SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
SYSTEMS BIOLOGY
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS
EPIGENOMICS
DATA INTEGRATION

Received
8 April 2013

Accepted
16 August 2013

Published
3 September 2013

Correspondence and requests for materials should be addressed to Q.W. (kigo@hit.edu.cn) or Y.Z. (yanyou1225@gmail.com)

* These authors contributed equally to this work.

Quantitative epigenetic co-variation in CpG islands and co-regulation of developmental genes

Hongbo Liu[1]*, Yanjun Chen[2]*, Jie Lv[1]*, Hui Liu[3], Rangfei Zhu[3], Jianzhong Su[3], Xiaojuan Liu[4], Yan Zhang[3] & Qiong Wu[1]

[1]School of Life Science and Technology, State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150001, China, [2]Deptartment of Cardiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin 150001, China, [3]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China, [4]Department of Rehabilitation, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, China.

The genome-wide variation of multiple epigenetic modifications in CpG islands (CGIs) and the interactions between them are of great interest. Here, we optimized an entropy-based strategy to quantify variation of epigenetic modifications and explored their interaction across mouse embryonic stem cells, neural precursor cells and brain. Our results showed that four epigenetic modifications (DNA methylation, H3K4me2, H3K4me3 and H3K27me3) of CGIs in the mouse genome undergo combinatorial variation during neuron differentiation. DNA methylation variation was positively correlated with H3K27me3 variation, and negatively correlated with H3K4me2/3 variation. We identified 5,194 CGIs differentially modified by epigenetic modifications (DEM-CGIs). Among them, the differentially DNA methylated CGIs overlapped significantly with the CGIs differentially modified by H3K27me3. Moreover, DEM-CGIs may contribute to co-regulation of related developmental genes including core transcription factors. Our entropy-based strategy provides an effective way of investigating dynamic cross-talk among epigenetic modifications in various biological processes at the macro scale.

DNA methylation and various types of histone modifications are widely studied epigenetic modifications that play important roles in regulation of cell development and differentiation[1]. The fulfillment of these functions depends on designated genome regions. CpG islands (CGIs) are specific regions in mammalian genomes with a high frequency of CpG dinucleotides and GC content[2]. CGIs are interspersed in different genome locations including the gene promoter, gene body, and intergenic regions. Approximately 70% of mammalian genes have CGIs in their promoter regions[3].

Mounting evidence has indicated that promoter CGIs are important epigenetic regulatory elements[4]. Hypomethylation is a noticeable feature of CGIs in mammal genomes and large number of experiments have confirmed that the hypermethylation of promoter CGIs is involved in inhibition of gene expression[2]. Promoter CGIs undergo dynamic methylation changes during cell development and differentiation[5]. In addition, recent studies also revealed new roles for CGIs in chromatin reconstitution. Vavouri et al.[6] found that human genes with CGI promoters had a distinct transcription-associated chromatin organization. Hypomethylated promoter CGIs can influence chromatin remodeling by recruiting functional proteins related to histone modifications. For example, promoter CGIs can directly recruit the histone H3 lysine 36 demethylase KDM2A[7], and the CpG-binding protein Cfp1 associated with the H3K4 methyltransferase Setd1[8]. In mammalian embryonic stem cells (ESCs), promoter CGIs can recruit PRC2, which catalyzes H3 lysine 27 trimethylation (H3K27me3)[9]. A systematic analysis of the epigenetic modifications in CGIs may contribute to the understanding of epigenetic regulation of gene transcription.

Moreover, several lines of evidence suggest cross-talk among multiple epigenetic modifications in the regulation of gene expression[10–16]. A typical example is bivalent chromatin that contains both activating and repressing epigenetic modifications in the same region and plays important roles in maintaining the pluripotency of ESCs and in determining cell fate. Specifically, the bivalent chromatin of H3K4me3/H3K27me3 is characteristic of important developmental genes in ESCs[10]. The allelic bivalent chromatin enriched in both H3K4me2 and H3K27me3 in early embryonic stages is resolved upon neural commitment, which plays important roles in regulating tissue-specific imprinting at Grb10[11]. Orford et al.[12] reported an association between H3K4me2 and

H3K4me3 on a genome-wide scale, with differential distribution in the genes that were transcriptionally silent and uniquely susceptible to differentiation-induced H3K4 demethylation. Combinatorial histone modifications have also been used to model expression levels and infer mRNA stability[14]. Recently, H3K27me3 and DNA methylation were found to be mutually exclusive and antagonistic in CGIs in mouse ESCs[15]. However, the co-regulation of different kinds of epigenetic modifications, including DNA methylation and histone modifications in CGIs during cell differentiation, has not been studied systematically and quantitatively.

Promoter CGIs undergo dynamic methylation changes during cell development and differentiation[5]. Histone modifications in CGIs also change greatly during cell differentiation[17]. For example, the bivalent histone modifications are enriched in the main developmental genes in ESCs, but tend to resolve during cell diffe rentiation[10]. In a recent study, the systematic assessment of the modification variations of H3K4me3, H3K27me3 and H3K36me3 for transcription factors across various cellular differentiation states revealed cell lineage-specific functions[18]. Epigenetic variation is required for normal development, while abnormal epigenetic changes often lead to dysregulation of the developmental processes, which causes developmental abnormalities and diseases[19]. The quantification of epigenetic variation is vital for exploring the real roles of epigenetic modifications in the regulation of development processes[20]. By studying the cross-talk among distinct epigenetic modifications and investigating the co-variation of different kinds of epigenetic modifications during cell differentiation, insights into the molecular mechanisms behind cellular programming and reprogramming may be revealed.

The genome-wide CGIs differentially modified by epigenetic modifications (DEM-CGIs) create functional regions of epigenetic modifications during cell differentiation. Several computational methods, such as ChIPDiff[21] and DIME[22], have been proposed to identify differential histone modification sites from chromatin immunoprecipitation coupled with sequencing (ChIP-seq) data. However, these methods can only be applied to two ChIP-seq datasets at a time and cannot be used to detect quantitative variations across multiple samples. In a previous study, we developed an entropy-based method named QDMR for the quantification of methylation variation and identification of differentially methylated regions[23]. Differentially methylated CGIs (DNAm-DEM-CGIs) proximal to the promoters of genes involved in pluripotency and differentiation have been identified[24]. The quantitative identification of DEM-CGIs may provide a new strategy for the analysis of epigenetic variation across multiple samples.

CGIs in gene promoters have been studied substantially in most epigenetic studies and DNA methylation and histone modifications in promoter CGIs that are involved in regulation of gene expression have been widely reported. However, we have estimated (as detailed below) that the CGIs that are located in promoters of known genes account for only about 50% of all the CGIs in the mouse genome. The distinct functions of epigenetic modifications in other genome regions have recently been noted in several studies[25–27]. Medvedeva et al.[25] studied the CGIs located in different regions of the human genome, and found location preferences and potential functions of the CGIs in different regions of the genome. Cell type-specific DNA methylation at intragenic CGIs was reported to regulate differential gene expression during the early stages of lineage specification[26]. However, detecting dynamic epigenetic modifications in the non-promoter CGIs, especially the Intergenic CGIs, and understanding their functions during differentiation have been elusive.

Here we optimized our entropy-based QDMR strategy to quantify the variation of epigenetic modifications (including DNA methylation and three specific histone modification patterns) across mouse ESCs, neural precursor cells (NPCs) and adult brain, and investigated the relationship among different kinds of epigenetic variations in CGIs during the differentiation of neurons at the macro scale. The identification of DEM-CGIs and the exploration of their roles in regulating developmental genes confirmed that CGIs with dynamic epigenetic modifications have a role in neuron differentiation. Our results revealed the genome-wide quantitative co-variation of epigenetic modifications in CGIs and their co-regulation of developmental genes.

Results

Genome-wide epigenetic modification pattern in CGIs in different development stages. We obtained 15,948 mouse CGIs from the UCSC Table Browser[28] and classified each CGI into one of seven genome regions: Up2kb, 5′UTR, CodingExon, Intron, 3′UTR, Down2kb and Intergenic regions according to their positions relative to the RefSeq genes (see Materials and Methods for details). As reported previously, the CGIs were located, for the most part, in gene-related areas, especially the Up2kb, 5′UTR and CodingExon regions (Supplementary Figure S1), indicating their role as functional regulatory regions for genes[29]. We selected 8337 of the CGIs that had four epigenetic modifications (DNA methylation, H3K4me2, H3K4me3 or H3K27me3) in the three development stages, ESCs, NPCs and adult brain. On a global scale, the stacked histograms of epigenetic modifications (obtained using Circos[30]) revealed that the four different epigenetic modifications underwent dynamic changes across the three development stages (Figure 1a). However, the four epigenetic modifications showed different variation tendencies; namely, DNA methylation and H3K4me2 were higher in the NPCs, while H3K4me3 and H3K4me27 were lower in NPCs compared with in ESCs and brain (Supplementary Figure S2).

Genome-wide combinatorial variation (co-variation) of epigenetic modifications in CGIs. To study the dynamics of epigenetic modifications in CGIs during cell differentiation, we improved our entropy-based QDMR method and quantified the epigenetic variations among the different development stages for all 8,337 CGIs (see Materials and Methods for details). For each epigenetic modification, each CGI was assigned an entropy value, with lower entropy indicating greater epigenetic variation among the three development stages. We visualized the quantified variation of the four epigenetic modifications in the CGIs in the different genome regions using Circos (Figure 1b) and found that the CGIs with low methylation entropy had low H3K27me3 entropy but high H3K4me2/3 entropy in all genome regions studied.

Next, we explored the combinatorial variation of the four epigenetic modifications during development and found that H3K4me2 and H3K4me3 shared a similar unimodal entropy distribution, while DNA methylation and H3K27me3 entropy shared a similar bimodal distribution (Figure 1c). The correlation analysis between methylation entropy and the three kinds of histone modification entropy revealed that the methylation variation was significantly and positively correlated with H3K27me3 variation, but negatively correlated with H3K4me2/3 variation (Figure 1d). Further, H3K4me2 variation was positively and significantly related to H3K4me3 variation (Supplementary Figure S3). These results implied a genome-wide universal co-variation among different epigenetic modifications during differentiation.

CGIs differentially modified by epigenetic modifications (DEM-CGIs) during neural differentiation. To investigate the pattern of co-variation among different epigenetic modifications during differentiation, we identified the DEM-CGIs during neural differentiation using a threshold (0.962) that was obtained from a probability model for three samples[23] (see Materials and Methods for details). We found that more than 62% (5,194/8,337) of the DEM-CGIs were differentially modified by at least one of the four epigenetic

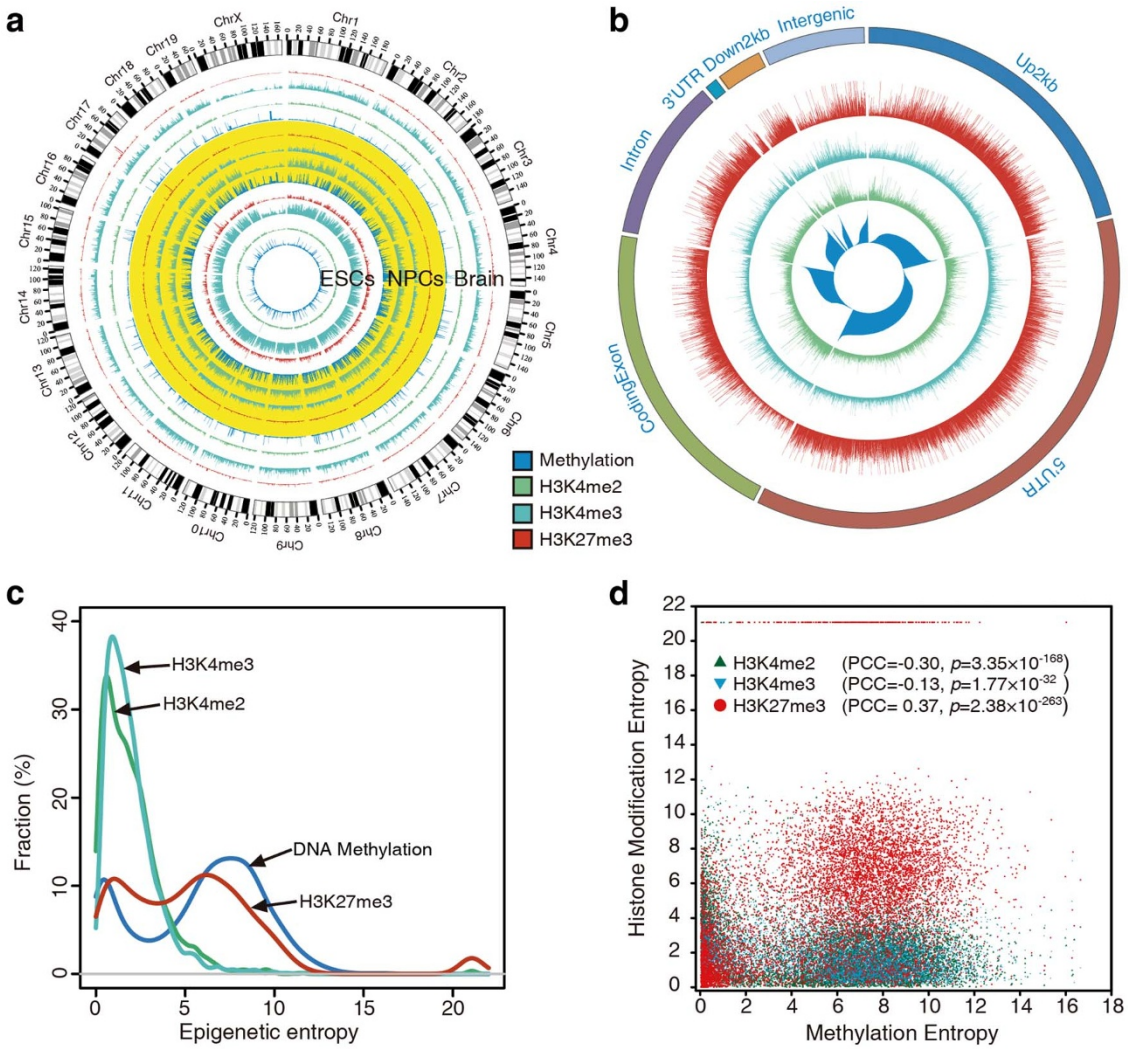

Figure 1 | Dynamic epigenetic modifications in CGIs. (a) Circos plot of the epigenetic modification profiles for CGIs in the whole genome. The tracks, from outermost to innermost, show the ideogram for the mouse karyotype (using genome build mm8), and the four epigenetic modifications in brain, NPCs and ESCs. The tracks are scaled separately to show modification fluctuations. (b) Distribution of the epigenetic entropies representing the variation of epigenetic modifications during neural differentiation, with lower entropy representing greater epigenetic variation. (c) Circos plot of the entropy of the four kinds of epigenetic modifications in the different genomic regions. The tracks, from outermost to innermost, show the genome region, H3K27me3, H3K4me3, H3K4me2 and DNA methylation. (d) Scatter diagram of DNA methylation entropy and the three kinds of histone modification entropy. PCC is the Pearson correlation coefficient between DNA methylation entropy and one kind of histone modification entropy; p is the significance of the PCC.

modifications, indicating a dramatic epigenetic variation in CGIs during mouse development (Figure 2a and Supplementary Table S1). Some DEM-CGIs were differentially modified by two or more epigenetic modifications (Supplementary Figure S4) and six DEM-CGIs were differentially modified by all four kinds of epigenetic modification, while five of them were located near the transcriptional start sites of genes (Supplementary Table S2 and Supplementary Figure S5). One of these CGIs was located in the promoter region of the Fzd9 gene. The epigenetic dynamics in this CGI may be responsible for the role of the Wnt signaling pathway in embryonic development and in abnormal development because Fzd9 encodes a receptor for Wnt in the Wnt signaling pathway[31–33] (Supplementary Figure S5).

In addition, the identified DEM-CGIs were distributed widely in the whole genome, and 92% (4,778/5,194) of them were located near 4,508 known genes. Here, we termed these genes as genes differentially modified by epigenetic modification (DEMGs) (Figure 2b and Supplementary Table S1). Some of the DEMGs were related to two or more DEM-CGIs (Supplementary Figure S4); for example, four DEM-CGIs were in or near gene Isl2, a LIM-homeodomain transcription factor that is important for terminal differentiation of motoneurons[34] (Figure 2c). We performed a functional enrichment analysis for the DEMGs related to each kind of DEM-CGI and found they were enriched in gene ontology biological process terms related to embryonic development, especially neuron differentiation (Figure 2d).

Differentially DNA methylated CGIs overlap with those differentially modified by H3K27me3. DNA methylation and H3K27me3 have recently been found to be mutually exclusive and antagonistic in CGIs in mouse ESCs[15]. This finding prompted us to investigate the relationships between DNAm-DEM-CGIs and H3K27me3-DEM-CGIs during neuron development (Figure 3a and Supplementary Table S3). We found that DNAm-DEM-CGIs were more prone to H3K27me3 changes than to H3K4me2/3 changes compared with nonDNAm-DEM-CGIs (Figure 3b, c). Further analysis revealed that DNAm-DEM-CGIs overlapped significantly with H3K27me3-DEM-CGIs, which is consistent with the

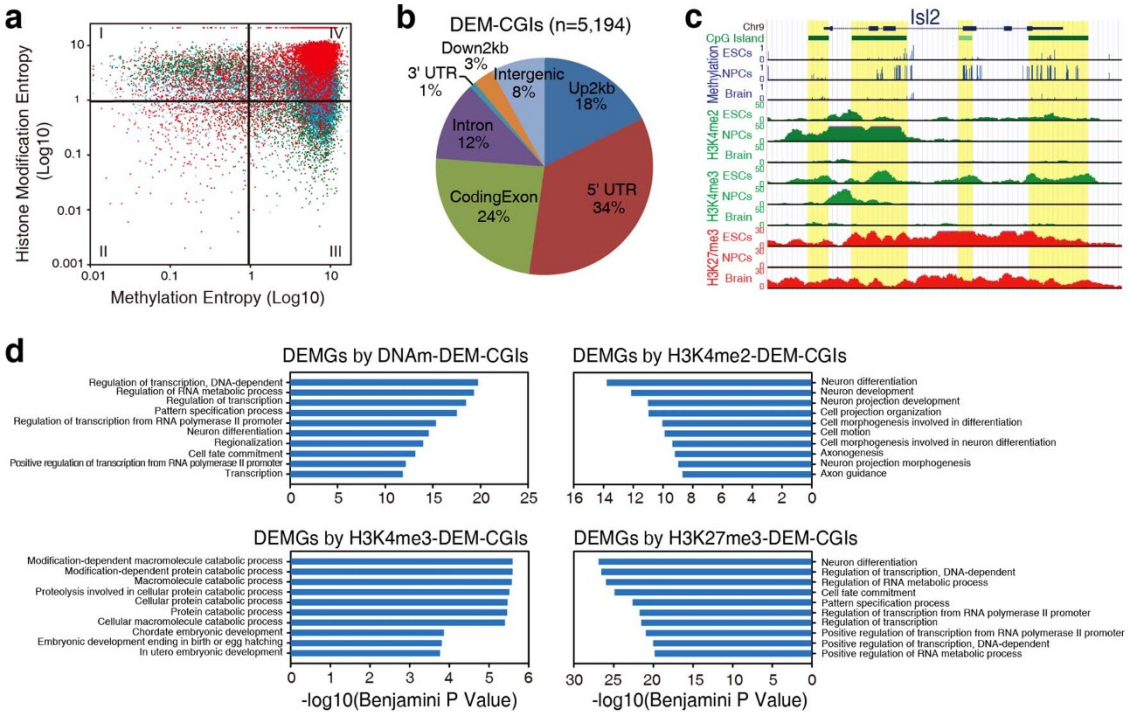

Figure 2 | CGIs differentially modified by epigenetic modifications. (a) Scatter diagram of DNA methylation entropy and the three kinds of histone modification entropy on a log-log scale. The entire space is divided into four parts (I, II, III and IV) by two black lines representing the DEM-CGI threshold (0.962). (b) Distribution of DEM-CGIs in seven genome regions. (c) UCSC Genome Browser view of epigenetic modification in four DEM-CGIs near the Isl2 gene. (d) Enrichment analysis of gene function of DEMGs. The top 10 terms based on the Benjamini p values are listed.

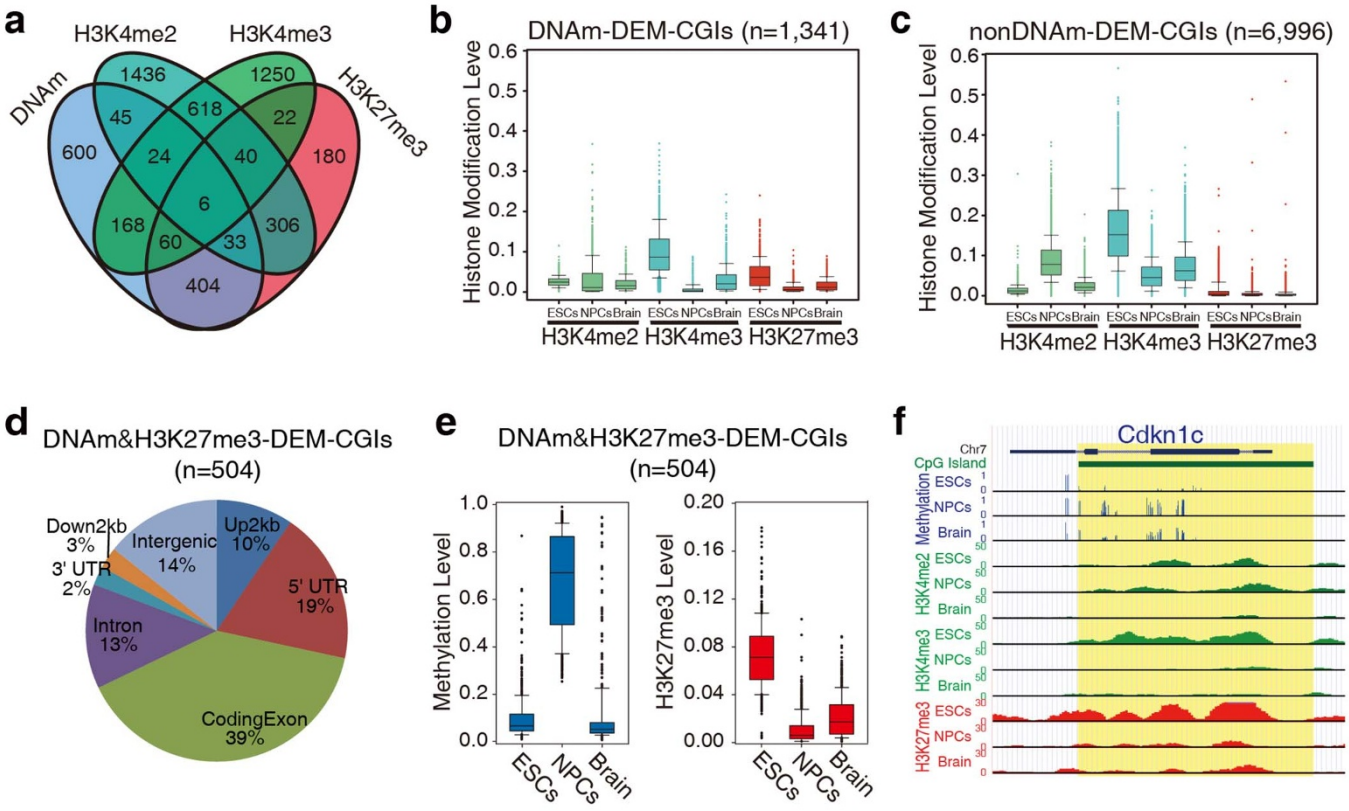

Figure 3 | CGIs differentially modified by DNA methylation and H3K27me3. (a) Venn diagram visualizing the DEM-CGI shared by double, triple, and quadruple combinations among DNAm-DEM-CGIs, H3K4me2-DEM-CGIs, H3K4me3-DEM-CGIs, and H3K27me3-DEM-CGIs. (b) Pattern of histone modifications on DNAm-DEM-CGIs. (c) Pattern of histone modifications on nonDNAm-DEM-CGIs. (d) Distribution of DNAm&H3K27me3-DEM-CGIs in seven genome regions. (e) Methylation and H3K27me3 pattern in the DNAm&H3K27me3-DEM-CGIs. (f) UCSC Browser view of epigenetic modification in a DNAm&H3K27me3-DEM-CGI near the Cdkn1c gene.

co-variation between two repressive markers (Supplementary Table S3). We identified 504 CGIs that were differentially modified by both DNA methylation and HK27me3 (DNAm&H3K27me3-DEM-CGIs); 40% (199/504) of them were located in the CodingExon region of the genome while only 21% (1,710/8,337) of the other CGIs were located in this region (Figure 3d and Supplementary Table S4). This finding indicated the potential co-variation between two repressive epigenetic modifications in the gene body in addition to the promoter region. Further analysis revealed most of the DNAm&H3K27me3-DEM-CGIs increased methylation levels in NPCs and decreased levels in brain and ESCs, while they decreased H3K27me3 in NPCs and increased levels in brain (Figure 3e), reflecting the antagonism between the two repressive markers during the dynamic development process. For example, as shown in Figure 3f, one of the DNAm&H3K27me3-DEM-CGIs overlapped with most of an imprinted gene Cdkn1c (also known as p57KIP2), whose abnormal expression induced by DNA methylation and H3K27me3 may lead to Beckwith-Wiedemann syndrome and multiple cancers[35,36].

Multiple epigenetic modifications co-regulate the developmental genes during neural differentiation. Previous studies have revealed the functions of epigenetic modifications in regulation gene expression[13,37,38]. To address the roles of epigenetic modifications in co-regulation of gene expression at the macro scale, we obtained the expression levels of 6,026 genes in all three developmental stages and

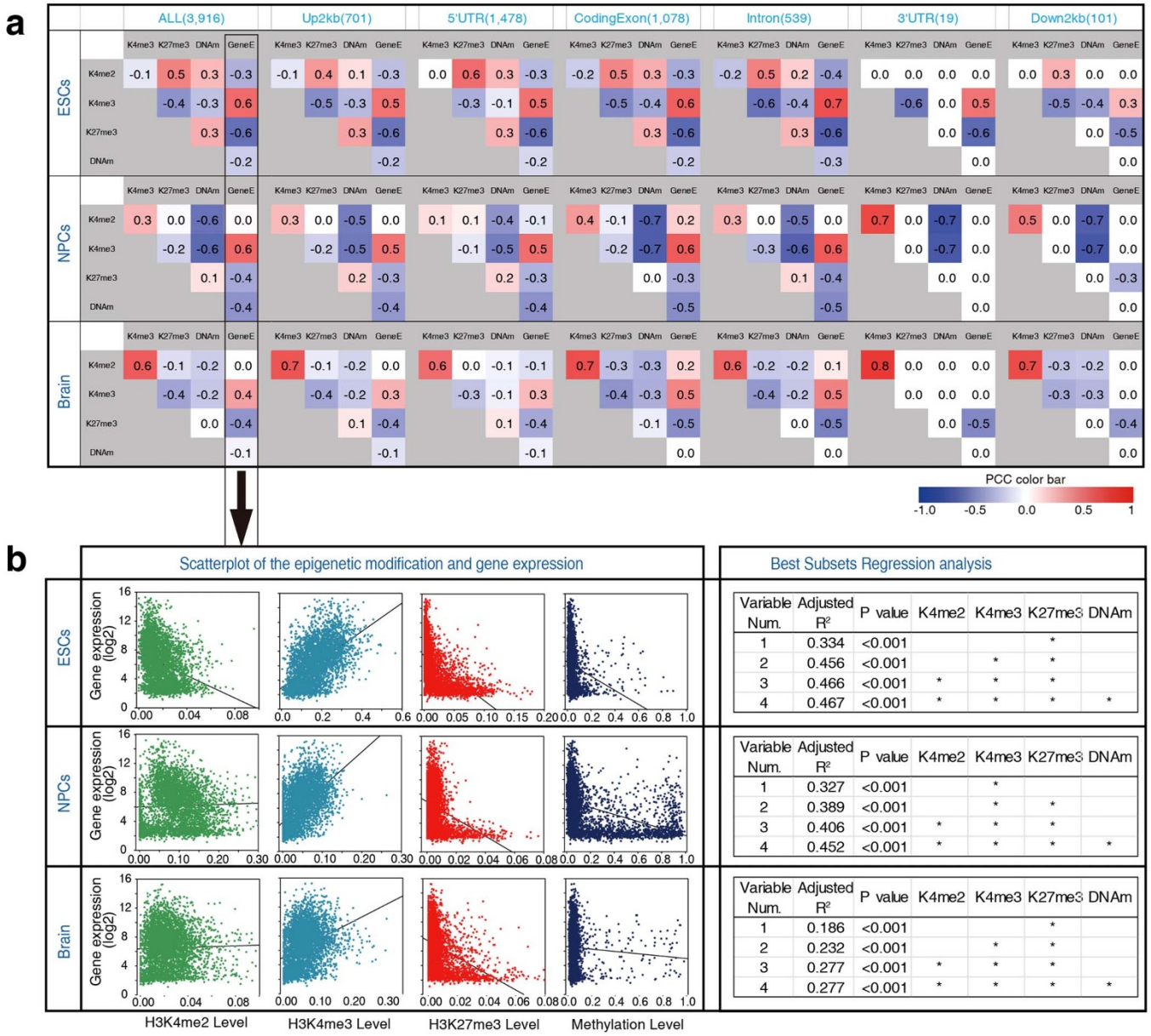

Figure 4 | Expression of developmental genes regulated by DEM-CGIs at different developmental stages. (a) Relationships among multiple epigenetic modification on DEM-CGIs and expression of the genes related to them in different genome regions in different development stages. PCC is the Pearson correlation coefficient between the corresponding row and column factors as represented by numbers ranging from −1.0 to 1.0 and color mapped from blue (negative correlation) to red (positive correlation). (b) Relationship between epigenetic modifications and gene expression. Scatter plot (left panel) of epigenetic modification and gene expression in different development stages. Best subsets regression analysis (right panel), where gene expression is the dependent variable and epigenetic modifications are independent variables. * indicates the independent variables in the optimal combination for a given variable number.

analyzed the correlation between epigenetic modifications in 3916 DEM-CGIs and the expression levels of 3699 (82%, 3,699/4,508) DEMGs related to these CGIs (Supplementary Figure S6). We found that epigenetic factors were significantly correlated with each other at nearly all the stages (Figure 4a). For example, the active chromatin marker H3K4me3 showed significant negative correlation with two repressive markers H3K27me3 and DNA methylation in nearly all genome regions studied. A correlation analysis between epigenetic modifications and gene expression revealed that H3K4me3 was positively correlated with gene expression, while H3K27me3 and DNA methylation were negatively correlated with gene expression in all three developmental stages (Figure 4b). A best subsets regression analysis revealed a combination of different kinds of epigenetic modifications may interpret gene expression better than one modification alone; however, the optimal combination varied in the different stages. We suggest that epigenetic modifications correlated with each other in DEM-CGIs related to specific genes, and these modifications may contribute to co-regulation of gene expression.

Differentially expressed genes tend to be differentially modified by epigenetic modifications. We quantified variations in the expression levels of 6,026 genes across the three developmental stages and identified 429 differentially expressed genes (DEGs) (Supplementary Table S5). Interestingly, we found that 80% (341/429) of the DEGs were also DEMGs, termed as DEMGs&DEGs, compared with only 61% (3,699/6,026) expected by chance ($p < 0.0001$; Supplementary Table S6 and Supplementary Figure S6). Functional enrichment analysis revealed the DEGs were enriched in three main clusters of gene ontology biological processes, cell cycle, cell differentiation, and neuron differentiation (Table 1), while only the DEMGs&DEGs were enriched for biological processes related to cell differentiation, especially neuron differentiation. This finding indicated that the DEGs induced by DEM-CGIs are likely to be involved in developmental processes. For example, the DEMG&DEG Ascl1 (also known as Mash1), which encodes a transcription factor essential to neuronal commitment and differentiation during embryogenesis[39], was highly and specifically expressed in NPCs, perhaps because of the increase

Table 1 | Functional enrichment of DEGs based on gene ontology biological process terms

Term Name	Gene Num	Benjamini p value	Term Name	Gene Num	Benjamini p value
Function clusters of 429 DEGs					
Cluster 1 (Enrichment Score: 7.30)					
Cell cycle	44	2.26E-08	Nuclear division	20	8.06E-06
Cell division	28	1.15E-07	M phase of mitotic cell cycle	20	1.01E-05
Cell cycle process	31	1.84E-06	Cell cycle phase	26	1.14E-05
M phase	25	4.33E-06	Organelle fission	20	1.08E-05
Mitosis	20	8.06E-06	Mitotic cell cycle	21	6.02E-05
Cluster 2 (Enrichment Score: 5.89)					
Nervous system development	47	5.25E-06	Developmental process	97	9.05E-05
Anatomical structure development	87	4.56E-06	Cellular developmental process	64	5.25E-04
System development	82	7.36E-06	Cell differentiation	62	5.40E-04
Multicellular organismal development	94	1.47E-05	Organ development	64	7.41E-04
Cluster 3 (Enrichment Score: 5.89)					
Nervous system development	47	5.25E-06	Neuron development	20	1.56E-03
Brain development	22	1.72E-04	Neuron differentiation	24	1.52E-03
Cell development	34	1.83E-04	Neuron projection development	17	1.50E-03
Generation of neurons	29	4.97E-04	Cell morphogenesis	19	8.65E-03
Cellular developmental process	64	5.25E-04	Cell morphogenesis involved in neuron differentiation	14	9.16E-03
Cell differentiation	62	5.40E-04	Cell morphogenesis involved in differentiation	15	1.13E-02
Central nervous system development	24	6.09E-04	Axonogenesis	13	1.12E-02
Neurogenesis	30	6.26E-04	Neuron projection morphogenesis	13	2.01E-02
Function clusters of 341 DEMGs&DEGs					
Cluster 1 (Enrichment Score: 4.05)					
Nervous system development	35	8.34E-03	Developmental process	75	7.25E-03
Multicellular organismal development	73	5.37E-03	Cell development	26	1.16E-02
Anatomical structure development	65	3.82E-03	Cell differentiation	47	3.47E-02
System development	61	4.08E-03	Cellular developmental process	47	4.72E-02
Cluster 2 (Enrichment Score: 2.91)					
Nervous system development	35	8.34E-03	Generation of neurons	21	4.56E-02
Cell development	26	1.16E-02	Brain development	15	4.77E-02
Neuron differentiation	19	3.13E-02	Neuron development	15	4.65E-02
Cell differentiation	47	3.47E-02	Cellular developmental process	47	4.72E-02
Neuron projection development	13	4.82E-02	Forebrain development	11	4.55E-02
Function clusters of 88 other DEGs					
Cluster 1 (Enrichment Score: 8.90)					
Cell cycle	21	1.03E-09	Mitotic cell cycle	13	1.36E-07
Cell cycle process	17	5.69E-09	M phase of mitotic cell cycle	12	1.38E-07
Cell division	14	8.64E-08	Organelle fission	12	1.40E-07
Nuclear division	12	1.66E-07	Cell cycle phase	14	2.15E-07
Mitosis	12	1.66E-07	M phase	13	4.06E-07
Cluster 2 (Enrichment Score: 3.47)					
Chromosome segregation	8	1.24E-06	Sister chromatid segregation	4	5.21E-03
Mitotic sister chromatid segregation	4	4.74E-03	Mitotic metaphase plate congression	3	7.22E-03

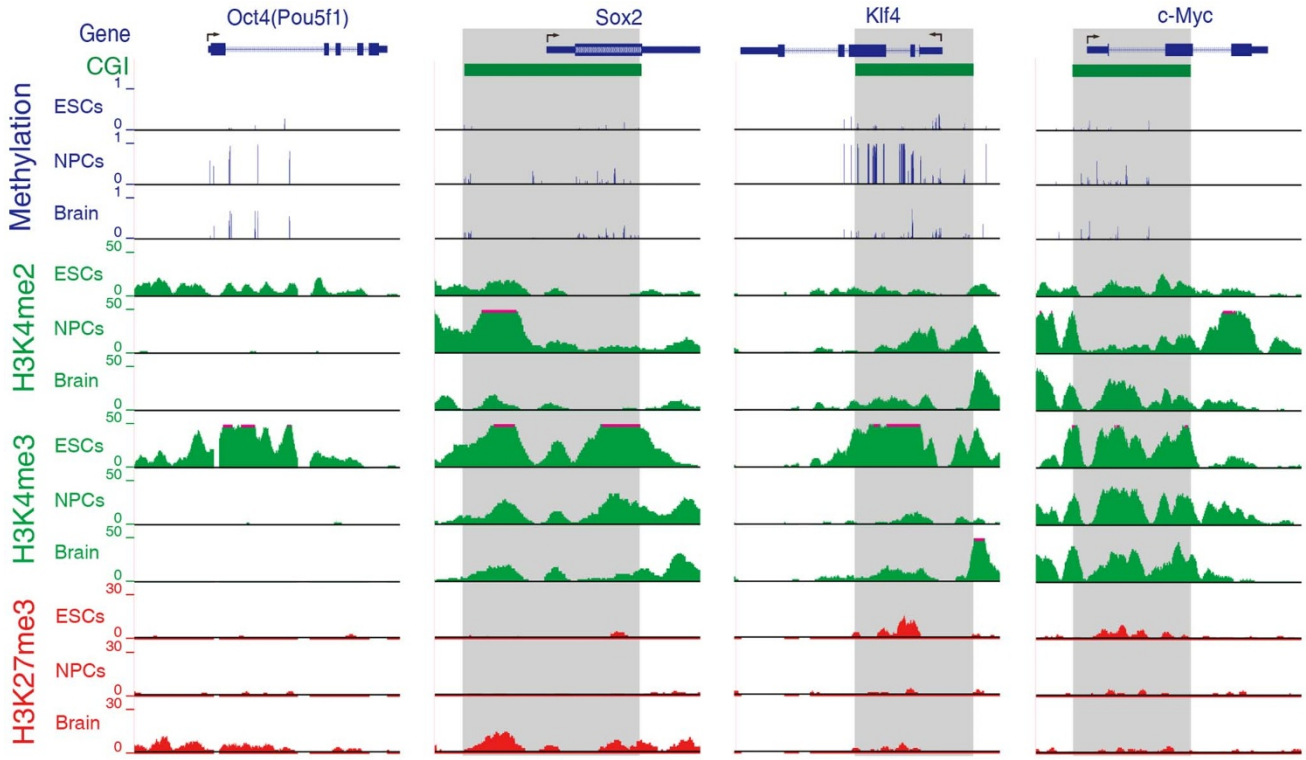

Figure 5 | Epigenetic modification pattern in the CGI/CpGs related to core transcription factors. UCSC Browser view of epigenetic modifications in the CGI/CpGs related to four core transcription factors.

of H3K4me2 and decrease of H3K27me3 modification of the CGIs in the 5′UTR region of this gene (Supplementary Figure S7).

Discussion

Various studies have focused on the interrelationships among epigenetic modifications and the extensive combination of DNA and histone modifications using correlative and direct approaches[5,15]. Here, we proposed a quantitative strategy to decipher the general question of how epigenetic modifications vary cooperatively and determine their roles in the regulation of developmental genes. We found evidence for the quantitative co-variation of genome-wide epigenetic modifications in CGIs and their co-regulation of developmental genes. The implications of these findings are discussed below.

The estimation of epigenetic variation-based entropy made it feasible to explore the quantitative and positive correlation between DNA methylation and H3K27me3 difference. Recently, mutual exclusiveness between H3K27me3 and DNA methylation in CGIs was reported in mouse ESCs using sequential ChIP-bisulfite-sequencing[15]. Our data strongly suggested antagonism between the two repressive markers during the dynamic development process may contribute to long-term repression of developmental genes, which would be activated in specific cell types, followed by alterations in epigenetic modifications[40]. Aberrant epigenetic alterations such as global DNA hypomethylation and formation of repressive chromatin domains may be a potential epigenetic pathway for gene regulation in cancer cells[19]. Thus, we propose that antagonism between H3K27me3 and DNA methylation in CGIs exists widely in multiple cell lines and may play irreplaceable roles in the regulation of the main genes related to pluripotency maintenance and committed differentiation.

Dynamic epigenetic modifications may participate in the regulation of important developmental genes such as core transcription factors. The fundamental roles of four core transcription factors (Oct4, Sox2, Klf4 and c-Myc) in programming and reprogramming have been established in an increasing number of studies[41]. Three of the transcription factors, Sox2, Klf4 and c-Myc, have CGIs in their promoter regions (Figure 5a). A recent study in human revealed the differentiation-associated differential methylation of pluripotency-associated transcription factors including OCT4 and KLF4[42]. Consistent with this observation, we found that the CGI in the Klf4 promoter and the CpGs in an intron of Oct4 (also known as Pou5f1) underwent dynamic DNA methylation and H3K4me3 during the differentiation from ESCs to adult brain (Figure 5). The CGI in the Sox2 promoter represented the transition from H3K4me3 to H3K4me2 during differentiation from ESCs to NPCs. The CGI in the c-Myc promoter showed stable epigenetic modifications during differentiation, which may explain why c-Myc is dispensable for direct reprogramming of mouse fibroblasts[43]. We propose that epigenetic modifications may participate in mediating cellular programming and reprogramming by dynamically regulating indispensable differentiation-associated transcription factors.

The dynamics of epigenetic modification in CGIs may be indispensable for genomic imprinting, which is a feature of mammalian development. There is increasing evidence that genomic imprinting is an epigenetic paradigm that involves DNA methylation and histone modifications, which can affect neuron development[44,45]. In this study, there were 30 imprinted genes among the 7,244 genes related to 8,337 CGIs. Interestingly, about 83% (25/30) of the imprinted genes were related with DEM-CGIs, while only 62% (4,449/7,214) of the non-imprinted genes were related to DEM-CGIs. Thus, the imprinted genes overlapped with DEMGs much more than expected (Chi-square test, $p < 0.05$, Supplementary Table S7 and Supplementary Figure S8). Most of the 25 imprinted genes have been verified as expressed in a parent-of-origin-specific manner in ESCs and brain in several studies (Supplementary Table S8). Dynamic epigenetic modifications in CGIs during cell differentiation may be markers of imprinted genes, which may provide a novel way for identification of more imprinted genes in mammalian genomes.

Previous studies focused on epigenetic modifications of CGIs in gene promoter regions, and CGIs with a dynamic epigenetic state

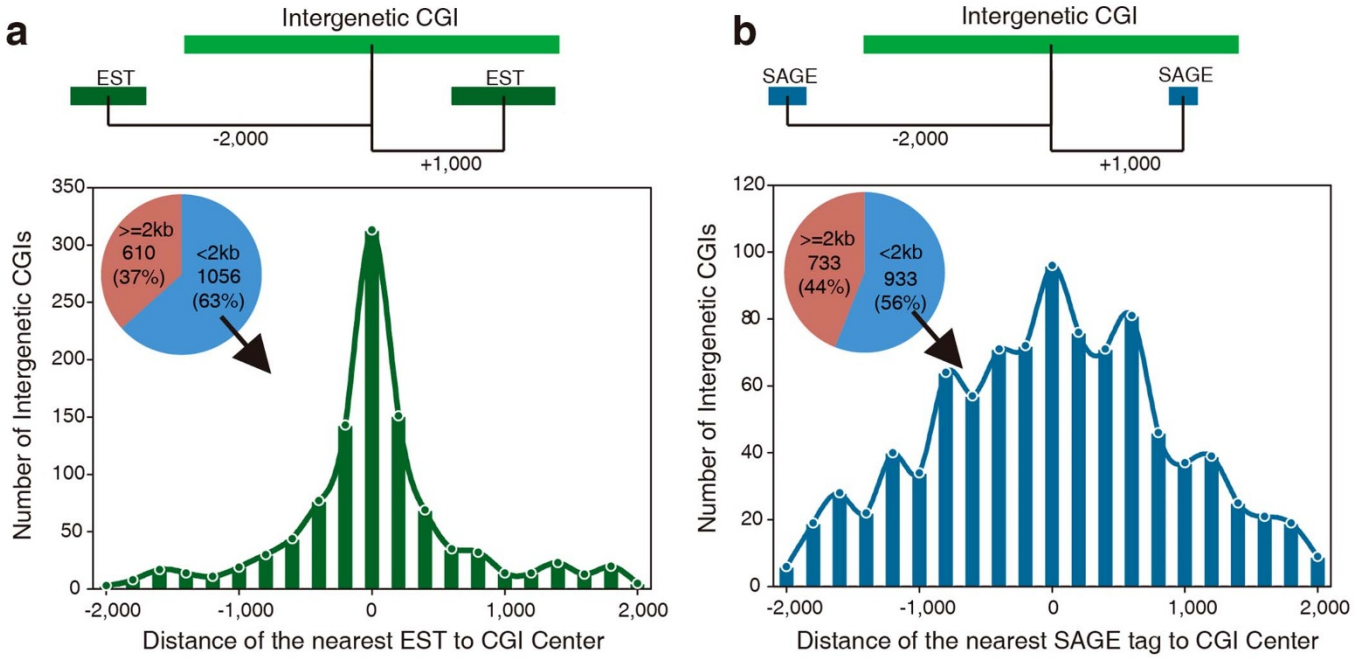

Figure 6 | Relative distances between Intergenic CGIs, and ESTs and SAGE tags. (a) Distribution of the distance of the nearest EST to the center of Intergenic CGIs in mouse. The pie chart shows the proportion of CGIs with different distances to ESTs. The histogram shows the number of CGIs within a distance of less than 2 kb from an EST. (b) Distribution of the distance of the nearest SAGE tag to the center of Intergenic CGIs in mouse.

were regarded as the functional regions involved in regulation of gene transcription[6,46]. Consistent with these observations, we found a significant co-variation of epigenetic modifications in promoter CGIs (CGIs in Up2kb and 5′UTR regions), supporting their probable roles in co-regulation of gene expression. Recent studies revealed that intragenic DNA methylation may also play important roles in the regulation of alternative promoters and in differential gene expression[26,47]. These findings were also confirmed in the present studies. We also suggested that the Intergenic CGIs, which were generally ignored in previous gene regulation studies, experienced dynamic combinatorial epigenetic changes similar to the gene-related CGIs. A possible explanation is that the Intergenic CGIs are functional regions related to chromatin structure. Intergenic CGIs may also be the regulatory elements for novel coding or non-coding genes, which was supported by our finding that most of the Intergenic CGIs were localized near gene transcripts from expressed sequence tag (EST) and serial analysis of gene expression (SAGE) data (Figures 6a and b). Medvedeva et al. have reported that Intergenic CGIs were enriched in the binding sites of Sp1, which can be recruited by accessible chromatin structure to regulate gene expression[25,48]. In a recent study, Aran et al. found that methylation of distal regulatory sites was related closely to gene expression levels and cell-specific enhancer methylation may modulate cell-specific transcription levels[27]. These observations revealed genome-wide universal synergy among different epigenetic modifications during differentiation. We suggest that CGIs may be an essential feature of chromatin structure defining dynamic gene expression in mammals.

Methods

CGIs and genomic annotation. 15,948 mouse CGIs (mouse genome mm8) were downloaded from the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables)[28]. The CGIs were classified into seven genome regions: Up2kb (from 2-kb upstream to the transcription start site of a gene); 5′UTR (from the transcription start site to the end of the 5′UTR); CodingExon (all the coding exons except the exon in the 5′UTR); Intron (all the introns of a gene); 3′UTR (from the start of the 3′UTR to the transcription termination site of a gene); Down2kb (the transcription termination site to 2 kb downstream of a gene); and Intergenic (2 kb distant from any gene). For each CGI, the RefSeq gene closest to it on the genome was identified and then the CGI was classified into one of the seven genome regions as described in our previous work[23].

DNA methylation data. DNA methylation data from mouse (mouse genome mm8) were downloaded from ftp://ftp.broad.mit.edu/pub/papers/rrbs/Meissner2008/[5]. This dataset contains mouse genome-wide methylation profiles of about 1 million distinct CpG dinucleotides detected by reduced representation bisulfite sequencing. The methylation level of a CGI in each of three tissue/cells (ESCs, NPCs and brain) was estimated as the mean methylation level across all CpG dinucleotides with ≥5-fold coverage overlapping the same CGI, requiring at least five fulfilled CpGs. In this way, we obtained 8,337 CGIs with their associated methylation data in the three tissue/cells for DNA methylation analysis.

Histone modification data. The histone modification data used in this study were downloaded from the Gene Expression Omnibus (GEO) repository (accession numbers GSE12241 and GSE11172)[5,49,50]. Three histone modifications (H3K4me2, H3K4me3 and H3K27me3), which have been detected in all three development stages (ESCs, NPCs and brain), were used to study dynamic changes of histone modification during differentiation. For each CGI, the histone modification tags that were centered in the CGI were counted. The tag count was normalized by the total number of bases in the region to obtain normalized histone modification levels for histone modification analysis.

Gene expression data. The gene expression data used in this study were downloaded from GEO; accession numbers GSE8024 (ESCs and NPCs) and GSE10246 (brain)[50,51]. All these expression data were detected using the same Gene Expression Array (Affymetrix Mouse Genome 430 2.0 Array). The annotations of probes were also downloaded from GEO (accession number GPL1261). For each probe, the expression value was the mean of the GCRMA-normalized fluorescence intensities in two replicates per cell/tissue. The mean expression value was used when multiple probes were available for a single RefSeq gene. Finally, the log2 transformed expression values of 6,026 RefSeq genes related to 7,771 CGIs were used for further analysis.

Quantification of epigenetic variation and identification of DEM-CGIs. Modified Shannon entropy was used to quantify dynamic epigenetic variation during neural differentiation and to identify the DEM-CGIs. For the DNA methylation data, the methylation difference for each CGI among different cells/tissues was quantified using QDMR[23]. QDMR is an entropy-based method for quantification of methylation difference and identification of differentially methylated regions. For a CGI, the methylation value in it varies across ESCs, NPCs and Brain. The methylation values of a CGI across multiple samples can be regarded as a dataset. As Shannon entropy is a quantitative measure of difference and uncertainty in a dataset[52], the methylation difference of can be measured by entropy-based method QDMR. Because Shannon entropy is independent of data distribution, QDMR can be used to DNA methylation data which follows bimodal distribution[53]. The QDMR entropy ranges from zero for regions differentially methylated in a single sample to a maximum value for regions with uniform methylation levels in all samples considered. A default threshold (0.962 ± 0.024) for three samples was obtained from the probability model described in QDMR. The threshold was used to identify DNAm-DEM-CGIs. CGIs with entropy

below the threshold were identified as DNAm-DEM-CGIs; the other CGIs were assigned as NonDNAm-DEM-CGIs.

There were several extremely large entropy values in the histone modification data for each sample, which may reflect the real modification intensity but which cannot be used to quantify histone modification variations. To quantify the variation of histone modification across cells/tissues, we optimized the entropy method used in QDMR as follows: (i) the data in each sample were preprocessed by computing the mean (μ) and the standard deviation (σ) in each sample and then replacing any that were over three standard deviations away from the mean by $\mu + 3\sigma$; (ii) for each type of histone modification, the maximum (*MAX*) and minimum (*MIN*) values of all preprocessed modification levels ($L_{hm,cgi,s}$) in the three cells/tissues were used to obtain standardized modification levels $SL_{hm,cgi,s} = (L_{hm,cgi,s} - MIN)/MAX$ that ranged from 0 to 1. The standardized modification levels were used to quantify the modification difference across the three stages; and (iii) the CGIs, which were differentially modified by histone modifications, were identified using the same threshold that was used for DNAm-DEM-CGIs. This optimized entropy method for the pretreatment and analysis of histone modification data has been introduced into the QDCMR software and the command line version is available at http://github.com/hbliu/QDCMR.

Quantification of gene expression variation and identification of DEGs. Because the characteristics of gene expression data are similar to histone modification density data, QDCMR was also used to quantify gene expression variation and to identify DEGs during mouse differentiation.

The association between Intergenic CGIs and ESTs and SAGE tags. Mouse ESTs and SAGE tags were downloaded from the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables)[28]. For each of the 1,666 Intergenic CGIs, the nearest EST was identified and the distance between them was calculated. This process was repeated for the SAGE data.

Statistical analysis and gene ontology analysis. SigmaPlot version 11.0 was used for the Wilcoxon signed rank test, Pearson correlation, best subsets regression analysis, and to draw the figures. SPSS version 19.0 was used for the chi-square test. The DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/) was used to analyze the gene functional enrichment under the gene ontology biological process[54].

1. Bibikova, M., Laurent, L. C., Ren, B., Loring, J. F. & Fan, J. B. Unraveling epigenetic regulation in embryonic stem cells. *Cell Stem Cell* **2**, 123–134 (2008).
2. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).
3. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412–1417 (2006).
4. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).
5. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
6. Vavouri, T. & Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol* **13**, R110 (2012).
7. Blackledge, N. P. *et al.* CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* **38**, 179–190 (2010).
8. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
9. Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS genetics* **6**, e1001244 (2010).
10. Rumble, S. M. *et al.* SHRiMP: accurate mapping of short color-space reads. *PLoS computational biology* **5**, e1000386 (2009).
11. Sanz, L. A. *et al.* A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *EMBO J* **27**, 2523–2532 (2008).
12. Orford, K. *et al.* Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Developmental cell* **14**, 798–809 (2008).
13. Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* **107**, 2926–2931 (2010).
14. Kugler, K. G., Mueller, L. A., Graber, A. & Dehmer, M. Integrative network biology: graph prototyping for co-expression cancer networks. *PLoS One* **6**, e22843 (2011).
15. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128–1138 (2012).
16. Lv, J. *et al.* Discovering cooperative relationships of chromatin modifications in human T cells based on a proposed closeness measure. *PLoS ONE* **5**, e14219 (2010).
17. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662–673 (2002).
18. Tian, R., Feng, J., Cai, X. & Zhang, Y. Local chromatin dynamics of transcription factors imply cell-lineage specific functions during cellular differentiation. *Epigenetics* **7**, 55–62 (2012).
19. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* **22**, 246–258 (2012).
20. Deal, R. B. & Henikoff, S. Capturing the dynamic epigenome. *Genome Biol* **11**, 218 (2010).
21. Xu, H., Wei, C. L., Lin, F. & Sung, W. K. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics* **24**, 2344–2349 (2008).
22. Taslim, C., Huang, T. & Lin, S. DIME: R-package for identifying differential ChIP-seq based on an ensemble of mixture models. *Bioinformatics* **27**, 1569–1570 (2011).
23. Zhang, Y. *et al.* QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res* **39**, e58 (2011).
24. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
25. Medvedeva, Y. A. *et al.* Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC genomics* **11**, 48 (2010).
26. Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* **21**, 1074–1086 (2011).
27. Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biology* **14** (2013).
28. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493–496 (2004).
29. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740–3745 (2002).
30. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645 (2009).
31. Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15–31 (2012).
32. Jiang, Y. *et al.* Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood* **113**, 1315–1325 (2009).
33. Ranheim, E. A. *et al.* Frizzled 9 knock-out mice have abnormal B-cell development. *Blood* **105**, 2487–2494 (2005).
34. Thaler, J. P. *et al.* A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. *Neuron* **41**, 337–350 (2004).
35. Diaz-Meyer, N. *et al.* Silencing of CDKN1C (p57KIP2) is associated with hypomethylation at KvDMR1 in Beckwith-Wiedemann syndrome. *J Med Genet* **40**, 797–801 (2003).
36. Yang, X. *et al.* CDKN1C (p57) is a direct target of EZH2 and suppressed by multiple epigenetic mechanisms in breast cancer cells. *PLoS One* **4**, e5011 (2009).
37. Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123–1134 (1991).
38. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
39. Pattyn, A. *et al.* Ascl1/Mash1 is required for the development of central serotonergic neurons. *Nature neuroscience* **7**, 589–595 (2004).
40. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295–304 (2009).
41. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
42. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res* **20**, 320–331 (2010).
43. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**, 10–12 (2008).
44. Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* **12**, 565–575 (2011).
45. Wilkinson, L. S., Davies, W. & Isles, A. R. Genomic imprinting effects on brain development and function. *Nature reviews. Neuroscience* **8**, 832–843 (2007).
46. Yagi, S. *et al.* DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res* **18**, 1969–1978 (2008).
47. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
48. Lee, J. J. *et al.* Accessible chromatin structure permits factors Sp1 and Sp3 to regulate human TGFBI gene expression. *Biochem Biophys Res Commun* **409**, 222–228 (2011).
49. Barrett, T. *et al.* NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* **37**, D885–890 (2009).
50. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
51. Lattin, J. E. *et al.* Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res* **4**, 5 (2008).
52. Shannon, C. E. The mathematical theory of communication. 1963. *MD Comput* **14**, 306–317 (1997).
53. Rakyan, V. K. *et al.* DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* **2**, e405 (2004).

54. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

Acknowledgments

This study was supported financially by grants from the National Natural Science Foundation of China (31171383, 31371478, 61075023, 31371334), the Fundamental Research Funds for the Central Universities (HIT. NSRIF.2010027), the Natural Science Foundation of Heilongjiang Province (C201217), and the Scientific Research Fund of Heilongjiang Provincial Education Department (12511272, 12521270).

Author contributions

Q.W. and Y.Z. conceived the idea. HHL designed the experiments, performed the bioinformatics analyses, prepared the figures, and wrote the manuscript. Y.J.C. and J.L. contributed to gene expression analysis. H.H.L., Y.J.C. and H.L. performed the statistical analysis. H.H.L., H.L. and R.F.Z. developed the software. J.L., J.Z.S. and X.J.L. participated in gene annotation. All authors have read and approved the final manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, H.B. *et al*. Quantitative epigenetic co-variation in CpG islands and co-regulation of developmental genes. *Sci. Rep.* **3**, 2576; DOI:10.1038/srep02576 (2013).