

RESEARCH ARTICLE

Open Access

Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions

Bo Wen^{1,4†}, Hao Wu^{2†}, Yuin-Han Loh^{3†}, Eirikur Briem¹, George Q Daley³ and Andrew P Feinberg^{1*}

Abstract

Background: The organization of higher order chromatin is an emerging epigenetic mechanism for understanding development and disease. We and others have previously observed dynamic changes during differentiation and oncogenesis in large heterochromatin domains such as Large Organized Chromatin K (lysine) modifications (LOCKS), of histone H3 lysine-9 dimethylation (H3K9me2) or other repressive histone posttranslational modifications. The microstructure of these regions has not previously been explored.

Results: We analyzed the genome-wide distribution of H3K9me2 in two human pluripotent stem cell lines and three differentiated cells lines. We identified > 2,500 small regions with very low H3K9me2 signals in the body of LOCKs, which were termed as euchromatin islands (EIs). EIs are 6.5-fold enriched for DNase I Hypersensitive Sites and 8-fold enriched for the binding of CTCF, the major organizer of higher-order chromatin. Furthermore, EIs are 2–6 fold enriched for differentially DNA-methylated regions associated with tissue types (T-DMRs), reprogramming (R-DMRs) and cancer (C-DMRs). Gene ontology (GO) analysis suggests that EI-associated genes are functionally related to organ system development, cell adhesion and cell differentiation.

Conclusions: We identify the existence of EIs as a finer layer of epigenomic architecture within large heterochromatin domains. Their enrichment for CTCF sites and DNase hypersensitive sites, as well as association with DMRs, suggest that EIs play an important role in normal epigenomic architecture and its disruption in disease.

Keywords: Epigenetics, H3K9me2, Euchromatin islands, CTCF, DNA methylation

Background

Epigenetics involves information retained during cell division other than DNA sequence per se, and both DNA methylation and post-translational modifications of histones are fundamental in understanding normal development and disease [1-3]. Genome-scale localization of histone modifications had been extensively mapped in mammalian genomes [4-10]. While most of these studies focused on local regulatory elements such as promoters and enhancers, global organization of the chromatin has not been well understood.

Recent evidence indicates that repressive histone modifications form large scale domains in both mouse and human genomes. We had previously identified large blocks of H3 lysine 9 dimethylation (H3K9me2), termed Large Organized Chromatin K9-modifications (LOCKS), which affect more than 40% of the mouse genome in liver cells [11]. LOCKs significantly overlap with lamina-associated domains (LADs) [12] and are associated with domain-wide gene silencing in a tissue-specific manner. Importantly, both coverage and domain size of LOCKs increase upon differentiation of mouse embryonic stem cells (ESCs) [11]. On the other hand, genome-scale reduction of LOCKs was seen in epithelial-to-mesenchymal transition (EMT) induced by TGF- β treatment of mouse hepatocytes, a process in which cells gain stem cell-like and malignant-type traits [13]. Similarly, large blocks of other repressive marks (H3K9me3 and H3K27me3) are also found to expand in human lung fibroblasts compared

* Correspondence: afeinberg@jhu.edu

[†]Equal contributors

¹Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Full list of author information is available at the end of the article

with human ESCs [14], and those blocks/LOCKS expand in breast cancer cells relative to normal epithelial cells [15]. Furthermore, large H3K9me3 and H4K20me3 blocks specifically coat olfactory receptor (OR) gene clusters in mouse olfactory epithelium but not in liver [16]. Taken together, these data demonstrate that large heterochromatin domains are highly dynamic in differentiation and tumorigenesis.

DNA methylation has been tightly linked to development and disease [1]. We previously reported that differentially methylated regions (DMRs) related to tissue specificity (T-DMRs), colon cancer (C-DMRs) and reprogramming (R-DMRs) have largely common targets in the genome and are strongly associated with local regulation of adjacent genes [17,18]. Whole genome bisulfite sequencing had found partial methylated domains (PMDs) which are highly methylated in human ESCs but partially methylated in fibroblasts [19]. Similar large hypomethylation blocks relative to normal cells have been identified in colon cancer [20] and breast cancer cells [15], and loss of methylation in these regions is accompanied by acquisition of large domains of H3K9me3 and H3K27me3 [15].

Surprisingly, the relationship of H3K9me2 LOCKs/blocks to DMRs has not been previously assessed. In the course of this investigation, we identified a new chromatin unit we term “euchromatin island” which may serve as a fulcrum between DNA methylation and chromatin in development.

Results

We analyzed whole genome distribution of H3K9me2 by ChIP-chip using a highly specific monoclonal antibody in two human pluripotent stem cell (PSC) lines (human ESC H1, human iPSC ADA-38) and three primary differentiated cell lines: human astrocytes (HA), human aortic endothelial cells (HAEC) and human pulmonary fibroblasts (HPF). For differentiated cells, we used early passages of primary cells instead of immortalized cell lines to avoid potentially aberrant epigenetic changes due to long time cell culture and immortalization of the cells [21]. These differentiated lines represent three germ layers: ectoderm (HA), mesoderm (HAEC) and endoderm (HPF).

We normalized the ChIP-chip data as described [11] to calculate the log₂ ratios of ChIP/Input comparable among cell types. By using the 90th quantile as a cutoff to define large domains, the genome coverage of LOCKs was found to increase from 17.5-24% in PSC lines, to 39.3-44.8% in differentiated cells, and the average sizes of LOCK expanded from 142–171 kb in PSC lines, to 233–315 kb in the differentiated. The trends were the same when we used different cutoffs to define LOCKs (Additional file 1: Table S1), consistent with our previous

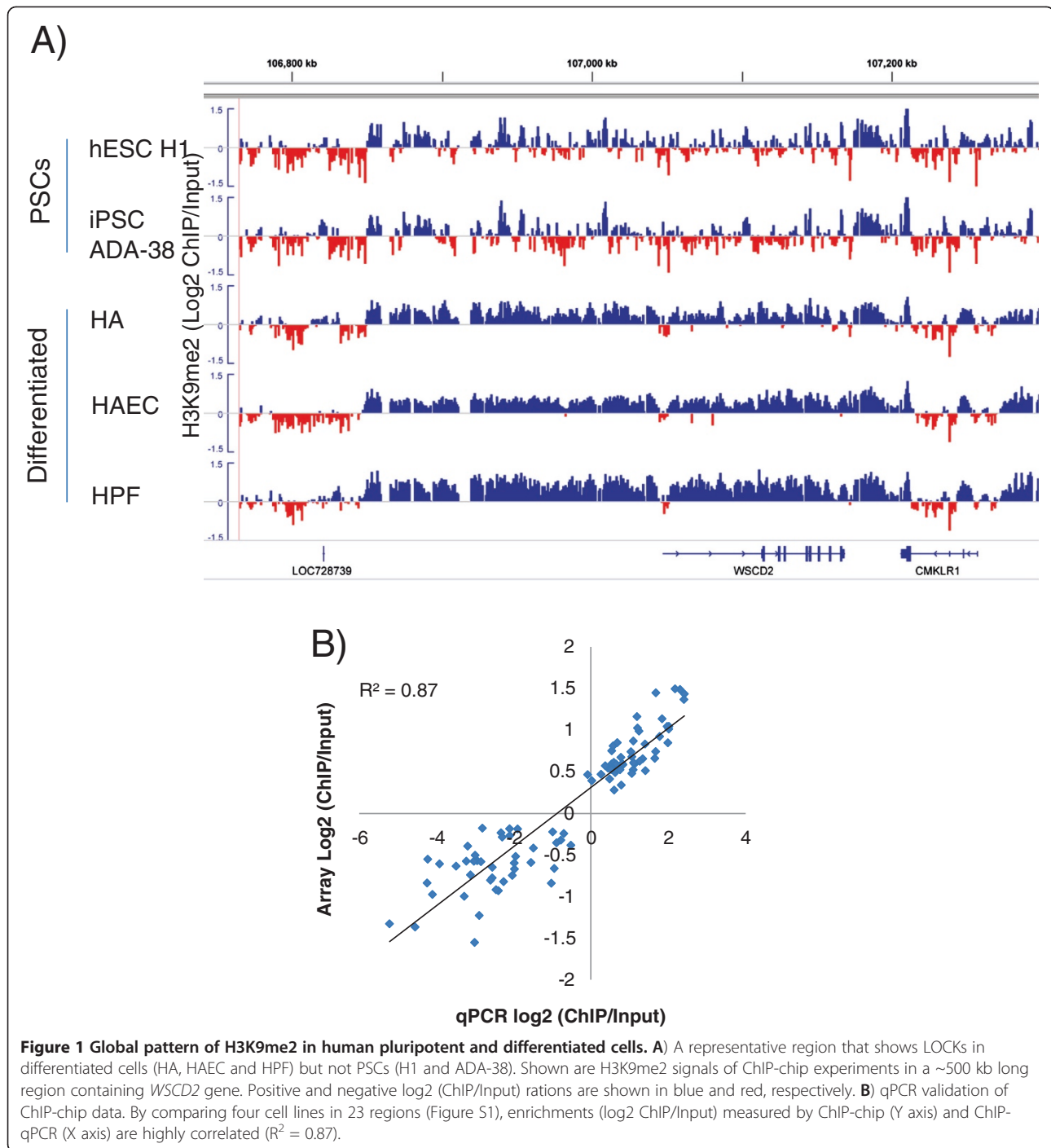
findings that LOCKs increase after mouse ESC differentiation [11]. For example, in the *WSCD2* gene locus, only some small H3K9me2 peaks can be seen in the PSCs, but the H3K9me2 enriched regions expanded to ~350 kb long and cover the whole gene body and its flanking regions in the differentiated cells (Figure 1A).

To validate the ChIP-chip data, we performed quantitative PCR (qPCR) on 23 loci using independently prepared ChIP and input DNA samples from four cell types. For all the cases, the quantitative differences of H3K9me2 enrichments within and among samples detected by ChIP-chip were well validated by qPCR (Additional file 2: Figure S1). Overall, the ChIP/Input log₂ ratios of microarray (ChIP-chip) and qPCR were strongly correlated ($R^2 = 0.87$, Figure 1B), indicating that the ChIP-chip data are of high quality.

To reveal the relationship between dynamics of H3K9me2 and DNA methylation on a large scale, we compared genome-wide distributions of LOCKs (this study), PMDs in fibroblasts [14], and DNA hypomethylation blocks in colon cancer [20]. LOCKs in fibroblasts (HPF) largely overlap PMDs (Additional file 3: Figure S2A), and overall 61.5% regions of LOCKs in HPF coincide with PMDs ($p < 0.001$, based on 1,000 permutations), and H3K9me2 signals in the regions of PMDs are higher than non-PMD regions (Additional file 3: Figure S2B). Furthermore, more than 80% LOCK regions in HPF were contained within DNA hypomethylation blocks found in colon cancer tissues (Additional file 3: Figure S2). Thus, our data support a strong correlation between LOCKs and DNA hypomethylation blocks in human cells.

On closer examination of the microstructure of the LOCKs, we noticed that many small H3K9me2-depleted regions are located in the body of LOCKs. These regions are a few kb in length, and away from the LOCK boundaries. We found that these regions are abundant in the genome, and they appear to be associated with open chromatin (see below). Thus, we termed these regions Euchromatin Islands (EIs). As an example, an EI was found near the transcription start sites (TSSs) of the cadherin 11 gene (*CDH11*, Figure 2A), of which epigenetic disruption was associated with metastasis of human cancers [22]. Other examples of EIs include within the gene body of *PDILT*, a testis-specific gene; and downstream of the glycoprotein 2 (*GP2*) gene (Figure 2B).

Then we developed a statistical algorithm to identify EIs genome-wide (see Methods). We identified 758 to 2,465 EIs across cell types, with average sizes from 4.4 to 5.9 kb (Table 1 and Additional file 4: Table S2). These EIs form strong dips relative to adjacent LOCK regions as demonstrated by average H3K9me2 densities (Additional file 5: Figure S3). We have performed replicates on one array of the “Mouse ChIP-chip 2.1M



Whole-Genome Tiling sets”, which covers 10% of the genome. The EIs detected from the replicate experiments have high concordance with the ones from whole genome arrays (Additional file 6: Figure S4). Percentage of EIs detected from whole genome arrays that can also be detected from replicate arrays are 76.3% for H1, 74.1% for ADA-38, 63.3% for HA, 84.4% for HAEC and 71.4% for HPF. To exclude the possibility that EIs resulted from lack of histones in these regions, we

plotted nucleosome density around EIs, and no depletion of nucleosomes was observed in EIs (Additional file 7: Figure S5), indicating that the observation of EIs is not due to nucleosome positioning.

Among the five cell lines, 4.6% to 12.7% of EIs coincided with transcriptional start sites (TSSs), which associated with 60 to 409 genes across cell types. Compared to random, the enrichment at TSS ranged from 2.7 (in ADA-38) to 7.8 (HA), with randomization p-values

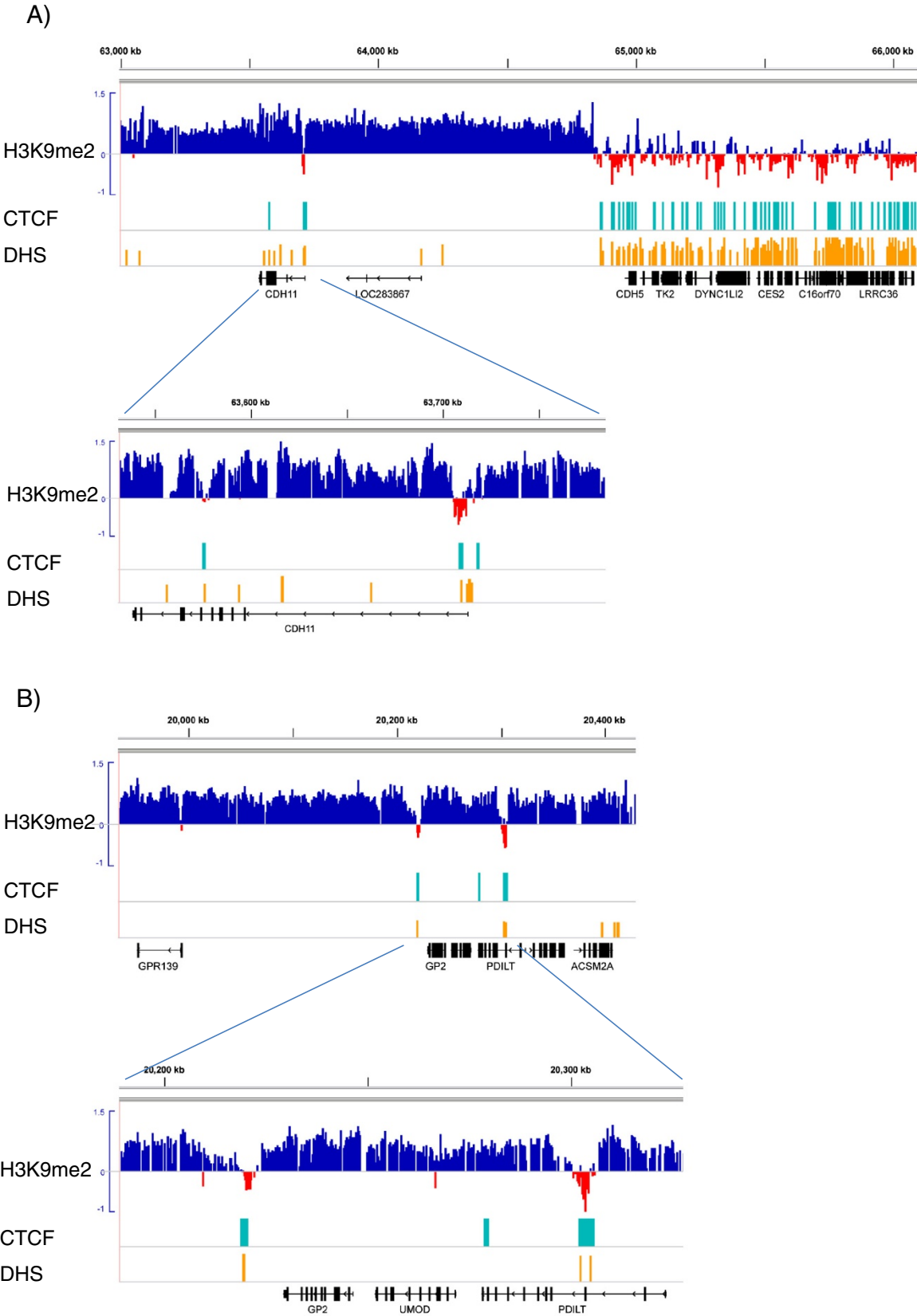


Figure 2 (See legend on next page.)

(See figure on previous page.)

Figure 2 Euchromatin islands (EIs) in LOCKs overlap CTCF interacting regions and DNase hypersensitive sites (DHSs). H3K9me2 log ratios of HAEC are shown on the top track. CTCF binding regions and DHSs of HUVEC are denoted as light blue and orange bars, respectively. EIs are small regions with strong negative signals within the body of LOCKs. **A)** Shown is a 3 Mb long region (top) containing CHD11 genes (zoomed-in view on the bottom), a member of the cadherin gene family. CTCF interacting regions and DHSs are highly depleted in the H3K9me2 blocks (LOCKs), but overlap the EI located near the TSS of *CDH11* gene. **B)** Additional examples of EIs near GP2 and PDILT genes.

$< 10^{-3}$ for all cell lines (Table 1). We further investigated the spatial relationship between EI and CpG islands (CGI). We found that 4.7% to 17% of EIs overlapped with CGIs, with enrichment ranging from 0.6 to 2.1. The randomization test suggested that EIs significantly overlapped with CGI in differentiated cells, but not in ES and iPSC cells (Table 1).

To probe the chromatin features of EIs, we compared locations of EIs in H1, HAEC and HPF with public datasets of comparable cell lines [10,23]. Interestingly, EIs highly coincide with regions interacting with CCCTC-binding factor (CTCF), the major organizer of higher-order chromatin in mammalian genomes (Figure 2). Overall, up to 61.3% of EIs overlap with CTCF binding regions, which are 8.2-fold enriched compared with the random pattern ($P < 10^{-3}$, Table 2). Furthermore, up to 49% of EIs overlap with DNase hypersensitive sites (DHSs), the hallmark of open chromatin, which is 6.5-fold enrichment compared with the random ($P < 10^{-3}$, Figure 2 and Table 2). We further explored the overlaps of EIs with other histone modifications, and found that EIs highly overlaps with H3K4me3 (Enrichment up to 5.3) and H3K9ac (Enrichment up to 3.3), but less enrich for H3K27me3 (Enrichments from 1.7 to 2.2) and H3K36me3 (Enrichments from 0.7 to 2.1). The enrichments are similar among the three cell types. In addition, we investigated the enrichment by comparing EIs with random pattern within LOCK regions, and got similar results and even stronger enrichments for CTCF (up to 13.9 fold, Table 2).

We then asked whether there is any association between EIs and DMRs. For this purpose, we compared genomic locations of EIs with DMRs identified by CHARM array [17,18]. We found that EIs are highly enriched for DMRs distinguishing tissue types (T-DMRs). For example, EIs near TSSs of nitric oxide synthase 1 (*NOS1*), xylosyltransferase I (*XYLT1*) and heparan sulfate (glucosamine) 3-O-sulfotransferase 1 (*HS3ST1*) all overlap T-DMRs (Figure 3). Overall, a large fraction of EIs (39-62% across the five cell types) overlap with T-DMRs, with enrichment from 2.1 to 2.9 folds relative to random patterns (Table 3; $p < 10^{-3}$).

We further tested the relationship between EIs and DMRs associated with reprogramming (R-DMRs). Similar to T-DMRs, R-DMRs were more methylated in iPSC cells compared to fibroblasts (Hyper R-DMRs) were also

significantly enriched in EIs of all the cell types, whose enrichments ranging from 2.2 to 3.2 fold. However, R-DMRs less methylated in iPSC cells (Hypo R-DMRs) were highly enriched in EIs of differentiated cells (5.1 to 6.2 folds enrichment, P values all $< 10^{-3}$), but much less enriched in PSCs (1.2 to 3.1 folds of enrichment). Importantly, EIs in HPF, the same cell type of parental cells in reprogramming, are strongly enriched for hypomethylated R-DMRs (enrichment = 6.1, $P < 10^{-3}$), whereas those in iPSCs did not significantly overlap with hypomethylated R-DMRs (enrichment = 1.2, $P = 0.33$), indicating a coordinated hypomethylation in these EIs during reprogramming.

We then compared EI locations with colon cancer-associated DMRs (C-DMRs) and observed an opposite trend to that of R-DMRs. EIs in 4 out of 5 cell lines were significantly enriched for C-DMRs more methylated in colon cancers (hypermethylated C-DMRs), and the enrichment ranged from 3.8 to 5.4 fold (Table 3). In contrast, all five cells types were not significantly enriched for C-DMRs less methylated in cancers (hypomethylated C-DMRs). These results were further confirmed by comparing EIs with an independent list of C-DMRs discovered by whole genome bisulfite sequencing [20]. EIs of all five cell lines significantly overlapped hypermethylated C-DMRs (enrichment from 4.1 to 7.6 fold, P values all $< 10^{-3}$), whereas none of them were significantly enriched for hypomethylated C-DMRs (Table 3). Interestingly, almost all EIs (98%) that associated with hypermethylated C-DMRs also overlap CGIs. These data suggest that EIs in normal cells may become hypermethylated in cancers.

To explore the biological role of EIs, we compared expression levels of genes associated with EIs, of genes with LOCKs but not EIs, and of genes not overlapping LOCKs (Figure 4A). It is clear that expressions of genes overlapping EIs are significantly higher than those of within LOCKs but not EIs (t -test, $p < 2 \times 10^{-16}$). To further test whether EI associated genes are regulated by other histone marks, we investigated the relationship between H3K36me3/H3K27me3 and genes with EIs, with LOCKs and without LOCKs (Figure 4B). In either category (with or without K36me3/K27me3), genes at LOCK regions always have the lowest expression and genes at non-LOCK regions have the highest. However, genes with EIs have expressions in the middle, and positively (negatively) associated with H3K36me3 (H3K27me3),

Table 1 Overlap of EIs with CpG islands (CGIs) and transcription start sites (TSSs)

Cell line	Number of EIs	Average Size of EIs (bp)	% of EIs overlap CGIs (En. ^b , P ^c)	% of EIs overlap TSSs (En. ^b , P ^c)	# of genes associated with EIs ^a
H1	1,060	4880	9.7 (1.3, 0.06)	7.9 (5.0, <10 ⁻³)	119
ADA-38	758	4401	4.7 (0.6, 0.99)	4.6 (2.7, <10 ⁻³)	60
HA	2,254	5029	14.2 (1.8, <10 ⁻³)	12.4 (7.8, <10 ⁻³)	338
HAEC	2,359	5867	13.6 (1.6, <10 ⁻³)	11.3 (5.4, <10 ⁻³)	373
HPF	2,465	5477	17 (2.1, <10 ⁻³)	12.7 (6.5, <10 ⁻³)	409

^a 1 kb up and downstream of TSS overlapping EIs; ^b Enrichment compared to random patterns; ^c p values calculated by 1,000 permutations.

indicating that EI related genes could be regulated by these two marks.

Then we conducted Gene Ontology (GO) analysis with genes whose TSSs are associated with EIs. EI-associated genes in differentiated cells were strongly associated with 1) biological processes such as system development, cell adhesion and cell differentiation; 2) cellular compartments of plasma membrane and synapse, and 3) molecular function of ion binding and channel activity (Table 4).

Finally, to test whether EIs are associated with specific cellular functions, we compared the location of EIs among the three differentiated cell lines. Based on our current strategy to define EIs, ~50% of them are cell type specific (Figure 5A). It should be noted that the detection of EI is based on the definition of LOCKs as well

as the amount of reduction of H3K9me2 levels within LOCK bodies. Some tissue specific EIs may be due to differential LOCKs or different amount of H3K9me2 reductions among cell types. Due to these reasons the number of tissue specific EIs is likely an over-estimate. New technology with higher resolution and dynamic range, such as CHIP-seq, will help achieve better accuracy and specificity in tissue comparisons. Nevertheless, we found that some tissue specific EIs are biologically meaningful. For example, an EI is located near the TSS of Down syndrome cell adhesion molecule gene (*DSCAM*) in astrocytes (HA) but not the other two cell types (Figure 5B). It was shown that *Dscam* diversity is essential for neuronal circuit assembly [24], and genetic variations of this gene were associated with Down syndrome and congenital heart disease (DSCHD) [25] and

Table 2 Overlaps (%) of EIs with chromatin marks^a

	EIs overlap with	Observed	Random within LOCKs	Fold enriched ^c	P value ^b	Random at WG	Fold enriched ^c	P value ^b
HAEC	CTCF	61.3	4.4	13.9	<10 ⁻³	7.5	8.2	<10 ⁻³
	DHSs	48.9	10.1	4.8	<10 ⁻³	7.5	6.5	<10 ⁻³
	H3K4me3	23.3	6.8	3.4	<10 ⁻³	5.6	4.2	<10 ⁻³
	H3K27me3	46.0	34.1	1.4	<10 ⁻³	27.8	1.7	<10 ⁻³
	H3K36me3	20.0	14.1	1.4	<10 ⁻³	9.7	2.1	<10 ⁻³
	H3K9ac	17.7	6.0	2.9	<10 ⁻³	5.3	3.3	<10 ⁻³
H1	CTCF	46.9	9.0	5.2	<10 ⁻³	11.8	4.0	<10 ⁻³
	DHSs	36.0	7.1	5.1	<10 ⁻³	18.1	2.0	<10 ⁻³
	H3K4me3	13.2	1.8	7.4	<10 ⁻³	1.4	9.8	<10 ⁻³
	H3K27me3	21.6	10.8	2.0	<10 ⁻³	9.8	2.2	<10 ⁻³
	H3K36me3	11.0	16.1	0.7	1	15.2	0.7	1
	H3K9ac	28.9	11.8	2.4	<10 ⁻³	11.5	2.5	<10 ⁻³
HPF	CTCF	56.9	7.9	7.2	<10 ⁻³	9.6	5.9	<10 ⁻³
	DHSs	71.0	27.7	2.6	<10 ⁻³	19.2	3.7	<10 ⁻³
	H3K4me3	30.8	8.1	3.8	<10 ⁻³	13.0	2.4	<10 ⁻³
	H3K27me3	53.1	34.2	1.6	<10 ⁻³	29.0	1.8	<10 ⁻³
	H3K36me3	22.6	16.5	1.4	<10 ⁻³	24.9	0.9	0.85
	H3K9ac	13.2	2.5	5.2	<10 ⁻³	6.2	2.1	<10 ⁻³

^a EIs of HAEC, H1 and HPF were compared with chromatin marks of HUVEC, H1 and normal lung fibroblasts, respectively (ref.10).

^b P values were calculated by 1000 permutations.

^c Enrichment is calculated as the ratio of observed to random.

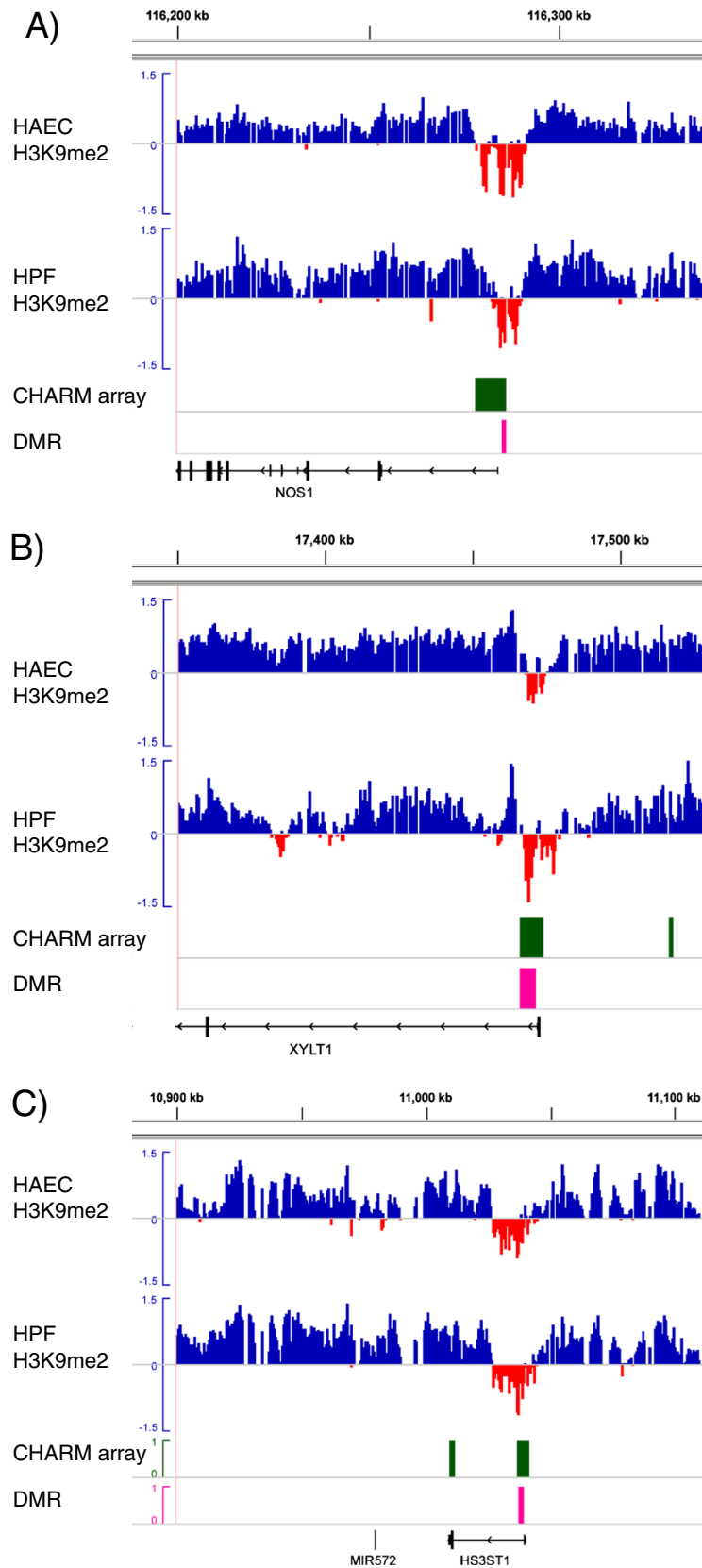


Figure 3 (See legend on next page.)

(See figure on previous page.)

Figure 3 EIs are enriched for differential methylation regions (DMRs). The H3K9me2 signals of HAEC and HPF are compared with regions of T-DMRs (pink bars). Regions of CHARM array are denoted by green bars. EIs (red dips) clearly overlap T-DMRs near the TSSs of *NOS1* (A), *XYLT1* (B) and *HS3HT1* (C).

bipolar disorder [26]. Furthermore, an EI is found on the 5' end of myocardin (*MYOCD*) gene in HA and HPF but HAEC (Figure 5C). Myocardin is a coactivator of serum response factor which specifically expressed in cardiac and smooth muscle cells [27], and promoter variation of this gene was proposed as a biomarker of cardiac hypertrophy [28]. These data suggest that EIs may be important in regulating specific cellular functions.

Discussion

In summary, by examining the genome-wide distribution of H3K9me2 in human PSCs and differentiated cells, we found a novel microstructure within heterochromatin domains of thousands of small euchromatin islands (EIs) located within large H3K9me2 blocks (LOCKS). EIs are strongly associated with open chromatin regions (DHSs), active chromatin marks (H3K4me3 and H3K9ac) and higher-order chromatin organizers (CTCF). Furthermore, EIs are highly enriched for DMRs associated with tissue specificity (T-DMRs), reprogramming (R-DMRs) and cancers (C-DMRs). This association is particularly strong for hypomethylated R-DMRs and hypermethylated C-DMRs. Genes associated with EIs are enriched for annotations of system development, cellular differentiation and cell adhesion. These results suggested that EIs may coordinate higher order chromatin and mediate co-regulation of DNA methylation in reprogramming and tumorigenesis. However, further experimental work is needed to address the functional relevance of EIs and their strong association with CTCF and DMRs.

In this study, we compared H3K9me2 profiles with publicly available epigenomic data generated from similar cell types. This strategy may lead to biased estimation of the enrichments of EIs with other epigenetic marks,

because patterns of EIs may be different between the two samples. Comparison of exactly matched cell lines and cultures could assess the association between them more accurately.

Note that a previous literature used the term "euchromatic islands" in a completely different context, simply to describe chromatin regions with H3K4me3 and CpG islands, essentially describing promoter regions of active genes [4,29]. As that term was rarely used previously, and to convey a completely different meaning, we do not think there will be confusion with our newly defined (and differently spelled) "euchromatin islands" or EIs, namely H3K9me2 depleted regions/islands within an ocean of heterochromatin (LOCKS), enriched for regulatory elements such as enhancers (DHSs) and insulators (CTCF). Thus, EIs are novel units of the genomic "toolbox" which may be important in epigenetic regulation as suggested by their strong association with DMRs (Table 3).

Higher-order organization of the genome remains a highly active area to be explored. Recent evidence indicates the presence of spatial compartments of active and repressive chromatin domains as general principles of genome organization in mammalian cells [30,31], and CTCF mediates intra- and inter-chromosomal interactions by tethering chromatin regions binding CTCF [32]. It would be interesting to explore the possibility that euchromatin islands act as "anchors" for the interactions among heterochromatin domains or between heterochromatic and euchromatic regions. Moreover, the relationships between EIs and heterochromatin formation, and the biophysical features of EIs are interesting questions for future investigation.

Although evidence provided in this and other studies have indicated that large heterochromatin domains

Table 3 Percentage of EIs overlap with DMRs

Cell line	Overlaps with DMRs (%)																				
	T-DMR			R-DMR						C-DMR						Whole genome C-DMR					
				Hyper			Hypo			Hyper			Hypo			Hyper			Hypo		
	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p
H1	53.4	2.8	<10 ⁻³	19.4	3.1	<10 ⁻³	5.8	3.1	0.02	11.7	3.8	<10 ⁻³	4.9	1.4	0.24	5.6	5.2	<10 ⁻³	0.9	1	0.31
ADA-38	39.1	2.1	<10 ⁻³	17.4	3.2	<10 ⁻³	2.2	1.2	0.33	6.5	2.2	0.05	6.5	2	0.08	4.1	4.1	<10 ⁻³	0.4	0.5	0.84
HA	60.4	2.9	<10 ⁻³	19.3	3.1	<10 ⁻³	11	5.5	<10 ⁻³	16.7	5.4	<10 ⁻³	4.5	1.3	0.25	8.6	7.3	<10 ⁻³	1	1.1	0.34
HPF	59.5	2.8	<10 ⁻³	17.5	2.8	<10 ⁻³	12.8	6.1	<10 ⁻³	17.5	5.1	<10 ⁻³	5.7	1.5	0.09	9.4	7.6	<10 ⁻³	1.3	1.4	0.04
HAEC	62.3	2.9	<10 ⁻³	14.3	2.2	<10 ⁻³	11.7	5.6	<10 ⁻³	15.7	4.5	<10 ⁻³	4.6	1.2	0.29	7	5.3	<10 ⁻³	1.1	1.1	0.27

Notes: En. = Enrichment. Statistical significances were tested by 1000 permutations.

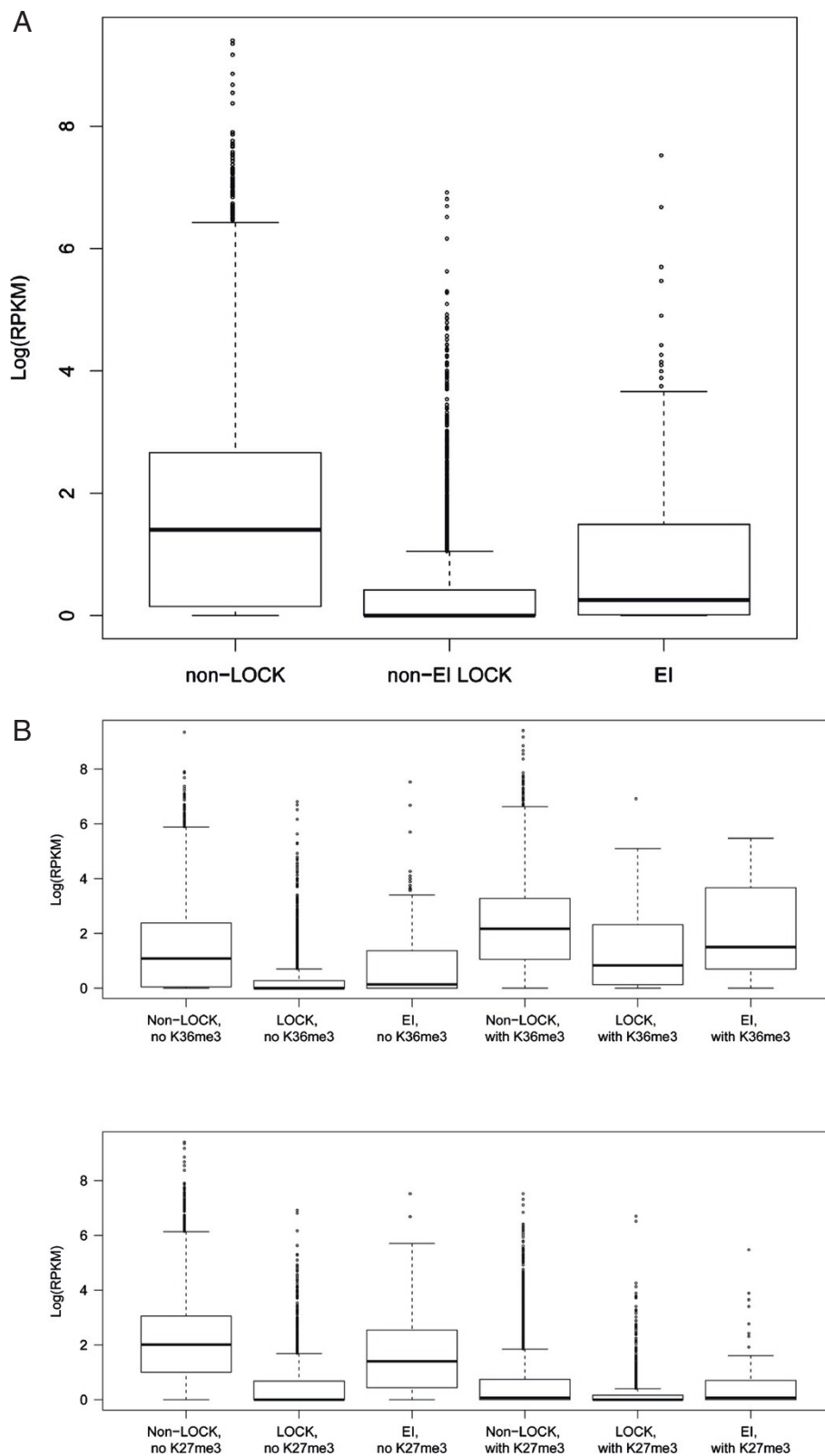


Figure 4 Expression of genes associated with EIs. We compared expression levels for genes with TSS at different regions. Expression values are RPKM (read per kb per million reads) for lung fibroblast IMR90 [7]. **A**) boxplot of expression level of genes with TSS 1) overlapping EIs; 2) overlapping LOCKs but not EIs; and 3) not overlapping EIs. **B**) Relationship between H3K36me3/H3K27me3 and expression of EI associated genes.

Table 4 Top 10 GO terms of genes associated with EIs

GO terms	% of genes	Fold en.	FDR
Biological processes			
nervous system development	12.6	2.2	9E-11
developmental process	25.5	1.6	4E-09
system development	20.5	1.7	4E-09
anatomical structure development	21.5	1.6	1E-08
multicellular organismal development	23.4	1.6	2E-08
cell adhesion	8.7	2.4	8E-08
biological adhesion	8.7	2.4	8E-08
cell differentiation	15.2	1.8	4E-07
cell development	8.0	2.4	4E-07
multicellular organismal process	31.1	1.4	4E-07
Cellular compartment			
plasma membrane part	21.1	1.9	4E-14
plasma membrane	30.3	1.6	1E-12
membrane part	44.2	1.3	1E-09
integral to plasma membrane	12.6	2.1	3E-09
intrinsic to plasma membrane	12.7	2.1	5E-09
membrane	47.0	1.3	2E-08
intrinsic to membrane	37.3	1.4	2E-07
integral to membrane	35.9	1.4	1E-06
synapse	5.3	3.0	2E-06
extracellular region	16.7	1.7	6E-06
Molecular function			
calcium ion binding	9.7	2.1	4E-06
gated channel activity	4.5	2.9	9E-05
molecular transducer activity	17.5	1.5	6E-04
signal transducer activity	17.5	1.5	6E-04
substrate specific channel activity	5.1	2.5	6E-04
ion channel activity	4.9	2.5	8E-04
channel activity	5.1	2.4	1E-03
passive transmembrane transporter activity	5.1	2.4	1E-03
cation channel activity	3.9	2.8	2E-03
ligand-gated channel activity	2.3	3.6	1E-02

are highly dynamic in stem cell differentiation and tumorigenesis [11-15], Lienert et al. indicated that genome coverage of H3K9me2 domains do not increase globally during neuronal differentiation of mouse ES cells [33]. First of all, lineage specificity of differentiated cells may explain the conflicts. As reported in our earlier work [11] the amount of LOCKs detected from brain and ES cells are comparable (9.8% vs. 4%), whereas the amount is very high in liver (45.6%). The Lienert study used in vitro differentiated neurons as differentiated cells which is more similar to brain. Furthermore, the inconsistency may be due to sensitivities of

different statistical methods for finding large domains, heterogeneity of stem cells, and so on. Notably, extensive deduction of LOCKs during EMT suggested that quantitative differences of these large domains may be functionally important [13]. Nevertheless, further studies on homogenous stem cell populations may be helpful to address these debates. Whatever it holds, functionally investigations of these large domains should provide important insight toward how higher-order chromatin affects normal development and disease.

Conclusions

In conclusion, we have explored the microstructure of LOCKs and identified thousands of euchromatin islands (EIs), which may be served as a finer layer of epigenomic architecture within large heterochromatin domains. The strong association of EIs with CTCF sites, DNase hypersensitivity sites, and DMRs suggests that EIs play an important role in normal epigenomic architecture and its disruption in disease.

Methods

Cell culture

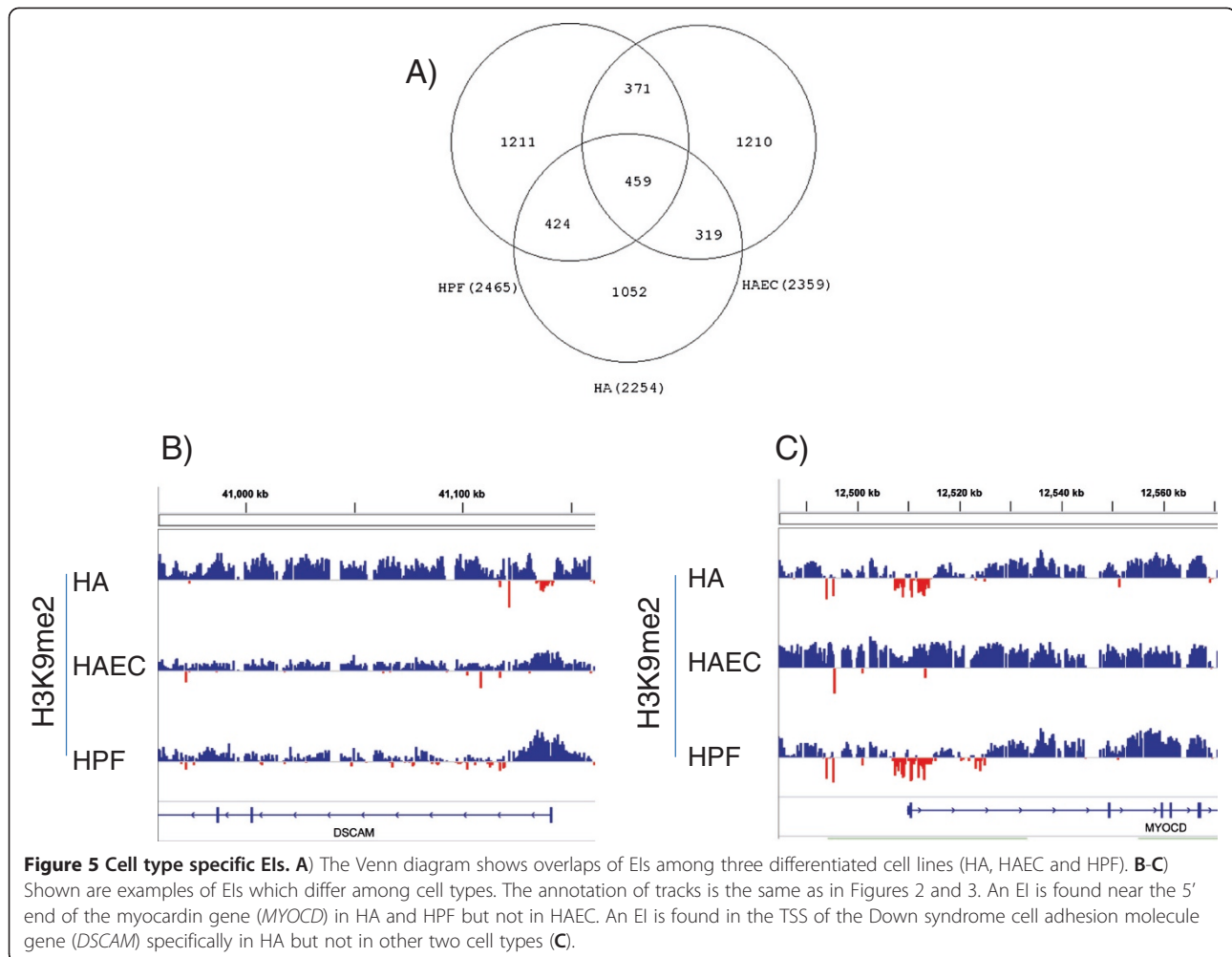
Human H1 ESCs and ADA-38 iPSCs were cultured as described [34]. Primary Human Pulmonary Fibroblasts (HPF), Human Aortic Endothelial Cells (HAEC) and Human Astrocytes (HA) were purchased from ScienCell Research Laboratories (San Diego, CA), and cultured as recommended by ScienCell.

ChIP-chip

ChIP-chip experiments were performed as described [11], using a commercial monoclonal antibody (Abcam, ab1220), which specifically recognizes H3K9me2 but not other modifications [35]. The passage numbers for cells used for ChIP analysis were P46 for H1, P59 for ADA-38 and P2 for primary cells from ScienCell. We first mapped whole genome distribution of H3K9me2 using "Mouse ChIP-chip 2.1M Economy Whole-Genome Tiling arrays (4 arrays per set) from NimbleGen", with 203 bp of median probe spacing. Then we repeated the microarray experiments on one of the Mouse ChIP-chip 2.1 M Whole-Genome Tiling sets, whose median probe spacing is 100 bp. The replicate array covers 10% of the genome, including part of chromosome 6 (111,920,005-170,893,515), whole chromosome 7 and part of chromosome 8 (521-74,730,105). For the replicate experiments, cell cultures, ChIP sample preparation, labeling and hybridization were performed independently.

ChIP-chip data analysis

Data were first normalized by partial quantile normalization, then LOCKs were detected based on the smoothing values of normalized log2 ratios of data



between ChIP and input channels [11]. The euchromatin islands (EIs) are defined as short regions within LOCK body that have low H3K9me2 methylation levels. To detect such regions we designed the following smoothing based approach. The log₂ ratios for probes within LOCKs were first smoothed using 5000 bp window. The relatively short smoothing window is used to capture the signal variations in small regions. Genomic regions with smoothed value less than 1% of all the smoothed values were defined as EIs. It is required that EIs are at least 1000 bps long and contain at least 10 probes. EIs less than 1000 bps apart will be merged into one. It is also required that the EIs are at least 20000 bps away from the LOCK boundaries. This is because the log₂ ratios are smaller at LOCK boundaries. Such requirement prevents mistakenly taking LOCK boundaries as EIs. A flow diagram showing the algorithm for detecting EIs was provided in Additional file 8: Figure S6. Microarray data have been submitted to GEO database (accession numbers: GSE37335).

To compute the enrichment of EI overlapping other genomic features (CTCF, DHS, etc.), we first calculated the percent of EIs overlapping the feature. Then a set of genomic regions was randomly sampled. The number and lengths of the random regions match the EI list. The random regions were then compared with the feature to obtain a percentage of overlapping. Such process was repeated 1000 times. The percentages obtained from the process form the null distribution for percentage of overlapping. The p-values and enrichments were computed based on the null distribution. The p-values were then corrected for multiple testing using Bonferroni correction. Publicly available datasets used for analysis were listed in Additional file 9: Table S3.

Quantitative PCR (qPCR)

Experiments of qPCR were conducted as described [11]. Primer sequences are provided in Additional file 10: Table S4.

GO analysis

GO analysis was performed using DAVID tools as described [36], using the list of genes overlapping EIs of the three differentiated cell lines (HA, HAEC and HPF).

Additional files

Additional file 1: Table S1. Description: Genome coverage and average size of LOCKs in human PSCs and differentiated cells.

Additional file 2: Figure S1(A-D). Description: qPCR validation of H3K9me2 ChIP-chip data on 23 loci. Upper panels show log₂ (ChIP/Input) ratios of microarrays and green bars denote regions selected for qPCR validation; lower panels present qPCR enrichments of ChIP over input in the selected regions.

Additional file 3: Figure S2. Description: LOCKs overlap partial methylation domains (PMDs). (A) One representative region (on chromosome 17) where LOCKs and PMDs overlap, green and orange bars show locations of LOCK (green) and PMD (orange), and hypomethylation blocks (purple), respectively; (B) H3K9me2 density in and out of PMDs. X-axis is the probe log₂ ratios between ChIP and control samples. Y-axis is the the probability density.

Additional file 4: Table S2. Description: Coordinates of EIs (HG18).

Additional file 5: Figure S3. Description: Average H3K9me2 densities in EIs and their adjacent regions.

Additional file 6: Figure S4. Description: H3K9me2 ChIP-chip experiments in whole genome (WG) and replicate (rep) arrays.

Additional file 7: Figure S5. Description: Nucleosome density in EIs and adjacent regions. We compared common EIs of HA, HAEC and HPF with nucleosome maps of GM12878 (Supplementary Table S3), to overcome potential lineage specificity among those cell types.

Additional file 8: Figure S6. Description: Flow diagram of EI detection.

Additional file 9: Table S3. Description: Public datasets used for analysis.

Additional file 10: Table S4. Description: qPCR primer sequences.

Abbreviations

EI: Euchromatin island; DHS: DNase hypersensitive site; DMR: Differential methylation region; H3K9me2: H3 lysine 9 dimethylation; LOCK: Large organized chromatin k9-modification; ESC: Embryonic stem cell; PMD: Partial methylated domains; PSC: Pluripotent stem cell; HA: Human astrocytes; HAEC: Human aortic endothelial cell; HPF: Human pulmonary fibroblast; TSS: Transcription start site; CTCF: CCCTC-binding factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BW and APF conceived the project and designed the study. BW performed cell culture of primary cells, ChIP and qPCR; HW performed data analysis; YL and QGD generated PSC lines; BE conducted microarray analysis; BW, HW and APF prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Institute of Health (NIH) grant R37 CA54358.

Author details

¹Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ²Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ³Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Howard Hughes Medical Institute, Boston,

MA, USA. ⁴Current address: Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, China.

Received: 28 September 2012 Accepted: 19 October 2012

Published: 26 October 2012

References

1. Feinberg AP: Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007, **447**(7143):433–440.
2. Berger SL: The complex language of chromatin regulation during transcription. *Nature* 2007, **447**(7143):407–412.
3. Kouzarides T: Chromatin modifications and their function. *Cell* 2007, **128**(4):693–705.
4. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: A high-resolution map of active promoters in the human genome. *Nature* 2005, **436**(7052):876–880.
5. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: High-resolution profiling of histone methylations in the human genome. *Cell* 2007, **129**(4):823–837.
6. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al: Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007, **39**(3):311–318.
7. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al: Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007, **448**(7153):553–560.
8. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al: Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 2008, **40**(7):897–903.
9. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al: Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009, **459**(7243):108–112.
10. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al: Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011, **473**(7345):43–49.
11. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP: Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* 2009, **41**(2):246–250.
12. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al: Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 2008, **453**(7197):948–951.
13. McDonald OG, Wu H, Timp W, Doi A, Feinberg AP: Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat Struct Mol Biol* 2011, **18**(8):867–874.
14. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al: Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 2010, **6**(5):479–491.
15. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, et al: Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* 2012, **22**(2):246–258.
16. Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C, et al: An epigenetic signature for monoallelic olfactory receptor expression. *Cell* 2011, **145**(4):555–570.
17. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al: The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009, **41**(2):178–186.
18. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, et al: Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 2009, **41**(12):1350–1353.
19. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al: Human DNA methylomes at base

- resolution show widespread epigenomic differences. *Nature* 2009, **462**(7271):315–322.
20. Hansen KD, Timp W, Bravo HC, Sabuncyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, et al: **Increased methylation variation in epigenetic domains across cancer types.** *Nat Genet* 2011, **43**(8):768–775.
 21. Paz MF, Fraga MF, Avila S, Guo M, Pollan M, Herman JG, Esteller M: **A systematic profile of DNA methylation in human cancer cell lines.** *Cancer Res* 2003, **63**(5):1114–1121.
 22. Carmona FJ, Villanueva A, Vidal A, Munoz C, Puertas S, Penin RM, Goma M, Lujambio A, Piulats JM, Mesia R, et al: **Epigenetic Disruption of Cadherin-11 in Human Cancer Metastasis.** *J Pathol* 2012.
 23. ENCODE: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**(4):e1001046.
 24. Hattori D, Demir E, Kim HW, Viragh E, Zipursky SL, Dickson BJ: **Dscam diversity is essential for neuronal wiring and self-recognition.** *Nature* 2007, **449**(7159):223–227.
 25. Barlow GM, Chen XN, Shi ZY, Lyons GE, Kurnit DM, Celle L, Spinner NB, Zackai E, Pettenati MJ, Van Riper AJ, et al: **Down syndrome congenital heart disease: a narrowed region and a candidate gene.** *Genet Med* 2001, **3**(2):91–101.
 26. Amano K, Yamada K, Iwayama Y, Detera-Wadleigh SD, Hattori E, Toyota T, Tokunaga K, Yoshikawa T, Yamakawa K: **Association study between the Down syndrome cell adhesion molecule (DSCAM) gene and bipolar disorder.** *Psychiatr Genet* 2008, **18**(1):1–10.
 27. Wang D, Chang PS, Wang Z, Sutherland L, Richardson JA, Small E, Krieg PA, Olson EN: **Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor.** *Cell* 2001, **105**(7):851–862.
 28. Kontaraki JE, Parthenakis FI, Patrianakos AP, Karalis IK, Vardas PE: **Myocardin gene regulatory variants as surrogate markers of cardiac hypertrophy - study in a genetically homogeneous population.** *Clin Genet* 2008, **73**(1):71–78.
 29. Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, Zardo G, Nervi C, Bernard L, Amati B: **Myc-binding-site recognition in the human genome is determined by chromatin context.** *Nat Cell Biol* 2006, **8**(7):764–770.
 30. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**(5950):289–293.
 31. Kalthor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.** *Nat Biotechnol* 2012, **30**(1):90–98.
 32. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al: **CTCF-mediated functional chromatin interactome in pluripotent cells.** *Nat Genet* 2011, **43**(7):630–638.
 33. Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schubeler D: **Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells.** *PLoS Genet* 2011, **7**(6):e1002090.
 34. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451**(7175):141–146.
 35. Egelhofer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, Alekseyenko AA, Cheung MS, Day DS, Gadel S, Gorchakov AA, et al: **An assessment of histone-modification antibody quality.** *Nat Struct Mol Biol* 2011, **18**(1):91–93.
 36. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.

doi:10.1186/1471-2164-13-566

Cite this article as: Wen et al.: Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions. *BMC Genomics* 2012 **13**:566.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

