A3	STMN3			NOVA2		
	DSCAM	SIRPA			RTBDN		
	C21orf37				CACNA1A		
	SLC7A4	AIRE			FLJ22184		
	PKDREJ	LIF			PALM3		
	CABP7	AC008103.1			CAMSAP3		
	CABP7	PANX2			EFNA2		
	MPPED1	HIC2			GRIN2D		
	TMEM40	PRR5			IGLON5		
	GHSR	SEPT5-GP1BB			MUM1		
	ERC2	MYLK			CELF5		
	C3orf14	LAMP3			GIPC3		
	PPARG	CSPG5			GIPC3		
	FBXO40	AMT			ST6GAL2		
	NAALADL2	SST			CCDC74C		
	PPP2R2C	ZIC4			HAAO		
	CPLX1	ACPL2			TRIM54		
	ZNF732	CACNA1D			NTSR2		
	LRBA	PRKCD			ARHGEF4		
	ISL1	ZNF827			THSD7B		
	ADAMTS16	CPLX1			GALNT13		
	TMEN171	ATOH1			B3GNT7		
	PCGHDA2	NAT8L			GAREML		
	PCDHGA3	PFN3			CCDC74A		
	SIM1	ADAMTS16			AMER3		
	LAMA4	RXFP3			DCDC2C		
	KIAA0408	GRM4			RAB6C		
	C6orf58	HIST1H3I			KCNQ2		
	DLX5	PLEKHG1			HRH3		
	PDE1C	ATAT1			RSPO4		
	STAG3L2	ADAP1			PAK7		
	GPR85	IRF5			ANKRD60		
	RELN	MUC3B			OVOL2		
	NPTX2	MUC3A			TCF15		
	RNF19A	TMEM178B			TMEM74B		
	STC1	FAM86B2			ZNF512B		
	HAS2	TP53INP1			SOX18		
	AF186192.5	KCNK9			OXT		
	ELAVL2	MTSS1			SYNDIG1		
		SNTB1			SSTR4		
	RAB14	LOC100506990			TCEA2		
	PNCK	ENTPD2			DSCAM		
	HEPH						
	SLC6A8	CNTFR			TSPEAR		
	DUSP9	FRRS1L			GSC2		
	LPAR4	FBP1			SLC16A8		

	GM	HELA	HEPG2	HUVEC	HSMM	NHEK	K562
ESC	**143**	**154**	**32**	**62**	**207**	**32**	**47**
	GPR173	PAX5			PKDREJ		
	ARMCX4				USP41		
	EGFL6	ARID3C			PANX2		
		FGD3			CABP7		
		TOR1B			MGAT3		
		LMX1B			TBX1		
		SUSD3			SHANK3		
		FAM69B			NUP210		
		CERCAM			ZNF385D		
		PHYHD1			WNT7A		
		GRIN1			NEK10		
		COL15A1			CAMKV		
		ARAF			MST1R		
		NDP			SYN2		
					LRRC3B		
					TMIE		
					OTOP1		
					CPLX1		
					RNF212		
					ZNF732		
					RP11-542P2.1		
					LRAT		
					SHISA3		
					TPPP		
					KCNN2		
					DOCK2		
					PSD2		
					SIMC1		
					SPINK9		
					OLIG3		
					PPP1R11		
					TFAP2B		
					RIPK1		
					TMEM151B		
					HIST1H4I		
					MMD2		
					ZYX		
					UNCX		
					MUC17		
					SRRM3		
					IKZF1		
					GRM3		
					OPRK1		
					FAM86B2		
					KCNV1		
					KCNK9		
					LY6H		
					FAM86B1		
					DMTN		
					GPT		
					LOC100506990		
					XKR4		
					KIF12		
					SEMA4D		
					NOXA1		
					LMX1B		
					SUSD3		
					RP11-262H14.4		
					MAOB		
					MTMR8		
					GPC3		
					ATP2B3		
					NRK		
					GABRQ		
					PNMA3		
					CYorf17		

A total of 143 genes with me2+3 methylations in the H1-hESC cell line have been converted into me3 methylation when the H1-hESC differentiates into a GM 12878 cell line. The expression values of the exclusively me2 modified as well as both me2 and me3 modified genes that were identified to have lost their me2 modification and gained me3 modification during the conversion of ESC line into the GM cell was noted. The log 10 values of ratios of expression values of genes with exclusive me3 methylation in the GM cell line to that of the corresponding genes in the ESC line with both me2 and me3 methylations was generated. This ratio of FPKM values was generated for all the 30 genes that lost its me2 and gained me3 during differentiation of the H1-hESC to GM cell line. Both the ratios of FPKM values were plotted together, generating a CDF plot.

Table 2. *Gene Ontology and KEGG Pathway Data.*

CELL LINES	UCSCGENE NAME		KEGG PATHWAY ASSOSIATED WITH GENE	
	Protein Binding	**ATP Binding**	**GENE NAME**	**PATHWAYS**
GM12878	ZRANB1 HCN2 ATPAF1	RFC5 SRMS SRXN1	F8	Complement& coagulation cascades
HELA	DLG5 DNMT3L MYOZ3 NR5A1 GCK	SRXN1 GCK	FGF13	MAPKSignaling Pathway Regulation of actincytoskeleton Melanoma
HEPG2	PDX41 GCK	PDX41 GCK	GCK	Glycolysis/ Neoglucogenesis Galactose Metabolism Starch & Sucrose Metabolism Insulin Signalling Pathway Type 2 Diabetes Mellitus Maturity onset diabetes of young
HUVEC	GSN F8		GSN	Regulation of actin cytoskeleton
HSMM	MYBPH USP9Y KCNA7 GLRA1	RFC5 SRMS	PVPL1	Cell Adhesion molecules Adherens Junctions
K562	USHBP1 TANK RNF14 FOXP3 NYX FGF13		RHOH	Leucocytetrans-endothelial migration
NHEK	RHOH PVPL1 GCK	SRXN1 GCK		

Gene Ontology revealed that Protein binding and ATP binding genes are mostly responsible for me2 to me3 modification. Genes for which pathways can be detected from KEGG are also represented.

2.3. Gene Ontology Analysis

To determine the function of the genes, which under goes transfiguration, we performed gene ontology analysis. We also retrieved the related pathways from KEGG database. The genes and their related function and pathways with respect to all the seven cell lines are noted in Table 2. On the level of gene ontology molecular function classification, these genes are mostly associated with protein binding, nucleotide binding, DNA binding and ATP binding. Other than that, signaling and receptor activity, metal ion binding and phosphorylation-dephosphorylating action can be correlated with these genes. These genes can be either cellular (nucleus/cytoplasm) or part of extracellular environment. GCK (Glucokinase) gene is found to be the most common gene shared by almost all the cell lines.

3. Discussions

In this work, we investigated the dynamic relationship of histone modifications and genomic sequence contexts to DNA methylation patterns in 8 cell lines based on the marks at TSS. Although previous studies have found that histone modifications were correlated with DNA methylation, our work provides a genome-wide insight into their genomic

region-specific and cell type-specific relationships. Recently, many whole genome DNA methylation profiles have been produced by the ENCODE project has provided a wealth of histone modification profiles by ChIP-Seq. Compared with the relationship between H1-hESC and others demonstrate that DNA methylation landscapes of these two cell types change dramatically. Venn Diagrams generated to identify the exclusively H3K4me2 and H3K4me3 modified genes in all the above-mentioned cell lines concluded that H1-hESC have more exclusively H3K4me2 modified genes as compared to most of the non embryonic cells [Fig. 1, Table1]. Further, it has also been identified that most of the exclusively H3K4me2 modified as well as both H3K4me2+3 modified genes in the H1-hESC lose their me2 methylation and gain me3 methylation in non embryonic cells during stem cells differentiation into non embryonic stem cells [Fig. 2, Table S1]. However, the reverse process, that is, switch over of me3 marked tome2 or both me2+3 during the differentiation process was seen to be negligible [Fig. 5]. [Table S3]. It has been further identified that this loss of me2 methylation and the gain of me3 methylation during differentiation ultimately leads to an up regulation of the genes in the differentiated cell lines. We show here the prominent change of histone modification status, particularly Lysine 4 methylation, when there is a transition from embryonic stem cell to seven different cell lines. We observe here that when H1-hESC diverges to normal and cancer cell lines, the number of Lysine 4 tri-methylated genes increases compared to Lysine 4 di-methylated genes, along with the overall expression level of genes. It is expected that with transition from embryonic state to differentiated state, more genes start functioning and thus need to be active. As a result it is justified that in differentiated cell lines there will be a predominance of H3K4me3 [26], as it is well known that H3K4me3 is associated with only active state of genes whereas H3K4me2 marks both active and inactive genes [27]. On the other hand, genes responsible to maintain stemness of cell are expected to show prevalence of H3K4me3 in H1-hESC in comparison to other cell lines. To confirm this, we checked the expression status and modification pattern of SOX2 and NANAOG. We observed that SOX2 and NANOG are highly expressed in H1-hESC as compared to rest of the seven cell lines. Moreover, Lysine 4 of both these genes are tri-methylated abundantly in H1-hESC but not in others. Strikingly, for NANOG, repressive marker H3K27me3 is present in all cell lines except H1-hESC. This scenario reinforces that H3K4me3 is required for gene activation [28, 29]. Along with the association of Lysine 4 methylation (di- and tri-) state with gene expression status, we also correlate our observation with the gene ontology classification at individual gene level [30]. In differentiated cell lines, though mostly H3K4me2 persists along with H3K4me3, there are genes, which completely lose the di-methylation status and become tri-methylated when the cell line identity changes from H1-hESC to other cell lines. Although the phenomenon experienced by these seven cell lines are same, but the gene sets involved are mostly non-overlapping. By applying this method to find the contra-variation relevant genes between embryonic, normal and cancer cell types may help to obtain potential cancer-related marks. Therefore, it is essential to identify the contra-variations of paired cell types to gain new understanding of biological processes from the large amounts of data that is now publicly available. We expect a crosstalk between the change of methylation status and gene functionality, as all these functions can be allied to transcriptional regulation and gene activation, which once again is linked to H3K4me3 mark.

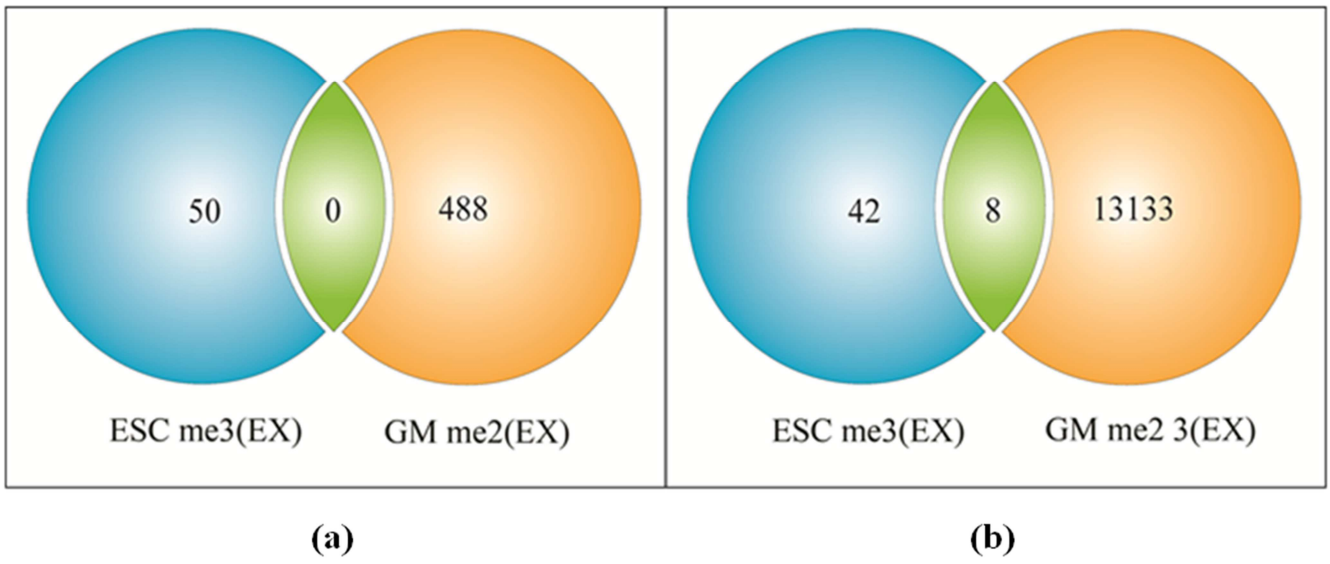


Figure 5. *Reverse epigenetic phenomenon.*

5a: Venn diagram showing no exclusively H3K4me3 methylated ESC genes converted to exclusive H3K4me2 methylation during the differentiation of ESC into GM cell line.
5b: Venn diagram showing only 8 exclusively H3K4me3 methylated ESC genes converted to H3K4me2+3 methylation during the differentiation of H1-hiESC into GM cell line.

Table S3. *ESC_ Me 3 CONVERTED TO NON- ESC_ Me 2.*

	GM	HELA	HEPG2	HUVEC	HSMM	NHEK	K562
ESC	**0**	**5**	**4**	**2**	**7**	**5**	**5**
		OR2T1	TXNDC2	TUSC5	CPLX4	CPLX4	FCN3
		ANXA8	SLC2A9	SLC2A9	AP002414.1	CGB8	IGFALS
		FAM25C	ARHGEF38		TPTE	SLC2A9	SPSB3
		IGF2BP2	TMEM184A		IGF2BP2	SORBS2	AP002414.1
		ARHGEF38			ARHGEF38		ANKRD62
					TMEM184A		

4. Methods

In accordance with the assumption made earlier in this study, that is, if a gene has at least one protein coding transcript then it is considered to be always protein coding, ENCODE transcript hg19 revealed 19,999 protein coding genes and 10,419 non protein coding genes in the human genome. For the purpose of this study, only the protein coding genes have been taken into consideration. The lengths of these 19,999 identified protein coding genes have been determined by the second assumption made in this study, that is, if a protein coding gene has multiple transcripts then the gene length is considered by connecting extreme 5′ and extreme 3′ coordinate among those transcripts.

UCSC provides Chip-Seq data of all the H3K4me2 and H3K4me3 present in 8 different cell lines of the human genome hg19. Computationally we selected those H3K4me2 and H3K4me3 markers that fall within the +/- 2KB region of the TSS of the protein coding genes. The genes with H3K4me2 modifications in their TSS have been considered "me2" modified, while those with H3K4me3 modifications in their TSS have been termed "me3" modified and finally, those with both me2 and me3 modifications in their TSS have been termed me2+3 modified. Such genes have been identified for the entire 8 cell lines used in this study, namely, H1-hESC, GM12878, HELA, HepG2, HUVEC, HSMM, NHEK, K562.

RNA-Seq methodology

Quantification of mRNA expression or reads in each cell lines were PolyA-trimmed. Mapping of reads to the human genome (hg19) was performed with TOPHAT (https://ccb. jhu. edu/software/tophat/index. shtml). The mapping coordinates of each read were overlapped with the refseq annotation track from the UCSC table browser (http://genome. ucsc. edu/cgibin/hgTables?command=start) to quantify mature mRNA expression. Normalization and test for differential expression was performed in cufflinks (http://cole-trapnell-lab. github. io/cufflinks/), cuffmerge and cuffdiff the statistical programming language R (www. rproject. org. GO-enrichment analysis was performed using the GO-enrichment toolkit from http://genxpro. ath. cx. The divergence of CDF plot was based on the KS test among those dynamic genes with a *P* value less than 0.05.

References

[1] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, andT. J. Richmond, "Crystal structure of the nucleosome core particle at 2. 8A resolution" Nature, vol. 389, pp. 251–260, September 1997.

[2] T. Kouzarides, "Chromatin modifications and their function", Cell, vol. 128, pp. 693–705, February 2007.

[3] R. Marmorstein, and R. C. Trievel, "Histone modifying enzymes: structures, mechanisms, and specificities" Biochim. Biophys. Acta, vol. 1789, pp. 58–68, January 2009.

[4] J. F. Couture, andR. C. Trievel, "Histone-modifying enzymes: encrypting an enigmatic epigenetic code" Curr. Opin. Struct. Biol, vol. 16, pp. 753–760, October 2006.

[5] V. W. Zhou, A. Goren, andB. E. Bernstein, "Charting histone modifications and the functional organization of mammalian genomes" Nat. Rev. Genet, vol. 12, pp. 7– 18, November 2011.

[6] J. C. Black, andJ. R. Whetstine, "Chromatin landscape: methylation beyond transcription" Epigenetics, vol. 6, pp. 9-15, January 2011.

[7] D. K. Pokholok, C. T. Harbison, andS. Levine, "Genome-wide map of nucleosome acetylation and methylation in yeast" Cell, vol. 122, pp. 517–527, August 2005.

[8] E. Vallas, S. Sanchez-Molina, andM. A. Martinex-Balbaz, "Role of histone modifications in marking and activating genes through mitosis" J Biol. Chem.,vol. 280, pp. 42592-42600, December 2005.

[9] T. Y. Roh, W. C. Ngau, and K. Cui, "High-resolution genome-wide mapping of histone modifications*"* Nat. Biotechnol, vol. 22, pp. 1013–1016, August 2004.

[10] W. E. Farrell, "Epigenetics of pituitary tumors: an update" CurrOpinEndocrinol Diabetes Obes. vol. 21, pp. 299-305, August 2014.

[11] N. D. Heintzman, R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, and K. A. Ching, "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, " Nat. Genet, vol. 39, pp. 311–318, March 2006.

[12] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, and K. Plath, "A bivalent chromatin structure marks key developmental genes in embryonic stem cells" Cell, vol. 125, pp. 315–326, April 2006.

[13] T. Y. Roh, S. Cuddapah, and K. Cui, "The genomic landscape of histone modifications in human T cells" Proc. Natl. Acad. Sci. USA, vol. 103, pp. 15782– 15787, October 2006.

[14] A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, andK. Zhao, "High-resolution profiling of histone methylations in the human genome" Cell, vol. 129, pp. 823–837, May 2007.

[15] G. E. Crawford, I. E. Holt, and J. Whittle, "Genome-wide mapping of DNasehypersensitive sites using massively parallel signature sequencing (MPSS)", Genome Res, vol. 16, pp. 123–131, January 2006.

[16] A. Visel, M. J. Blow, and Z. Li, "ChIP-seq accurately predicts tissue-specificactivity of enhancers" Nature, vol. 457, pp. 854–858, February 2009.

[17] N. D. Heintzman, G. C. Hon, and R. D. Hawkins, "Histone modifications at human enhancers reflect global cell-type-specific gene expression" Nature, vol. 459, pp. 108–112, May 2009.

[18] K. Ishihara, M. Oshimura, and M. Nakao, "CTCF-dependent chromatin insulator is linked to epigenetic remodeling" Mol. Cell, vol. 23, pp. 733–742, September2006.

[19] P. A. Jones, and D. Takai, "The role of DNA methylation in mammalian epigenetics", Science, vol. 293, pp. 1068–1070, August 2001.

[20] T. K. Kim, M. Hemberg, J. M. Gray, and et al "Widespread transcription at neuronal activity-regulated enhancers" Nature, vol. 465, pp. 182–187, May 2010.

[21] K. Polyak, "Breast cancer: origins and evolution," J. Clin. Invest,vol. 117, pp. 3155– 3163, November 2007.

[22] D. S. Johnson, A. Mortazavi, and R. M. Myers, "Genome-wide mapping of in vivo protein-DNA interactions" Science, vol. 316, pp. 1497–1502, June 2007.

[23] T. Y. Roh, andK. Zhao, "High-resolution genome wide chromatin modifications by GMAT"Methods Mol. Biol, vol. 387, pp. 95-108, May 2007.

[24] A. Sanyal, B. R. Lajoie, G. Jain and J. Dekker. "The long-range interaction landscape of gene promoters" Nature, vol. 6, pp. 109-113, September 2012.

[25] T. Kim, Z. Xu, S. Clauder-Münster, L. M. Steinmetz, and S. Buratowski, "Set3HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics, " Cell, vol. 150 (6), pp. 1158-69, September2012.

[26] R. Schneider, A. J. Bannister, F. A. Myers, A. A. Thorne, C. Crane-Robinson, andT. Kouzarides, "Histone H3 lysine 4 methylation patterns in highereukaryotic genes, " Nature Cell Biology, vol. 6, pp. 73-77. January 2004.

[27] H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, E. N. C TolgaEmre, S. L. Schreiber, J. Mellor, andT. Kouzarides, "Active genes are tri-methylated at K4 of histone H3, " Nature, vol. 419, pp. 407-411, September2002.

[28] C. M. Koch, R. M. Andrews, and P. Flicek, "The landscape of histone modifications across 1% of the human" Genome Res, vol. 17, pp. 691-707, June 2007.

[29] B. E. Bernstein, E. L. Humphrey, R. L. Erlich, R. Schneider, P. Bouman, J. S. Liu, T. Kouzarides, and S. L. Schreiber, "Methylation of histone H3 Lys 4 in coding regions ofactive genes, " Proc. Natl. Acad. Sci. USA, vol. 99, pp. 8695–8700, June 2002.

[30] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium" Nat Genet, vol. 25 (1), pp. 25-9, May 2000.