Howard Hughes Medical Institute

Published as: *Cell Rep*. 2016 December 20; 17(12): 3369–3384.

Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain

Chongyuan Luo[1,2,5], **Madeline A. Lancaster**[3,4,5], **Rosa Castanon**[1], **Joseph R. Nery**[1], **Juergen A. Knoblich**[3,*], and **Joseph R. Ecker**[1,2,6,*]

[1]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, United States

[2]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, United States

[3]Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna 1030, Austria

Summary

Organoids derived from human pluripotent stem cells recapitulate the early three-dimensional organization of the human brain, but whether they establish the epigenomic and transcriptional programs essential for brain development is unknown. We compared epigenomic and regulatory features in cerebral organoids and human fetal brain, using genome-wide, base resolution DNA methylome and transcriptome sequencing. Transcriptomic dynamics in organoids faithfully modeled gene expression trajectories in early-to-mid human fetal brains. We found that early non-CG methylation accumulation at super-enhancers in both fetal brain and organoids marks forthcoming transcriptional repression in the fully developed brain. Demethylated regions (74% of 35,627) identified during organoid differentiation overlapped with fetal brain regulatory elements. Interestingly, pericentromeric repeats showed widespread demethylation in multiple types of *in vitro* human neural differentiation models but not in fetal brain. Our study reveals that organoids recapitulate many epigenomic features of mid-fetal human brain and also identified novel non-CG methylation signatures of brain development.

Introduction

The differentiation of human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs) into neural tissues provide promising models for studying human brain development, psychiatric disorders and will potentially enable stem cell therapies (Lancaster and Knoblich, 2014; Yin et al., 2016; Fatehullah et al., 2016; Johnson and Hockemeyer,

*Corresponding author.
[4]Current address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom
[5]Co-first author
[6]Lead Contact

Author Contributions

J.R.E. and J.A.K. designed and supervised the research. M.A.L. generated the cerebral organoids and collected samples. R.C. prepared MethylC-seq and RNA-seq libraries. J.R.N. sequenced the libraries. C.L. and M.A.L. analyzed data. C.L. prepared the manuscript. M.A.L., J.A.K. and J.R.E. revised the manuscript. s

2015). *In vitro* neural differentiation recapitulates the program of early brain development starting with the initial commitment to neuroepithelium followed by the sequential generation of neural progenitor cells, and formation of early and late born neurons. Three dimensional culture models allow the formation of more complex developing brain structures such as progenitor zones along the apical-basal axis (Eiraku et al., 2008; Danjo et al., 2011; Eiraku et al., 2012; Mariani et al., 2012; Lancaster et al., 2013; Qian et al., 2016). While previous studies aimed at characterization of *in vitro* neural differentiation systems have examined the organization of the differentiated tissues, as well as the identity of progenitor and neuron cells, and their transcriptomic profiles and electrophysiological properties (Eiraku et al., 2008; Chambers et al., 2009; Mariani et al., 2012; Lancaster et al., 2013; van de Leemput et al., 2014; Qian et al., 2016), none have performed an analysis at the level of the epigenome, Recently, epigenomic analysis such as whole genome cytosine methylome analysis has been applied in stem cell engineering to evaluate the epigenomic fidelity of iPSCs, somatic nuclear transfer (SCNT) stem cells and naive hESCs (Lister et al., 2011; Ma et al., 2014; Pastor et al., 2016; Theunissen et al, 2016) providing important insights into the fidelity of these *in vitro* system. These studies have identified mega-scale differentially non-CG methylated regions, between hESC and reprogrammed pluripotent cells generated with different techniques (Lister et al., 2011; Ma et al., 2014), which are robust epigenomic signatures of reprogramming methods that could not be discovered by transcriptome sequencing alone.

Human brain development is driven by a transcriptional program with exceptional temporal and spatial precision that is orchestrated by corresponding changes in the activity of numerous regulatory elements along with global remodeling of the epigenome (Bertrand et al., 2002). Here, we aim to examine epigenomic dynamics and regulatory element usage during *in vitro* neural differentiation using cerebral organoids (COs) as a model system and with the method of whole genome methylome sequencing. COs are three dimensional *in vitro* cultures that recapitulate key characteristics of human brain development including brain regional specification, the formation of progenitor layers and the generation of diverse types of functional neurons (Lancaster et al., 2013). Regulatory elements involved in human organ development can be identified by epigenomic features such as open chromatin, histone modifications or local depletion of DNA methylation (Crawford et al., 2006; Heintzman et al., 2011; Stedeler et al., 2011; Thurman et al., 2012; Ziller et al., 2013; Schultz et al., 2015). Cytosine DNA methylation (mC) is an important epigenomic mark that is highly dynamic during human development and across tissues (Lister et al., 2009; Ziller et al., 2013; Schultz et al., 2015). Genome-wide mC maps contain rich information about gene regulatory programs since differentially methylated regions (DMRs) are robust indicators of tissue-specific regulatory elements (Hon et al., 2013). The analysis of whole genome cytosine methylomes of most major human tissue types have identified over a million putative regulatory elements that strongly overlap with open chromatin or histone modification signatures (Ziller et al., 2013; Schultz et al., 2015).

Although mC is predominantly localized at CG dinucleotides in most human tissues, abundant mC in non-CG contexts (mCH) is found in hESCs, brain and a number of other somatic tissues (Lister et al., 2009; Schultz et al., 2015). Human and mouse brain development is associated with dramatic mC remodeling including a notable accumulation

of mCH starting at birth and reaching a plateau during adolescence (Lister et al., 2013; Guo et al., 2014). Strikingly, in mature human neurons, mCH outnumbers mCG and becomes the predominant form of mC (Lister et al., 2013). Unlike mCH in hESCs, which is predominantly located in a CAG context and positively correlates with gene expression, mCH in brain is enriched in a CAC context and is correlated with transcriptional repression (Lister et al., 2009; Lister et al., 2013; Guo et al., 2014). Recent studies have shown that mCH is recognized by MeCP2 which further recruits the NCoR/SMRT co-repressor complex (Lyst et al., 2013; Gabel et al., 2015; Chen et al., 2015). Additional mC features such as DNA methylation valleys (DMVs) are large genomic domains depleted of mC and are commonly colocalized at the promoter of approximately 1,000 genes in any given cell or tissue type (Xie et al., 2013; Jeong et al., 2014). DMVs associated with neuron type-specific transcription factors (TFs) become hyper-methylated in the neuron populations in which that specific TF is developmentally expressed (Mo et al., 2015).

In this study, we apply whole genome methylome profiling and transcriptome sequencing to examine epigenomic and transcriptomic remodeling during *in vitro* differentiation. We compare the transcriptomes of multiple *in vitro* neural differentiation models and find that the various methods could be distinguished from one another by the expression pattern of extracellular matrix component genes. Organoid differentiation, in particular, recapitulated the transcriptomic dynamics of early-to-mid fetal brain development. Further analysis of DNA methylation in COs revealed a novel type of mCH enrichment at fetal brain super-enhancers that was indicative of transcriptional repression in the later developed brain *in vivo*. Hyper-methylation of DNA methylation valleys (DMVs) in COs provided a robust epigenomic signature of forebrain specification. In addition, we found pervasive demethylation of pericentromeric repeats in *in vitro* human neural cultures which is not observed during fetal brain develpment, revealing the importance of using genomic/epigenomic methods to evaluate how well *in vitro* derived neuronal tissues, resulting from stem cell engineering, represent their *in vivo* counterparts. Finally, dynamic mCG was identified at over 65,000 genomic regions that marked the repression and activation of hESC and fetal brain regulatory sequences, respectively. Overall, the remodeling of mCG and mCH observed during CO differentiation led to the formation of an epigenomic state closely resembling human fetal brain.

Results

Cerebral organoids recapitulate transcriptomic signatures of early-to-mid fetal brain development

In order to access the capacity for a variety of *in vitro* neural differentiation methods to model human brain development at the gene regulatory level, we performed comparative transcriptomic analyses using mRNA-seq data generated for COs at several stages. We further compared these data with existing datasets for telencephalic aggregates (Mariani et al., 2015), two dimensional CORTECON cultures (van de Leemput et al., 2014), and the recently established multi-well bioreactor forebrain organoids (Qian et al., 2016). We analyzed H9 hESC, embryoid body (EB) and COs after 40 days and 60 days of differentiation in order to encompass several stages of neural development from early neural

induction to neurogenesis of various neuron populations. Principal component analysis (PCA) identified notable transcriptomic remodeling associated with differentiation from hESC to EB, and from EB to CO 40d (Figure 1A and B), whereas the further differentiation of CO 40d to CO 60d was coupled with much more moderate alterations of gene expression.

To determine the stages of *in vivo* cortical development that best matched these various *in vitro* models, we compared transcriptome data to the BRAINSPAN human brain developmental transcriptome RNA-seq dataset (Kang et al., 2011, Figure 1C and D). At a global level, all four *in vitro* systems showed comparable levels of correlation to early fetal (stage 1, 8–9 PCW) cortex (Figure 1D, box – a). However, mid-fetal cortex samples (stage 2 and 3, 12–16 PCW) were more strongly correlated with COs and multi-well organoids (Figure 1D box – b and Figure S1A box – b). To identify the gene groups that contributed to the difference between the *in vitro* culture systems and fetal cortex, we applied principal component analysis (PCA) to the transcriptome of *in vitro* and *in vivo* samples and found that the majority of variance could be explained by just two principal components, PC1 (40.2%) and PC2 (22.4%, Figure 1E). Consistent with the monotonic increase of PC1 scores along with the progression of *in vivo* cortical development, the expression of genes enriched for synaptic transmission and ion transport functions positively contributed to PC1, whereas the expression of translation machinery and cell cycle related genes were inversely correlated with PC1 (Figure 1E and F). PC2 was positively correlated with the expression of neurogenesis and axonogenesis genes, which explains the greater PC2 scores observed for mid-fetal cortex samples (Figure 1E and F). Intriguingly, the group of genes with the strongest negative contribution to PC2 was enriched for extracellular matrix (ECM) component functions (Figure 1F). The *in vitro* systems showed distinct gene expression dynamics that were captured by PC2. CO and multi-well organoids were associated with increased PC2 scores, whereas PC2 scores decreased during differentiation of telencephalic aggregates and CORTECON cultures (Figure 1E). Interestingly, distinct PC2 dynamics were also observed for fetal cortical development. Early-to-mid (from stage 1 to 2) and mid-to-late (from stage 3 to 4) fetal development stages were associated with a strong increase and decrease of PC2 scores, respectively (Figure 1E). The results suggest that characteristics of early-to-mid or mid-to-late fetal cortical development, represented by PC2, were differentially recapitulated by various *in vitro* differentiation models.

Multi-well organoids represent a modified CO method with the addition of dual SMAD inhibition in the early stage of differentiation (Qian et al., 2016). In line with their methodological similarities, the transcriptomic profiles of CO and multi-well organoids were highly correlated. CO 60d and multi-well organoid 54d samples colocalized in the space of PC1 and PC2 (Figure 1E). Extensively (100 days) cultured multi-well organoid samples showed significantly greater PC1 and PC2 scores than other *in vitro* cultures (hence more closely resembled mid-fetal cortex), including earlier time points of CO and multi-well organoid samples (Figure 1E). The results suggest that COs are capable of further differentiation and maturation, if the duration of the time in culture is extended (Qian et al., 2016).

Selected gene sets with strong contributions to PC1 (synaptic transmission) and PC2 (neurogenesis and ECM) were analyzed to identify genes that were differentially regulated

in *in vitro* systems (Figure 1F-J). Comparable up-regulation of genes with synaptic transmission and neurogenesis functions was observed during CO, multi-well organoid and telencephalic aggregate differentiation methods, whereas a substantially weaker induction was observed for CORTECON (Figure 1G, and H left panels). Notably, the expression of ECM genes was strongly activated in telencephalic aggregates and CORTECON, but not during CO and multi-well organoid differentiation (Figure 1I, left panel). The upregulation of ECM genes was a characteristic of mid-to-late fetal development that occurred between stages 3 and 4 of cortical development (Figure 1I, right panel). However, the expression level of the ECM genes was much greater in telencephalic aggregates and CORTECON than in the brain (Figure 1J). The expression of this group of genes was repressed and showed little dynamics during mid fetal development, which was similar to that observed in CO and multi-well organoids (Figure 1I and J). Therefore the differential transcriptomic dynamics of *in vitro* cultures captured by PC2 in the PCA analysis could be at least partially explained by the distinct expression patterns of ECM genes, which were unusually high in CORTECON and telencephalic aggregates compared with CO samples and fetal brain.

Co-expression gene modules with neuronal functions are preserved in cerebral organoids

To identify the biological processes that are conserved between the various differentiation methods and *in vivo* cortex development, we identified 28 co-expression gene modules across *in vivo* cortical development with WGCNA (Figure S1B, Langfelder and Horvath, 2008). The enriched gene ontology terms for each module are listed in Table S2. A number of modules were enriched in biological processes that undergo major changes during brain development. For example, module 15 was enriched in cell cycle functions and was drastically down-regulated in early-to-mid fetal development (Figure S1B). Modules 2, 5, 10 and 13 were enriched in synaptic transmission and ion transporter functions and were up-regulated during brain development (Figure S1B). Following identification of these *in vivo* co-expression modules, we determined their preservation in *in vitro* differentiation systems (Figure S1C-F, Langfelder et al., 2011). Using a Zsummary > 4 threshold, CO differentiation had the greatest number (17/28) of preserved modules compared to other *in vitro* systems (Figure S1C). Importantly, all modules (2, 5, 10 and 13) that were enriched for neuronal differentiation, synaptic transmission and ion transport functions were preserved in COs (Figure S1C).

We next examined the overlap of differentially expressed genes (fold change >= 2 fold) in *in vitro* culture systems with the 28 coexpression modules (Figure S1G). Interestingly, module 20 and 28, which were enriched for ECM components were significantly enriched for up-regulated genes in telencephalic aggregates and CORTECON, but were enriched in down-regulated genes in CO and multi-well organoids (Figure S1G). Further analyses of ECM gene expression in modules 20 and 28 supported the results above that ECM genes were induced in telencephalic aggregates and CORTECON differentiation and were overexpressed compared to cortical samples (Figure S1H and I).

Take together, the transcriptomic analysis of *in vitro* neural differentiation systems revealed extensive correlation between COs and multi-well organoids. Among the analyzed culture methods, the two organoid methods most strongly overlapped with mid-fetal human cortex

signatures including neuronal as well as ECM gene expression. The organoid methods further have the potential of generating neural tissues that better resemble fetal cortex through extended culture periods.

Non-CG methylation is enriched at super-enhancers in fetal cortex and cerebral organoids

Human pluripotent and somatic cell types are associated with dramatically different patterns of non-CG methylation (mCH) with respect to both sequence context and correlation with gene expression (Lister et al., 2009; Schultz et al., 2015). The CO method provides a unique opportunity to study mCH reconfiguration during human brain development. Differentiation of hESC to COs was associated with a drastic genome-wide reduction of mCH with only a minor alteration of the global mCG level (Figure 2A–C). CO 40d, 60d and fetal cortex contained low but significant levels of mCH (Figure 2B and C). Later stages of CO differentiation (from 40d to 60d) were associated with increased mCH levels (Figure 2B and C). Trinucleotide CAG and CAC sites are preferentially methylated in hESC and somatic tissues, respectively (Lister et al., 2009; Lister et al., 2013; Schultz et al., 2015). Consistent with the somatic identity of neural tissues, mCH sites in fetal cortex, CO 40d and 60d were enriched at CAC sites (Figure 2E and Figure S2A). Gene body mCH and gene expression were positively correlated in hESC (Lister et al., 2009), but were inversely correlated in developed somatic tissues (Lister et al., 2013; Schultz et al., 2015). The positive correlation between gene body mCH and gene expression was substantially reduced in CO 60d (Spearman r = 0.137 or 0.205) compared to hESC (Spearman r = 0.608 or 0.615, Figure S2B). Gene body mCH and gene expression were very weakly correlated in the fetal cortex (Spearman r = 0.024, $p=8.1\times10^{-4}$, Figure S2B). Hierarchical clustering of global mCH found that CO 40d and 60d were most similar to mid-fetal cortex (Figure 2D), which together formed a distinct cluster that was separated from the upper left cluster consisting of hESC and EB, and lower right cluster including further developed cortex samples (Figure 2D). Clustering by global mCG was dominated by the separation of *in vitro* versus *in vivo* samples, which could be contributed by the presence of large hypo-DMR blocks in COs as further discussed below (Figure S2C).

Interestingly, we found that the sparse mCH in COs and fetal cortex was enriched in continuous genomic domains modified by H3K4me1 (Figure 2H and Figure S2D). Large genomic domains with high levels of Mediator binding or enhancer histone marks H3K27ac and H3K4me1 are characteristic of super-enhancers (Whyte et al., 2013). To test whether mCH was enriched in super-enhancers in COs and fetal cortex, we identified putative super-enhancers with ChIP-seq data of H3K4me1 that were available for both fetal brain and adult cortex (Whyte et al., 2013; Roadmap Epigenomics Consortium et al., 2015). The super-enhancers identified using adult cortex H3K27ac and H3K4me1 highly overlapped (Figure S2E), which validated the continuous domain of H3K4me1 as a marker for super-enhancers. Most (82.1% 1,130/1,383) putative fetal brain super-enhancers overlapped with gene bodies. Enrichment for mCH was found at genes that overlapped with one or multiple putative super-enhancers, such as *Auts2*, *Nfia*, *Cux1* or *Prdm16* (Figure 2H and Figure S2D). In both COs and fetal cortex, putative super-enhancers were associated with significantly greater levels of mCH compared to shuffled regions (Wilcoxon rank sum test, $p<8\times10^{-113}$, Figure 2F). Notably, the enrichment of mCH at regulatory elements in COs and fetal cortex

contrasts with the depletion of mCH at putative adult cortical super-enhancers relative to shuffled regions ($p=9.5\times10^{-25}$ for 25 yr cortex, Figure 2G).

To address whether mCH was associated with active or repressive states at fetal brain super-enhancers, we analyzed the developmental expression of genes located closest to the predicted super-enhancer (Figure 2I–N). Consistent with this proposed repressive role of mCH, genes nearby fetal super-enhancers containing high mCH were more strongly down-regulated in neonatal and postnatal development, as illustrated by the developmental expression patterns of *Auts2*, *Nfia*, *Cux1* or *Prdm16* (Figure 2H, K and M). Intriguingly, genes nearby fetal super-enhancers containing high mCH were expressed actively in early fetal brain development (8–9 PCW, age group 1), whereas genes nearby fetal super-enhancers containing low mCH were repressed in 8–9 PCW and upregulated during mid-fetal development (age group 2, Figure 2L and N). In summary, mCH accumulation in fetal brain and COs marked loci that overlapped with super-enhancers, and were actively expressed during early fetal stages but became repressed after the transition to neonatal/postnatal development.

Hyper-methylation of DNA methylation valleys predicts the forebrain specification of cerebral organoids

Neuronal hyper-methylation of DMVs associated with developmental TF genes is informative about gene expression history and inference of cell-subtype (Mo et al., 2015 and Mo et al., 2016). We further asked whether hyper-methylation of DMVs in COs was informative of brain region specification by examining a panel of established brain region marker genes (Kirkeby et al., 2012). Notably, 39 out of 42 (92.9%) marker genes examined overlapped with a DMV, confirming their strong enrichment of developmental regulators. DMVs that overlapped with four of the rostral markers – *Otx2*, *Lhx2*, *Six3* and *Pax6* showed significant (>0.1 mean mCG/CG gain in CO 40d and 60d) hyper-methylation in COs compared to H9 hESC (Figure 3A and B). All non-rostral markers, except for *Gbx2*, showed insignificant increases in mC (Figure 3A and B). In addition, no hyper-methylation was found for DMVs overlapping with eight mesoderm and endoderm marker genes (Figure 3A). The DMV hyper-methylation of *Otx2*, *Lhx2*, *Six3*, *Pax6* as well as the caudal marker gene Gbx2 loci in COs were paralleled with mCG accumulation in fetal or developed cortex, as well as neurospheres explanted from cortex or ganglionic eminence of 13 and 17 PCW fetus (Roadmap Epigenomics Consortium et al., 2015; Gu et al., 2016, Figure 3A). The results suggest that the DMV hyper-methylation observed in COs was consistent with the developmental program observed in rostral brain and primary brain cells (Figure S3D).

Given the strong association between brain region marker genes and DMV hyper-methylation, we set out to identify additional genes that may be involved in regional specification of COs. Most genes that colocalized with DMVs were modified by either H3K4me3 or H3K27me3 histone modifications, or both (Xie et al., 2013; Mo et al, 2015). Genes with roles in brain differentiation are likely repressed in hESCs and modified by repressive mark H3K27me3. By stratifying DMVs according to their histone modification patterns in H9 hESC, we found that genes colocalized with H3K27me3+ DMVs, including DMVs with H3K4me3+/H3K27me3+ bivalent states, were repressed in hESC, whereas

genes colocalized with monovalent H3K4me3+ DMVs were instead actively expressed in hESC (Figure S3A and B). All (39/39) DMVs associated with brain regional marker genes were modified with H3K27me3. Consistent with the speculation that DMV hyper-methylation is associated with developmental gene expression, hyper-methylated H3K27me3+ DMVs were associated with increased gene expression, H3K4me3 signals and moderate decreases of H3K27me3 signals (Figure 3C). We further categorized H3K27me3+ DMVs by their mC states in COs and developed cortex. DMVs (73 of 619) were hyper-methylated in both COs and cortex (CO+/Ctx+) including rostral brain markers such as *Pax6*, *Otx1/2*, *Lhx2* and *Six3*. Gene ontology (GO) terms such as pattern specification process, forebrain and sensory organ development were consistently enriched for genes overlapping with CO+/Ctx+ DMVs. In contrast, genes overlapping with CO-/Ctx- DMVs (397 of 619) were enriched for anterior/posterior pattern formation and skeletal system development. Finally, DMV hyper-mCG in COs was predictive of the expression of DMV associated genes in fetal cortex. Genes overlapping with both CO+/Ctx- and CO+/Ctx+ DMVs were more actively expressed in the fetal cortex (Figure S3C).

Comparison of DMV hyper-methylation between COs and cortical development allowed an estimation of the developmental stage of COs by scoring the epigenomic state of DMVs colocalized with critical brain developmental genes. Hierarchical clustering of mCG at DMVs for 39 established regional marker genes grouped COs with mid-fetal and newborn (35 day) cortex (Figure 3A). Clustering analysis using all 619 H3K27me3+ DMVs revealed substantial similarity between COs, fetal and 35d cortex (Figure 3D). CO+/Ctx+ DMVs were strongly methylated in fetal and 35d cortex; CO-/Ctx+ DMVs were depleted for mCG in fetal cortex and showed intermediate levels of mCG in 35d Ctx. Overall, COs displayed DMV mCG signatures comparable with mid-fetal to newborn cortex at genes with established or predicted roles in brain development.

Pericentromeric repeats are demethylated in human neural cultures

Differentially methylated regions (DMRs) in CG context (CG-DMRs) are discrete genomic regions that are often associated with regulatory activity that is specific for tissues, cell types or developmental stages (Lister et al., 2013; Ziller et al., 2013; Schultz et al., 2015; Mo et al., 2015). We identified 50,099 and 30,574 CG-DMR regions showing decreased (hypo-DMRs) and increased mCG (hyper-DMRs) across CO differentiation, respectively. Analysis of the spatial distribution of CO hypo-DMRs across chromosomes revealed a number of regions that were highly enriched for hypo-DMRs, hereafter referred to as hypo-DMR blocks (Figure 4A and Figure S4E). We applied a two-step thresholding method and identified 140 hypo-DMRs blocks with a total length of 66 Mb (Figure 4A and Figure S4E, Table S3). With few exceptions, most hypo-DMRs blocks were located adjacent to a centromeric region (Figure 4A and Figure S4E.). The centromeres of all chromosomes were associated with at least one hypo-DMR block (Figure S4E). Consistent with their pericentromeric locations, centromere satellite sequences were 21.17 times more enriched in hypo-DMR blocks. LTR/ERVK and several different species of satellite sequences were also enriched in hypo-DMR blocks by more than two fold over the permutated background (Figure 4B).

To ask whether the demethylation of centromeric sequences is a characteristic of embryonic brain or is specific to COs, we analyzed additional methylomes including germinal matrix from 20wk PCW fetus and neurospheres (Gu et al., 2016). Hypo-DMR blocks were less methylated in all *in vitro* cultured samples including COs and neurospheres (Ns) derived from either cortex or ganglionic eminence (Figure 4A and C). For most tissues, including developing and mature cortex, the pericentromeric hypo-DMR block regions were not demethylated and showed mCG levels between 0.6 and 0.8, except for in pancreas and one of the fat tissues (Figure 4C and Figure S4B); pancreas and fat tissues have been shown to contain pronounced partially methylated domains (PMDs) (Schultz et al., 2015). Hypo-DMR block regions were also demethylated in immortalized cell lines GM12878, HepG2 (Figure S4C top panel). However, the phenomenon of hypo-DMR block was specific for *in vitro* neural cultures since COs and neurospheres had normal global mCG levels (~0.8) whereas GM12878 and HepG2 genomes were aberrantly lowly methylated (Figure S4C bottom panel). Furthermore, we examined methylomes of multiple hESC and iPSC lines which revealed that hypo-DMR blocks were not drastically demethylated in any of the pluripotent cell lines (Figure S4D). Therefore the demethylation of pericentromeric regions did not originate from the culture of pluripotent cells, but rather was induced during neural differentiation. Considering their vastly different cell origins of COs (hESC-derived) and neurospheres (primary cells), these data indicates that demethylation of pericentromeric regions is likely associated with the *in vitro* nature of the neural cultures.

Previous studies have demonstrated an association between large hypo-methylated domains found in *in vitro* cultured cells with repressive chromatin marked by H3K9me3 or H3K27me3, and proposed a passive demethylation mechanism biased towards late-replicating regions located near lamina-associated regions (Hon et al., 2012). To test whether such a mechanism might explain the formation of hypo-DMR blocks in COs, we compared the H3K9me3 level of hypo-DMR blocks between neurosphere cultures and fetal brain (Figure 4F and Figure S4F). Surprisingly, a substantial fraction of hypo-DMR blocks were associated with strongly reduced H3K9me3 levels in neurospheres, suggesting potential decompaction of heterochromatin in these regions. For example, decreased mCG and H3K9me3 levels were observed in pericentromeric regions of chromosome 11 (Figure 4H). We did not observe a widespread increase in the euchromatic mark H3K4me1 in hypo-DMR blocks (Figure 4G and Figure S4G). Thus the decrease in mCG and H3K9me3 in a subset of hypo-DMR blocks did not lead to active chromatin.

Transcript abundance of repeat elements in hypo-DMR blocks was next analyzed to address whether demethylation was associated with transcriptional alterations. Since a large amount of repeats were differentially expressed during early neural lineage commitment between hESC and EB, and pericentromeric demethylation was most pronounced in CO 40d and CO 60d, we focused on repeat elements that were differentially expressed between COs and EBs. We found a small number of eight repeat elements that were differentially expressed by more than two-fold between CO 60d and EB stages. Seven of the eight elements were located in a 786 kb region downstream of the chromosome 1 centromere; all elements showed an increase in expression (Figure 4D). These seven elements included four L1 LINE, one LTR/ERV1, one LTR/ERVL-MaLR and two satellite sequences (Figure 4E).

Cerebral organoids recapitulate the epigenomic landscape of human fetal brain

After excluding DMRs that overlapped with pericentromeric hypo-DMR blocks, 35,627 hypo-DMRs and 29,960 hyper-DMRs were identified during CO differentiation. mCG remodeling was mostly unidirectional during CO differentiation, with only 30 regions showing the reverse direction of mCG change. After grouping DMRs by their dynamics through CO differentiation (Figure 5B), further analyses were focused on nine DMR groups containing greater than 1,000 region, spanning 12.3 Mb. DMR groups 5, 9, 10 and 11 gained mCG whereas groups 17, 18, 19, 22 and 24 lost mCG during CO differentiation with distinct kinetics. Notably, COs and fetal cortex formed a distinct cluster that was separated from hESC, EB and NPC. Most CO hypo-DMRs (groups 17–24) were associated with depleted mCG in fetal cortex, and CO hyper-DMRs (groups 5–11) were associated with greater mCG levels in fetal cortex compared to hESCs (Figure 5B). DMRs surrounding the Emx1 locus provide examples of several DMR groups (Figure 5A). All hypo-DMRs (a, c, e and f) were depleted of mCG in fetal cortex. The lowly methylated state of DMR (a) was preserved in postnatal development, whereas DMRs (c, e and f) were partially or fully re-methylated in 35d and more developed cortex samples (Figure 5A). For hyper-DMRs (b and d), fetal and developed cortex showed fully methylated states that were similar to COs (Figure 5A).

To further determine the degree that CO differentiation mimics the epigenomic remodeling of brain development, chromatin accessibility and histone modifications were examined for the DMR regions with DNase-seq and ChIP-seq data generated from fetal brain (Figure 5A and B). Previous observations have revealed that hypo-DMRs are strong indicators of regulatory elements (Stadler et al., 2011; Hon et al., 2013; Mo et al., 2015). Strikingly, most CO hypo-DMRs shown in Figure 5A overlapped with open chromatin (DNase-seq peaks) and H3K4me1 enrichment in fetal brain, whereas hyper-DMRs were associated with elevated chromatin accessibility in hESCs. CO hyper-DMRs overlapped with open chromatin regions in hESC H9 and H1, becoming less accessible in fetal brain (Figure 5B). CO hypo-DMRs were predominantly depleted of mCG in fetal cortex and were associated with increased chromatin accessibility and enhancer chromatin mark H3K4me1 in fetal brain (Figure 5B). Quantification of the overlaps showed 67% of CO hyper-DMRs overlapped with hESC open chromatin; 48% and 69% of CO hypo-DMRs overlapped with fetal brain open chromatin and lowly methylated regions (UMRs+LMRs or ULMRs), respectively (Figure 5C). ULMRs have previously been shown to be robust markers for regulatory sequences in other systems (Stadler et al., 2011; Burger et al., 2013). If we defined fetal brain regulatory elements as the union of open chromatin regions and ULMRs in the fetal brain, 74% of CO hypo-DMRs overlapped with fetal brain regulatory sequences. These findings suggest DMRs in COs predict regulatory regions in the human fetal brain *in vivo*, pointing to robust recapitulation of the epigenomic signature.

In addition to the correlated regulatory sequence signatures found between COs and human fetal brain, we found a significant fraction of regulatory sequences, identified in fetal cortex, were absent in COs (Figure S5C). Only 32.5% of fetal cortex LMRs were also identified as lowly methylated in 40d or 60d CO (Figure S5C). Absence of a fraction of fetal cortex LMRs in COs can be explained by the intermediate mCG level of regulatory elements in COs, whereas the regions were generally strongly depleted of mCG in fetal cortex (Figure

S5D–H). A greater level of mC accumulation in COs than in fetal cortex can be observed at both promoter regions (indicated by UMRs, Figure S5G and H), and distal elements or enhancers (indicated by LMRs, Figure S5D). Interestingly, regulatory elements were strongly depleted of mCG in neurospheres to an extent similar to that of fetal cortex (Figure S5J–K). Therefore, *in vitro* culturing does not intrinsically inhibit DNA demethylation. Instead, cell type heterogeneity or sub-optimal demethylation machinery function may be responsible for the intermediate mCG level at regulatory elements in COs.

Overlapping CO hypo-DMRs with ULMRs identified in fetal and developed cortex samples strongly suggests the epigenomic state of COs matched that of mid-fetal brain. As mentioned above, 69% of hypo-DMRs overlapped with fetal cortex ULMRs. 45% of hypo-DMRs overlapped with neonatal (35d) cortex ULMRs and less than 30% of overlaps with ULMRs were identified in more developed cortex (Figure S5I). CO hypo-DMRs were strongly depleted of mCG in fetal cortex and a substantial fraction become hyper-methylated in later stages of cortical development (Figure S5J).

Regulatory elements in cerebral organoids and fetal brain are enriched for common TF binding motifs

The enrichment of TF binding motifs was examined for each DMR group to identify candidate TFs whose binding may significantly shape the CO and fetal cortex regulatory landscapes. Differential TF enrichment was found between DMRs with distinct dynamics. For example, PAX6 and OTX1/2 motifs were strongly enriched in group 24 and 22 hypo-DMRs that became demethylated at the EB stage (Figure 5D), consistent with the expression of these two factors which peaked at the EB stage (Figure 5E). Hypo-DMRs that become demethylated in later stages of CO differentiation (group 17) were enriched for Nuclear factor I (NFIB and NFIX) motifs, paralleling the elevated expression of *Nfib* and *Nfix* in 40d and 60d COs (Figure 5E). NFI family TFs have been shown to regulate neural stem cell quiescence and are considered core factors of the neural stem cell regulatory network (Martynoga et al., 2013; Mateo et al., 2015).

The list of TF binding motifs that were enriched in fetal cortex specific LMRs were highly similar to that enriched in CO hypo-DMRs, suggesting CO differentiation and human brain development are driven by a common set of TFs. Consistent with the mid-fetal stage of the fetal cortex sample, fetal cortex specific LMRs were weakly enriched for PAX6 and OTX1/2 motifs, but were strongly enriched for NFI motifs. TF binding motifs that were enriched in fetal cortex but not COs included homeobox protein TGIF, and bHLH factors NEUROD2, NEUROG2 and BHLHE22. Intriguingly, none of these TFs were under-expressed in COs, suggesting other factors such as the accessibility of binding sites or binding partner requirements may be responsible for the less enriched binding signatures observed in COs.

Cerebral organoids share promoter DNA methylation signatures of human cortical development

To further dissect the contribution of proximal and distal elements in regulating neural differentiation, the locations of DMRs were mapped relative to transcription start sites (TSSs). The analysis revealed that hyper-DMRs that gradually gain mCG (groups 9, 10 and

11) during CO differentiation were strongly enriched at the TSS, whereas early hyper-methylated (group 5) hyper-DMRs were located farther from the TSS (Figure 6A). The two types of hyper-DMRs also showed distinct enrichments for TF binding motifs. The binding motifs of pluripotent factors such as NANOG, POU5F1(OCT4) and SOX2 were strongly enriched in group 5 hyper-DMRs (Figure 5D). TF binding motifs commonly found in GC rich promoters, such as EGR1 and SP1 motifs, were instead enriched in group 9, 10 and 11 hyper-DMRs (Figure 5D). In contrast to hyper-DMRs, all groups of hypo-DMRs were predominantly located farther from TSS (Figure 6A). These results suggest that distal hESC enhancers are quickly methylated upon commitment to the neuronal lineage, whereas hESC specific promoters accumulate mCG more slowly during CO differentiation. The different kinetics of enhancer and promoter mCG during differentiation could reflect the process of integrating regulatory inputs from distal regulatory sequences at the TSS.

Correlating relative gene expression with the changes in mCG around the TSS revealed a sharp inverse correlation for initial differentiation of hESC to EB (Figure 6B). The inverse correlation was predominantly contributed by hyper-mCG at the TSS for genes becoming repressed in EBs, compared to hESC (Figure S6A). No apparent mCG remodeling was observed for genes upregulated in EBs (Figure S6A and B). The correlation between TSS hyper-mCG and gene down-regulation could also be observed in the later stages of CO differentiation but was much less pronounced (Figure 6B and Figure S6C and D).

In order to assess whether promoter mCG patterns established in COs recapitulate that in fetal brain, we first analyzed the more pervasive promoter hyper-mCG. Active TSSs in hESCs were first identified by reference guided transcript assembly and H3K4me3 peak signatures. We then identified ULMRs surrounding TSSs to define the boundary of regions exhibiting hyper-mCG during CO differentiation (Burger et al., 2013). 95.7% of annotated TSSs in hESC were associated with an ULMR. TSSs that were associated with ULMR and at least one hyper-DMR were K-means clustered according to their mCG dynamics during CO differentiation (Figure 6D). Five promoter clusters were identified to be associated with apparent hyper-mCG. Hyper-mCG was observed at the upstream boundary of promoter ULMR in clusters 1 and 3, and at the downstream boundary of promoter ULMR in clusters 2 and 4 (Figure 6D). mCG was accumulated across the promoter ULMR region for cluster 5. mCG remodeling of clusters 1–4 promoters during CO differentiation generally recapitulated promoter mCG patterns in cortex samples (Figure 6C, D and Figure S6E). In addition, the extension of mCG into upstream and downstream boundaries of promoter UMRs were associated with a decrease in H3K4me3 signal for the corresponding regions in the fetal brain (Figure 6E). The presence of matching signatures of mCG and histone modifications further supported the conclusion that the promoter epigenomic state of brain cortex was accurately modeled in COs. Promoters in cluster 5 accumulated greater levels of mCG than in the cortex of both fetal and adult brains (Figure 6D). Together with the observation above that regulatory elements were associated with intermediate mCG levels in CO (Figure S5D–H), these results suggest that COs may be associated with an altered equilibrium of *de novo* DNA methylation and demethylation activities compared to the cortex.

The analysis of gene expression dynamics during CO differentiation confirmed the inverse correlation between promoter hyper-mCG and gene repression. Clusters 3, 4 and 5 showed more drastic gene repression during the initial differentiation to EB (Figure 6F and Figure S6J). The methylation of cluster 5 promoters was correlated with strong gene repression in EBs, COs as well as in the fetal and adult cortex (Figure 6G). Lastly, mCG remodeling at a small number of demethylated promoters also resulted in mCG patterns more similar to that in cortex (Figure S6F–I,K).

Discussion

hESC and iPSC derived differentiation systems are fundamental to the modeling of human diseases and regenerative therapies. The assessment of the quality of various *in vitro* differentiation models requires comprehensive histological, physiological, cellular and molecular characterization and comparison to their *in vivo* counterparts. The safety requirement of stem cell therapies further necessitates deeper understanding of the genomic and epigenomic stability of *in vitro* cultured cells. In this study, we have carried out detailed transcriptomic and epigenomic characterization of a three-dimensional model of human brain development – cerebral organoids.

Cerebral organoids model the transcriptomic dynamics of early-to-mid fetal cortical development

Integrated transcriptomic analysis of multiple *in vitro* neural differentiation methods revealed that COs and multi-well organoids were more strongly correlated with mid-fetal human cortex, compared to other culture systems examined (Figure 1D). Since the multi-well organoid approach (Qian et al., 2016) is a modification of the CO method initially developed (Lancaster et al., 2013), it is perhaps expected that COs and multi-well organoids showed a similar pattern of both PC1 and PC2 scores (Figure 1E).

The differential correlation of *in vitro* models to mid fetal cortex can be explained in part by the distinct expression patterns of ECM genes. The repression of ECM genes in COs and multi-well organoids may be a result of the embedding of neuroepithelial tissues within Matrigel droplets in organoid culture protocols (Lancaster et al., 2013; Qian et al., 2016). Since Matrigel is mainly composed of ECM proteins such as laminin, collagen and heparin sulfate proteoglycans, the embedding step may provide neural tissue with the needed ECM framework and prevent cells from attempting to restore the ECM structure by over-expressing the component genes. The impact of externally supplying tissues with ECM may extend beyond preventing ECM gene over-expression and also affect the patterning and maturation of tissues through modulating signaling and stress responses.

Non-CG methylation may mediate primed repression of fetal regulatory elements

Although human mid-fetal cortex is largely free of mCH (Lister et al., 2013), we were able to identify distinct patterns of mCH accumulation at fetal brain super-enhancers in both COs and fetal cortex. High mCH super-enhancers in fetal cortex are associated with postnatal downregulation of associated genes. The accumulation of mCH at fetal brain specific regulatory elements may serve as an anticipating epigenomic mark to guide mCH binding

proteins such as MECP2 in later developmental stages. MECP2 protein level was previously shown to increase dramatically during postnatal development of the mouse brain (Skene et al., 2010). Super-enhancer mCH may not incur transcriptional repression in the fetal brain due to the absence of binding proteins, but instead marks the regions for future binding as soon as MECP2 level increases during early postnatal development. Such anticipating epigenomic mechanism may facilitate brain maturation by terminating the activity of regulatory elements required for fetal development. It is also quite striking that these predictive patterns of mCH methylation were recapitulated in COs, further pointing to the fidelity of organoids for modeling human brain development.

DNA methylome sequencing reveals large-scale epigenomic remodeling in *in vitro* neural differentiation systems

The finding of pericentromeric hypo-DMR blocks in *in vitro* neuronal cultures including COs and neurospheres, stresses the importance of using genomic and epigenomic methods to evaluate how well iPSC-derived cells mimic authentic neural tissues, especially for therapeutic purposes. The demethylation of a total of 66 Mb of genomic regions was not generally associated with transcriptional alterations. However, the notable loss of H3K9me3 from a subset of hypo-DMR blocks suggests potential heterochromatin decompaction in these regions. Further analyses will be needed to determine whether the presence of hypo-DMR blocks impairs genomic and epigenomic stability of cultured cells. The finding of hypo-DMR blocks in neuronal cultures suggests the possibility of spontaneous and perhaps culture-condition-dependent epigenomic remodeling during extended periods of differentiation, which may impact on possible future therapeutic applications.

Epigenomic profiling suggests directions for the improvement of cerebral organoid methods

Transcriptomic and epigenomic characterization of COs also revealed limitations of the current CO method, and pointed to directions for the improvement of the culture method. As shown by the stronger overlap between 100 day multi-well organoids and fetal brain transcriptomes, extending the duration of CO culture while preventing necrotic cell death using bioengineering approaches would be a fruitful direction for generating COs that more closely resemble the fetal brain.

An apparent difference between the DNA methylome of COs and fetal cortex is that regulatory elements in CO generally accumulated a greater level of mCG. Interestingly, neurospheres derived from primary fetal brain tissues were strongly depleted of mCG at these regulatory regions. Further studies will be required to examine whether the intermediate mCG at CO regulatory elements is a result of cell type heterogeneity or an altered equilibrium of *de novo* methylation and demethylation activities. The application of single cell transcriptomic and epigenomic methods will enable further comparison of organoids to human brain tissues at a cell type-specific level, allowing the ability to discern variation in cell maturation from cell-type composition (Camp et al., 2016).

Lastly, molecular markers can be designed from hypo-DMR block regions and facilitate the screening of CO culture conditions that eliminate or reduce the pericentromeric demethylation, which may contribute to greater long-term genomic stability of CO culture.

Experimental Procedures

Further details can be found in the Supplemental Experimental Procedures.

Cerebral organoid differentiation

COs were generated following exactly the previously established protocol and using an orbital shaker for agitation (Lancaster et al. 2013; Lancaster and Knoblich 2014). For RNA-seq analysis, the following samples were collected from each of two independent batches: H9 hES day 0 (remaining cells from EB generation), 5 pooled EBs day 18, 3 pooled COs day 40, 3 pooled COs day 60. For DNA methylation analysis, the following samples were collected from each of two independent batches: H9 hES day 0 (remaining cells from EB generation), 3 pooled EBs day 16, 3 pooled COs day 40, 3 pooled COs day 60.

Methyl-Seq and RNA-seq library preparation

A detailed protocol for MethylC-seq libraries preparation can be found in Urich et al., 2015. MethylC-seq libraries were sequenced on Illumina HiSeq 2500 with 100 or 130 (fetal cortex sample) bases single-ended reads. Stranded RNA-seq libraries were constructed with Truseq Stranded mRNA LT kit (Illumina, RS-122-2101 and RS-122-2102) and were sequenced on Illumina HiSeq 2500 with 100 bases single-ended reads.

Accession Numbers

Raw and processed data files for this study have been deposited to NCBI GEO accession GSE82022, and data files are displayed via an Annoj browser (http://neomorph.salk.edu/cerebral_organoids_methylome.php).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. M. Margarita Behrens for providing the human fetal cortex tissue (UMB_412); Dr. Carol S. Huang for critical reading of the manuscript; Dr. Eran A. Mukamel and Yupeng He for comments and suggestions; Angela Peer for technical assistance. The work in J.R.E's laboratory was supported by the Howard Hughes Medical Institute and Gordon and Betty Moore Foundation (GBMF3034). The work in J.A.K.'s laboratory was supported by the Austrian Academy of Sciences, the Austrian Science Fund (grants I_1281-B19 and Z_153_B09), and an advanced grant from the European Research Council (ERC). M.A.L. was supported by a Marie Curie post-doctoral fellowship, and by the Medical Research Council MC_UP_1201/9.

References

Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. Nat Rev Neurosci. 2002; 3:517–530. [PubMed: 12094208]

Burger L, Gaidatzis D, Schübeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. 2013; 41:e155. [PubMed: 23828043]

Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Bräuninger M, Lewitus E, Sykes A, Hevers W, Lancaster M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A. 2015; 112:15672–15677. [PubMed: 26644564]

Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009; 27:275–280. [PubMed: 19252484]

Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W, Zoghbi HY. MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc Natl Acad Sci U S A. 2015; 112:5509–5514. [PubMed: 25870282]

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, et al. Genome- wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. 2006; 16:123–131. [PubMed: 16344561]

Danjo T, Eiraku M, Muguruma K, Watanabe K, Kawada M, Yanagawa Y, Rubenstein JL, Sasai Y. Subregional specification of embryonic stem cell-derived ventral telencephalic tissues by timed and combinatory treatment with extrinsic signals. J Neurosci. 2011; 31:1913–1933.

Eiraku M, Watanabe K, Matsuo-Takasaki M, Kawada M, Yonemura S, Matsumura M, Wataya T, Nishiyama A, Muguruma K, Sasai Y. Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. Cell Stem Cell. 2008; 3:519–532. [PubMed: 18983967]

Eiraku M, Sasai Y. Self-formation of layered neural structures in three-dimensional culture of ES cells. Curr Opin Neurobiol. 2012; 22:768–777. [PubMed: 22405989]

Fatehullah A, Tan SH, Barker N. Organoids as an in vitro model of human development and disease. Nat Cell Biol. 2016; 18:246–254. [PubMed: 26911908]

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature. 2015; 522:89–93. [PubMed: 25762136]

Gu J, Stevens M, Xing X, Li D, Zhang B, Payton JE, Oltz EM, Jarvis JN, Jiang K, Cicero T, et al. Mapping of Variable DNA Methylation Across Multiple Cell Types Defines a Dynamic Regulatory Landscape of the Human Genome. G3 (Bethesda). 2016; 6:973–986. [PubMed: 26888867]

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci. 2014; 17:215–222. [PubMed: 24362762]

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. 2011; 39:311–318.

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; 22:246–258. [PubMed: 22156296]

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet. 2013; 45:1198–1206. [PubMed: 23995138]

Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, Zhang X, Chavez L, Wang H, Hannah R, et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat Genet. 2014; 46:17–23. [PubMed: 24270360]

Johnson JZ, Hockemeyer D. Human stem cell-based disease modeling: prospects and challenges. Curr Opin Cell Biol. 2015; 37:84–90. [PubMed: 26546888]

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011; 478:483–489. [PubMed: 22031440]

Kirkeby A, Grealish S, Wolf DA, Nelander J, Wood J, Lundblad M, Lindvall O, Parmar M. Generation of regionally specified neural progenitors and functional neurons from human embryonic stem cells under defined conditions. Cell Rep. 2012; 1:703–714. [PubMed: 22813745]

Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA. Cerebral organoids model human brain development and microcephaly. Nature. 2013; 501:373–379. [PubMed: 23995685]

Lancaster MA, Knoblich JA. Organogenesis in a dish: modeling development and disease using organoid technologies. Science. 2014; 345:1247125. [PubMed: 25035496]

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. [PubMed: 19114008]

Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011; 7:e1001057. [PubMed: 21283776]

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; 462:315–322. [PubMed: 19829295]

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011; 471:68–73. [PubMed: 21289626]

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013; 342:1237905.

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, et al. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat Neurosci. 2013; 16:898–902. [PubMed: 23770565]

Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, Hariharan M, Nery JR, Castanon R, Sabatini K, et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature. 2014; 511:177–183. [PubMed: 25008523]

Mariani J, Simonini MV, Palejev D, Tomasini L, Coppola G, Szekely AM, Horvath TL, Vaccarino FM. Modeling human cortical development in vitro using induced pluripotent stem cells. Proc Natl Acad Sci U S A. 2012; 109:12770–12775. [PubMed: 22761314]

Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, Amenduni M, Szekely A, Palejev D, Wilson M, et al. FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. Cell. 2015; 162:375–390. [PubMed: 26186191]

Martynoga B, Mateo JL, Zhou B, Andersen J, Achimastou A, Urbán N, van den Berg D, Georgopoulou D, Hadjur S, Wittbrodt J, et al. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. Genes Dev. 2013; 27:1769–1786. [PubMed: 23964093]

Mateo JL, van den Berg DL, Haeussler M, Drechsel D, Gaber ZB, Castro DS, Robson P, Crawford GE, Flicek P, et al. Characterization of the neural stem cell gene regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal. Genome Res. 2015; 25:41–56. [PubMed: 25294244]

Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, et al. Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron. 2015; 86:1369–1384. [PubMed: 26087164]

Mo A, Luo C, Davis FP, Mukamel EA, Henry GL, Nery JR, Urich MA, Picard S, Lister R, Eddy SR, et al. Epigenomic landscapes of retinal rods and cones. Elife. 2016; 5:e11613. [PubMed: 26949250]

Pastor WA, Chen D, Liu W, Kim R, Sahakyan A, Lukianchikov A, Plath K, Jacobsen SE, Clark AT. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. Cell Stem Cell. 2016; 18:323–329. [PubMed: 26853856]

Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, Yao B, Hamersky GR, Jacob F, Zhong C, et al. Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. Cell. 2016; doi: 10.1016/j.cell.2016.04.032

Roadmap Epigenomics Consortium. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015; 523:212–216. [PubMed: 26030523]

Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell. 2010; 37:457–468. [PubMed: 20188665]

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]

Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, Pontis J, Wang H, Iouranova A, Imbeault M, et al. Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell. 2016; pii: S1934-5909(16)30161-8. doi: 10.1016/j.stem.2016.06.011

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat Protoc. 2015; 10:475–483. [PubMed: 25692984]

van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, Lee C, Martinez RA, Levi BP, Thompson CL, et al. CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron. 2014; 83:51–68. [PubMed: 24991954]

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]

Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153:1134–1148. [PubMed: 23664764]

Yin X, Mead BE, Safaee H, Langer R, Karp JM, Levy O. Engineering Stem Cell Organoids. Cell Stem Cell. 2016; 18:25–38. [PubMed: 26748754]

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013; 500:477–481. [PubMed: 23925113]

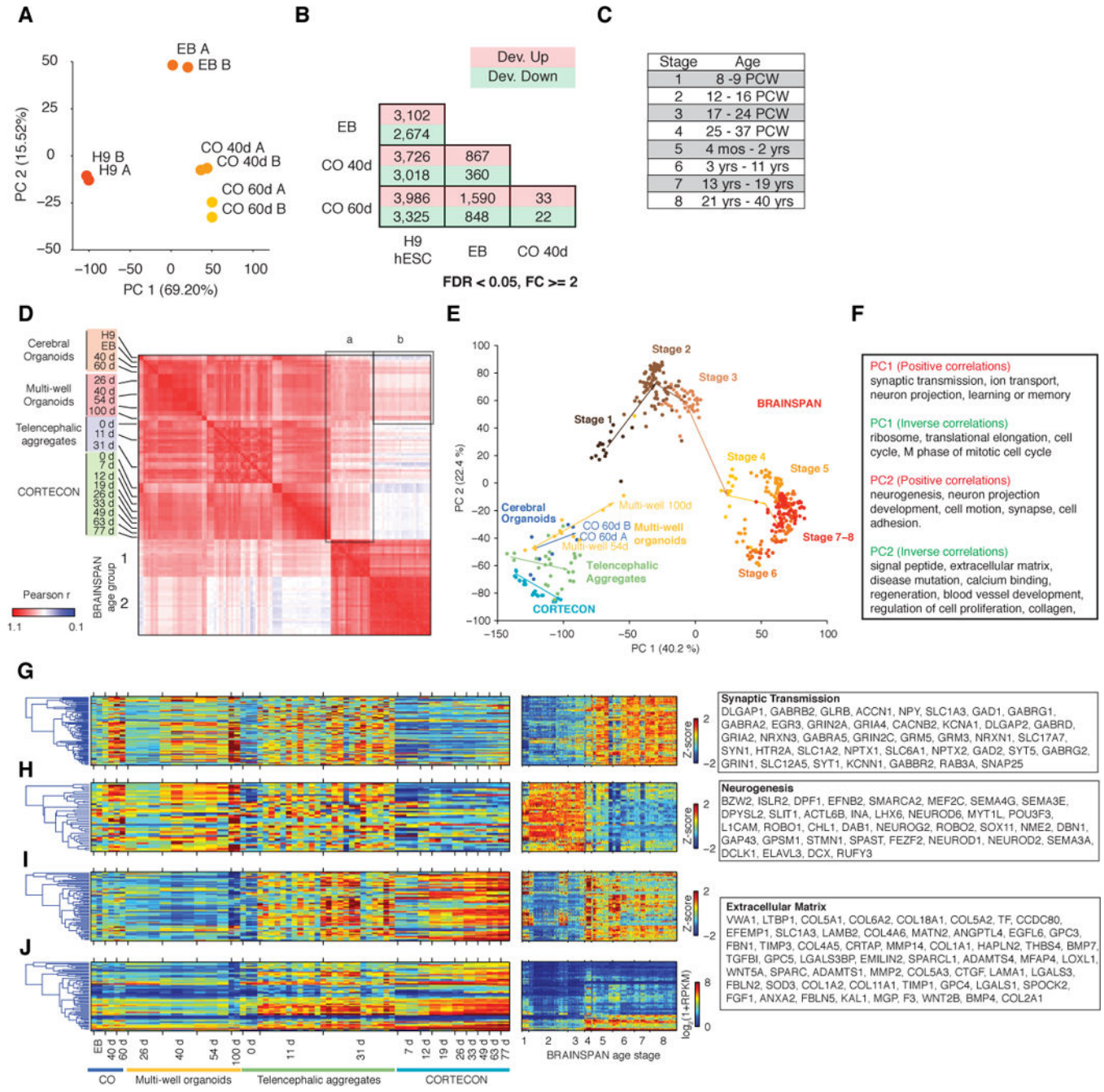

Figure 1. Transcriptional dynamics during cerebral organoid differentiation

(A) Principal component analysis (PCA) of gene expression during cerebral organoid differentiation. **(B)** The number of differentially expressed genes across CO differentiation with FDR < 0.05 and FC (fold change) >= 2. **(C)** The definition of human brain development stages used in this study. **(D)** Pearson correlation of the expression (RPKM) of protein coding genes across various *in vitro* neural differentiation systems and human cortical development (BRAINSPAN). **(E)** PCA of the gene expression of *in vitro* differentiation systems and human cortical development. **(F)** Enriched gene ontology terms of genes with top 1,000 positive or minus loading for PC1 and PC2 in (E). **(G)** The relative

expression (Z-score) of genes with synaptic transmission functions that were associated with top 1,000 positive loadings for PC1 in (E). **(H)** Z-score of neurogenesis genes that were associated with top 1,000 positive loadings for PC2 in (E). **(I)** Z-score of ECM genes that were associated with top 1,000 negative loadings for PC2 in (E). **(J)** The expression ($\log_2$(1+RPKM)) of ECM genes shown in (I).

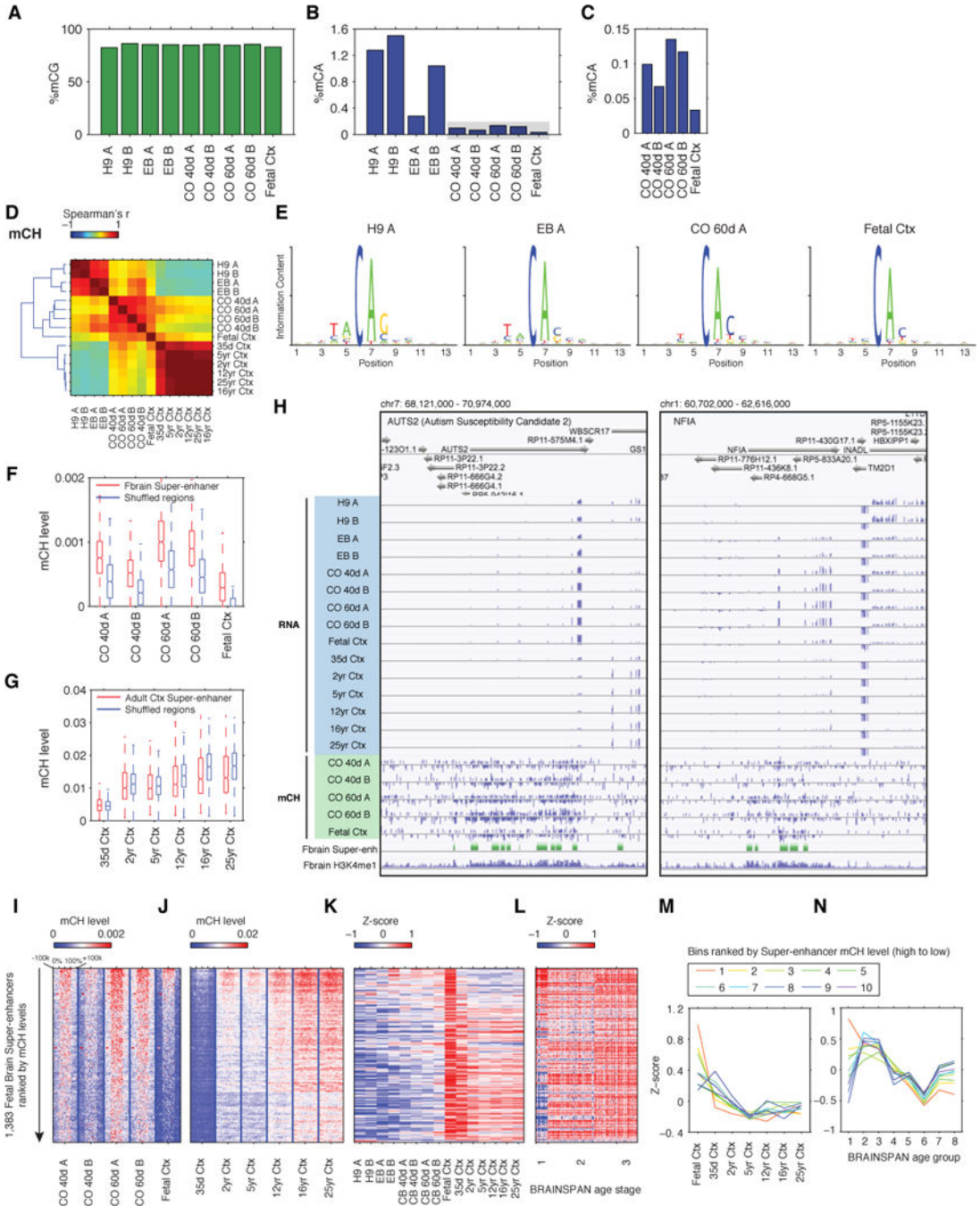

Figure 2. Cerebral organoids preserve fetal cortex non-CG DNA methylation patterns

(A–C) Genome-wide levels of DNA methylation in CG (A) and CA (B) contexts. mCA levels of CO and fetal cortex samples were plotted with a reduced scale in (C). **(D)** Hierarchical clustering of the pairwise spearman correlation matrix of mCH levels in 100 kb bins. **(E)** Sequence contexts of statistically significant mCH sites. **(F)** mCH levels of putative super-enhancers identified in fetal brain (Fbrain) and shuffled regions. **(G)** mCH levels of putative super-enhancers identified in adult cortex and shuffled regions. **(H)** The enrichment of mCH at putative super-enhancers overlapping with *Auts2* and *Nfia* loci. The scale of each

track was individually selected for optimal display. Ticks indicate statistically significant mCH sites (binomial tests, FDR<0.01). **(I-J)** Putative fetal brain super-enhancers were ranked by their mCH levels in the fetal cortex. The mCH levels of fetal-brain super-enhancers in postnatal cortical development were plotted (I). **(K-L)** Zscore of genes associated with fetal brain super-enhancers. **(M-N)** Z-score of genes associated with bins of super-enhancers ranked by their mCH levels in fetal cortex.

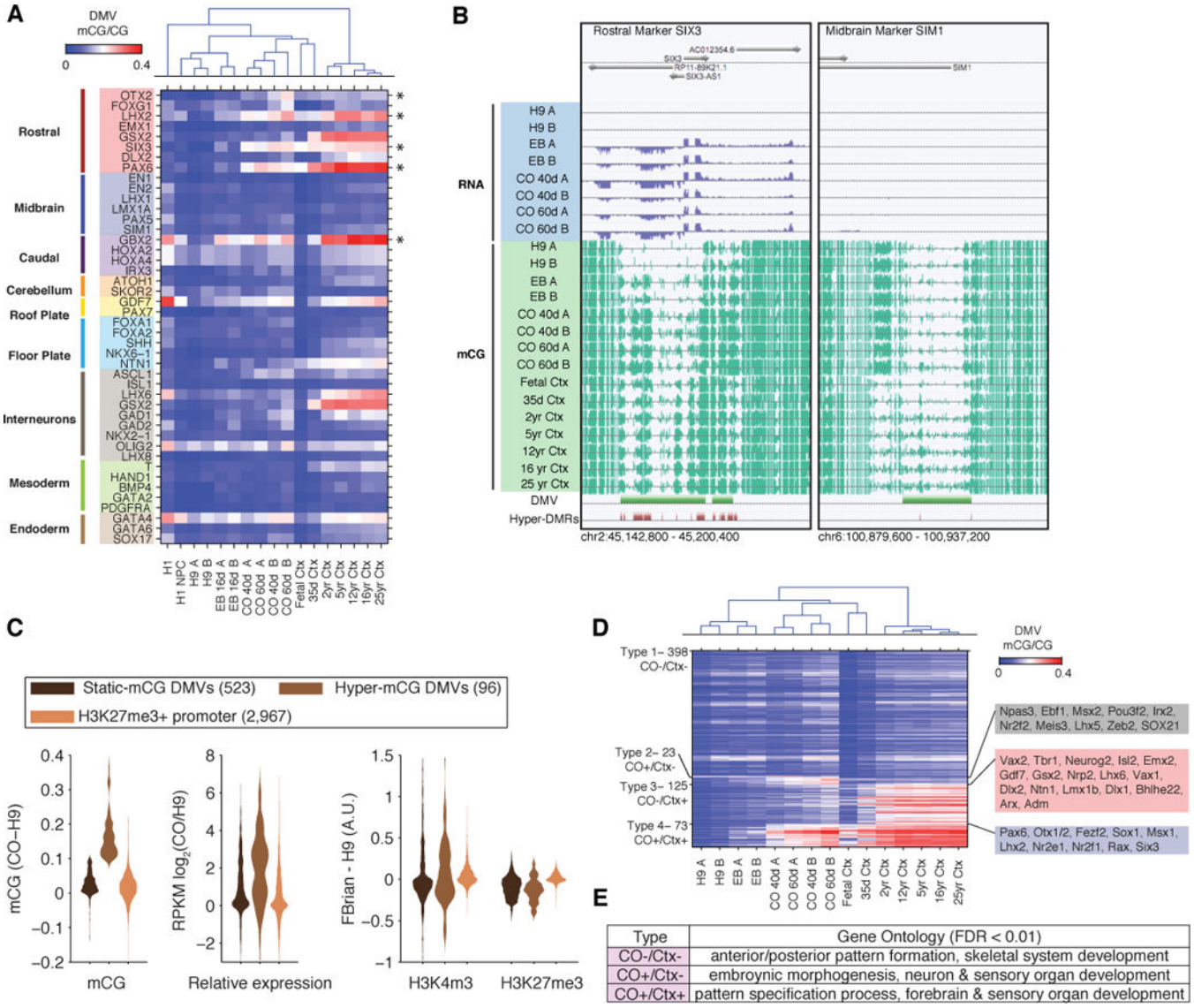

Figure 3. Brain region specific regulator genes are associated with distinct hyper-methylation signatures

(A) The average CG methylation level of DMVs colocalized with brain regional marker genes. **(B)** Browser views showing mCG remodeling at rostral maker *Six3* gene (left panel) and midbrain marker *Sim1* gene (right panel) during CO differentiation and cortical development. **(C)** The distribution of mCG, relative gene expression, H3K4me3 and H3K27me3 levels for static, hyper-mCG and all H3K27me3+ DMVs. **(D)** H3K27me3+ DMVs were categorized by hyper-mCG accumulation in COs during cortical development. For example, Type 3 DMVs were hyper-methylated in cortex but not in COs (CO-/Ctx+). **(E)** Gene ontology term enrichments for the genes associated with each type of DMV.

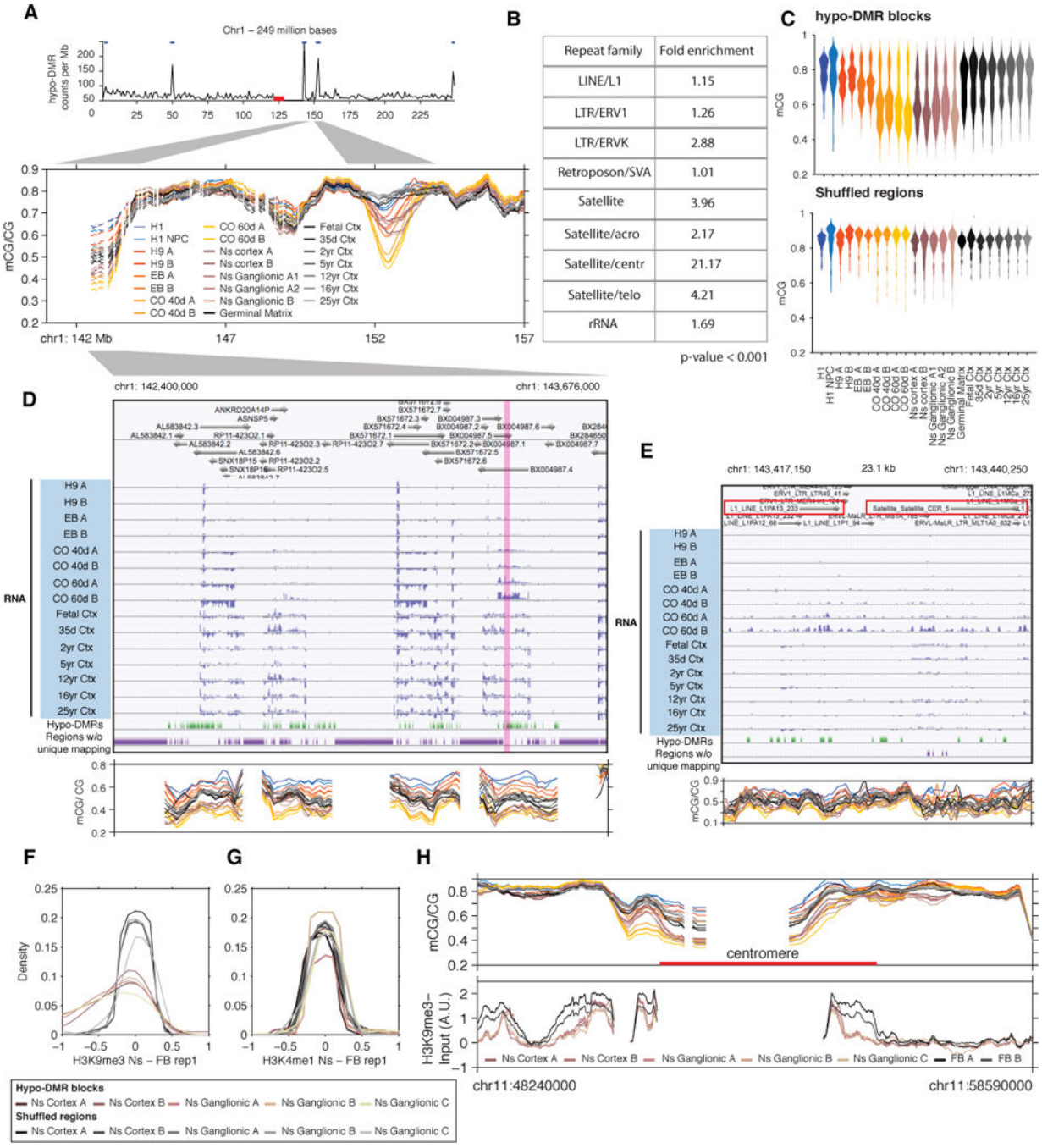

Figure 4. Enriched demethylation of pericentromeric repeats in cerebral organoids

(A) The top panel shows the number of hypo-DMRs per Mb window across chromosome 1. Red and blue bars indicate the locations of centromere and hypo-DMR blocks, respectively. The bottom panel shows mCG levels for COs, neurospheres (Ns) and cortex samples for a pericentromeric region chr1: 137 – 157 Mb. **(B)** Repeat families enriched in hypo-DMR blocks identified by relative enrichment compared to shuffled genomic regions repeated 1,000 times. **(C)** The distribution of mCG levels for all hypo-DMR blocks and shuffled regions. **(D and E)** Multiple regions showing increased transcripts abundances during CO

differentiation in a pericentromeric hypo-DMR block region (chr1: 142,400,000 – 143,676,000) of chromosome 1. The region with pink shade is expanded in (E). Two annotated repeat elements that were upregulated in CO 60d are highlighted by red rectangles. **(F, G)** The comparison of H3K9me3 (F) and H3K4me1 (G) levels between neurosphere cultures and the fetal brain (FB, replicate 1), at hypo-DMR blocks and shuffled regions. **(H)** Reduced mCG and H3K9me3 levels at a pericentromeric region of chromosome 11.

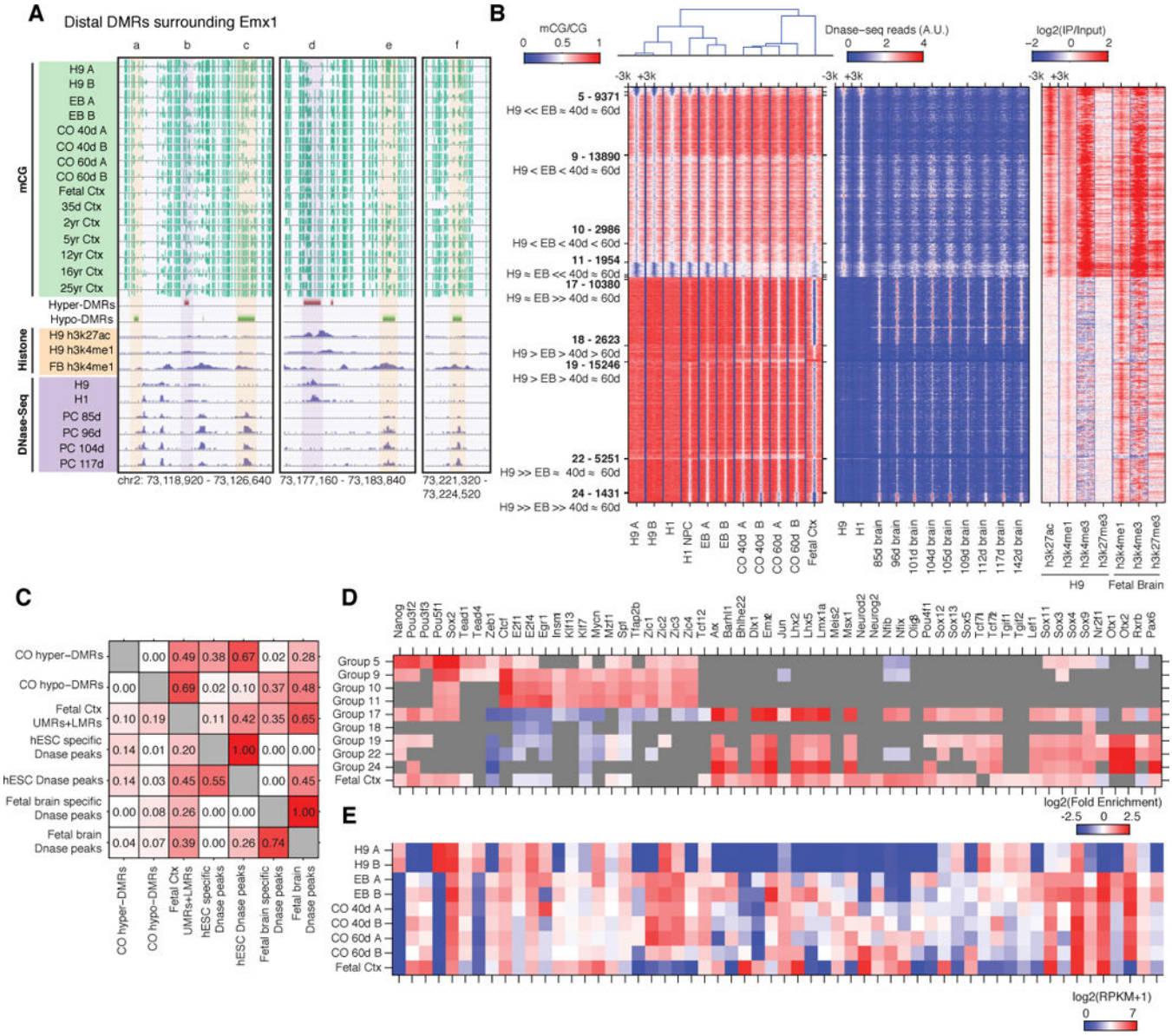

Figure 5. Epigenomic remodeling of regulatory regions during cerebral organoid differentiation

(A) Browser views of DMRs in regions surrounding *Emx1* locus. **(B)** DMRs found in CO differentiation grouped by the dynamics of mCG remodeling. Heatmaps show mCG, DNase-seq and ChIP-seq signals for regions +/- 3kb from the center of DMRs. **(C)** The overlap of DMRs with hESC and fetal brain regulatory features. The heatmap and numbers show the percentage of row features overlapping with column features. **(D and E)** TF binding motif enrichment in DMR groups and fetal cortex specific LMRs. The expression of TFs is shown in (E).

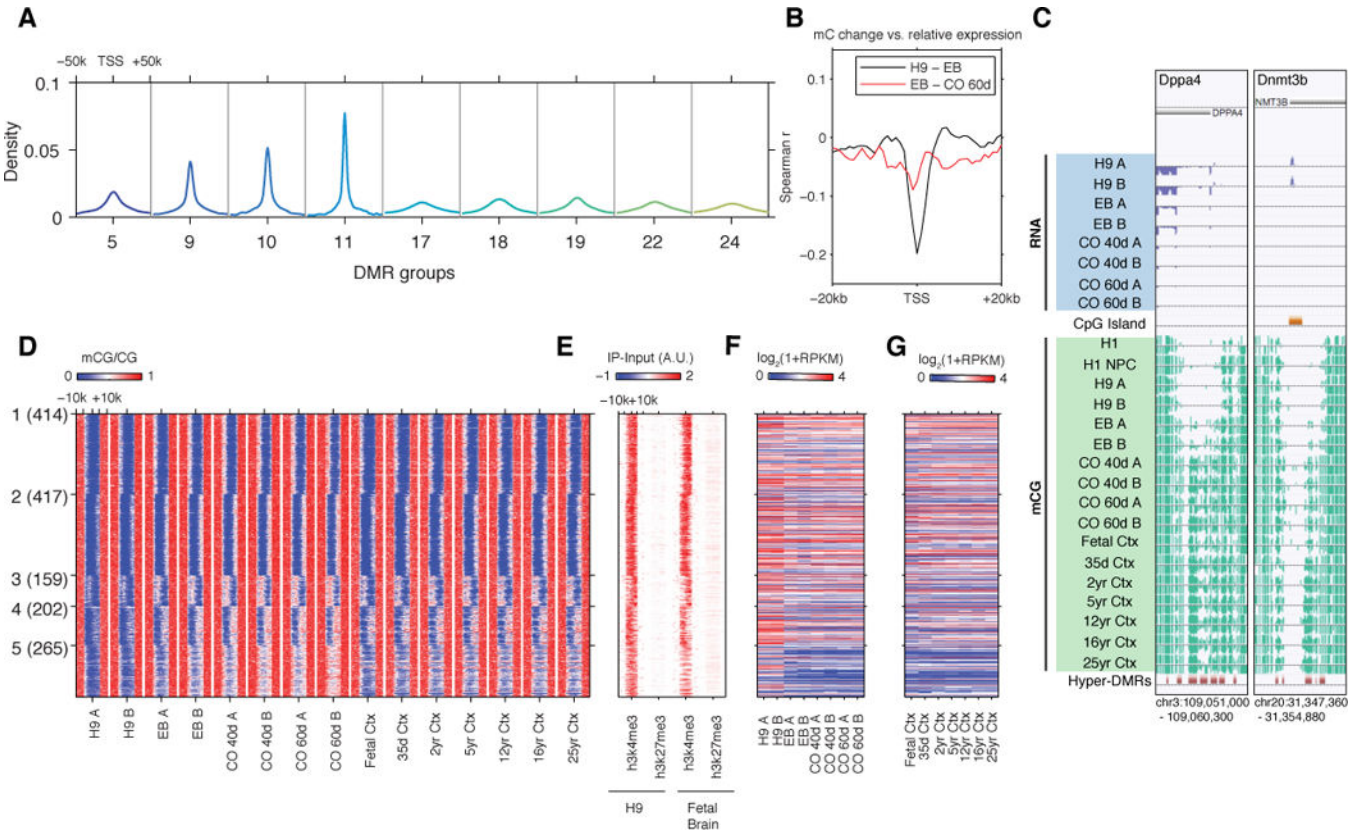

Figure 6. Cerebral organoids establish promoter DNA methylation patterns similar to human brain cortex

(A) The distribution of DMR groups in regions +/− 50kb from TSS. **(B)** Spearman correlation coefficients between mCG change and relative gene expression for the differentiation transitions between H9 – EB and EB – CO 60d. **(C)** Browers views of mCG remodeling at Dppa4 and Dnmt3b promoters during CO differentiation and human cortical development. **(D)** K-means clustering of mCG patterns at promoters that overlapped with TSS and CO hyper-DMRs. Clusters 1-5 with apparent hyper-methylation during CO differentiation are shown. **(E-G)** Histone modifications H3K4me3, H3K27me3, gene expression in CO differentiation and human cortex development were plotted for clustered promoters and associated genes in (D).