The Epigenetic Role of EGR1 during Postnatal Mammalian Brain Development

Zhixiong Sun

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biology

Hehuang Xie, Chair

Christopher B. Lawrence, Co-Chair

Liwu Li

Jinsong Zhu

Pawel Michalak

June 26, 2018

Blacksburg, Virginia

Keywords: DNA methylation, EGR1, TET1, bipolar DNA methylation, differentially methylated regions, bisulfite sequencing

The Epigenetic Role of EGR1 during Postnatal Mammalian Brain Development

Zhixiong Sun

ABSTRACT

DNA methylation is an epigenetic mechanism critical for tissue development, cell specification and cellular function. Mammalian brains consist of millions to billions of neurons and glial cells that can be subdivided into many distinct types of cells. We hypothesize that brain methylomes are heterogeneously methylated across different types of cells and the transcription factors play key roles in brain methylome programming.

To dissect brain methylome heterogeneity, in Chapter 2, we first focused on the identification of cell-subset specific methylated (CSM) loci which demonstrate bipolar DNA methylation pattern, i.e., hypermethylated in one cell subset but hypomethylated in others. With the genome-scale hairpin bisulfite sequencing approach, we demonstrated that the majority of CSM loci predicted likely resulted from the methylation differences among brain cells rather than from asymmetric DNA methylation between DNA double strands. Importantly, we found that putative CSM loci increased dramatically during early stages of brain development and were enriched for GWAS variants associated with neurological disorder-related diseases/traits. It suggests the important role of putative CSM loci during brain development, implying that dramatic changes in functions and complexities of the brain may be companied by a rapid change in epigenetic heterogeneity.

To explore epigenetic regulatory mechanisms during brain development, as described in Chapter 3, we adopted unbiased data-driven approaches to re-analyze methylomes for human and mouse frontal cortices at different developmental stages. We predicted *Egr1*, a transcriptional factor with important roles in neuron maturation, synaptic plasticity, long-term memory formation and learning, plays an essential role in brain epigenetic programming. We performed EGR1 ChIP-seq and validated that thousands of EGR1 binding sites are with cell-type specific methylation patterns established during postnatal frontal cortex development. More specifically, the CpG dinucleotides within these EGR1 binding sites become hypomethylated in mature neurons but remain heavily methylated in glia. We further demonstrated that EGR1 recruits a DNA

demethylase TET1 to remove the methylation marks at EGR1 binding sites and activate downstream genes. Also, we found that the frontal cortices from the knockout mice lacking *Egr1* or *Tet1* share strikingly similar profiles in both gene expression and DNA methylation. Collectively, the study in this dissertation reveals EGR1 programs the brain methylome together with TET1 during postnatal development. This study also provides new insights into how life experience and neuronal activity may shape the brain methylome.

The Epigenetic Role of EGR1 during Postnatal Mammalian Brain Development

Zhixiong Sun

GENERAL AUDIENCE ABSTRACT

DNA methylation is a widespread epigenetic mark on DNA, serving as a "switch" to turn on or off gene expression. It plays essential roles in cellular functions, tissue development. Mammalian brains contain millions to billions of neurons and glial cells, which can be further divided into many different types of cells. We hypothesize that brain cells have different methylation profiles across the genome, and transcriptional factors play important roles in programming methylation in the mammalian brain genome.

To study the diversity of methylation profiles across the genomes of different brain cells, in Chapter 2, we first focused on the identification of cell-subset specific methylated (CSM) genomic regions which show bipolar DNA methylation pattern, i.e., hypermethylated in one type of cell but hypomethylated in others. By applying a technique called the genome-scale hairpin bisulfite sequencing to mouse frontal cortices, we demonstrated that the majority of CSM genomic regions predicted likely resulted from the methylation differences among brain cells, rather than from methylation differences between DNA double strands. Surprisingly, we found that these predicted CSM genomic regions increased dramatically during early stages of brain development and were enriched for GWAS variants associated with neurological disorder-related diseases/traits. It suggests the importance of predicted CSM genomic regions, implying that dramatic changes in brain function and structure may be companied by a rapid change in DNA methylation diversity during brain development.

To explore underlying epigenetic mechanisms during brain development, as described in Chapter 3, we re-analyzed methylomes for human and mouse frontal cortices at different developmental stages, and predicted *Egr1*, a transcriptional factor with important roles in neuron maturation, synaptic plasticity, long-term memory formation and learning, plays an essential role in brain methylome programming. We found thousands of EGR1 binding sites showed cell-type specific methylation patterns, and were established during postnatal frontal cortex development. More

specifically, the methylation level of these EGR1 binding sites was low in mature neurons but pretty high in glial cells. We further demonstrated that EGR1 recruits a DNA demethylase TET1 to remove the methylation marks at EGR1 binding sites and activate downstream genes. Also, we found that the frontal cortices from the *Egr1* knockout or *Tet1* knockout mice show strikingly similar profiles in both gene expression and DNA methylation. Collectively, the study in this dissertation reveals EGR1 works together with TET1 to program the brain methylome during postnatal development. This study also provides new insights into how life experience and neuronal activity may shape the brain methylome.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisor, Dr. Hehuang (David) Xie, for his valuable guidance, support and confidence in me during my entire Ph.D. study. I, fortunately, joined Dr. Xie's lab at Biocomplexity Institute in May 2013 after one-year rotation. Dr. Xie is the one who introduced me into the exciting field of epigenetics and neuroscience. I greatly appreciate the training opportunities he provided, from which I benefit a lot. His intellectual rigor, enthusiasm for science and critical thinking greatly helped me to grow up as an independent researcher.

I would also like to thank the rest of my committee members: Dr. Christopher Lawrence, Dr. Liwu Li, Dr. Pawel Michalak and Dr. Jinsong Zhu for their insightful comments and advice in my experiments and project progress. I am very grateful for their support and help in my dissertation and job hunting. Special thanks go to my co-advisor, Dr. Christopher Lawrence, who has been giving me priceless advice on my career development, generously sharing his research experiences, offering help in my defense. I would also like to thank Dr. Jinsong Zhu, his postdoc Dr. Pengcheng Liu, and his Ph.D. student Xiaonan Fu for their generous help in job hunting and in my experiments.

I would like to thank our current and previous lab members. Special thanks go to Dr. Ming-an Sun, Xiguang Xu and Jianlin He for their generosity of time and help in bioinformatics analysis and experiments. I also want to thank Alex Murray, and Sharmi Banerjee for their help in experiments and bioinformatics analysis; it is my pleasure to work with them.

I would like to thank Dr. Chang Lu, his Ph.D. students Dr. Yan Zhu, and Dr. Sai Ma for their help in ChIP-seq, introducing our lab the microfluidic technique to profile epigenomes at ultra-low input.

Last but not the least, I would like to thank my family: my wife, my parents and my parents-in-law who supported my academic pursuits, and helped at every stage of my personal life. Many thanks to my wife, Miao Chen, who always encouraged me in every possible way to let me go through hard times.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

5hmC	5-hydroxymethylcytosine
ASM	allele-specific DNA methylation
BP	biological process
CC	cellular component
CGIs	CpG islands
Co-IP	co-immunoprecipitation
CREB	cAMP response element binding protein
CSM	cell-subtype specific methylated
DEGs	differentially expressed genes
DMEM	Dulbecco's Modified Eagle's Medium
DMRs	differentially methylated regions
DMSs	differentially methylated sites
E11.5	embryonic day 11.5
EGR1	early growth response gene-1
Egr1KO	Egr1 knockout
ESCs	embryonic stem cells
FACS	fluorescence-activated cell sorting
FFPE	formalin fixed paraffin embedded
GEO	Gene Expression Omnibus
GO	Gene Ontology
GWAS	genome-wide association studies

HATs	histone acetyltransferases
HBSS	Hank's balanced salt solution
IDR	irreproducible discovery rate
INTACT	isolation of nuclei tagged in specific cell types
KO	knockout
LMRs	low-methylated regions
MBD	methyl-CpG binding domain
mC	methylated cytosine
mCH	non-CpG methylation
MF	molecular function
MS-qPCR	methylation-sensitive qPCR
NeuN	neuronal-specific antibody
NP	neuronal progenitors
pCSM	predicted cell-subtype specific methylated
RT	room temperature
sc-BS-seq	single-cell genome-wide bisulfite sequencing
SRA	Sequence Read Archive
SVZ	subventricular zones
TET	Ten-Eleven Translocation
Tet1KO	Tet1 knockout
TFs	transcriptional factors
TKO	triple knockout
TSS	transcription start site

UHRF1	ubiquitin-like plant homeodomain and RING finger domain 1
WGBS	whole-genome bisulfite sequencing
WT1	Wilms tumor 1

Chapter 1 – Overview

1.1 DNA METHYLATION AND GENE EXPRESSION

DNA methylation, referring to the addition of a methyl (-CH_3) group to the fifth position of cytosine, which was first reported by Johnson and Coghill and then further confirmed by Hotchkiss and Wyatt (Johnson and Coghill 1925, Hotchkiss 1948, Wyatt 1951), is a highly conversed covalent modification widely found in prokaryotes (e.g. bacteria) and eukaryotes including fungi, plants, and animals, such as the model organisms like *Neurospora*, *Arabidopsis*, Mouse (Lee, Zhai et al. 2010). As a well-studied epigenetic mark, it plays import roles in the regulation of gene expression and genome stability, such as genomic imprinting, X-chromosome inactivation and silencing of transposable elements (Jones 2012), and serves as a crucial player in a wide spectrum of biological processes, including cell proliferation, cellular differentiation, and development.

As a stable epigenetic mark, DNA methylation is faithfully propagated through cell divisions and serves as a form of cellular epigenetic memory, maintaining cell identity (Li and Zhang 2014, Bogdanovic and Lister 2017). Recognizing that DNA polymerase is incapable of copying the methylation pattern from the parental stand to the daughter stand by the hemi-conservative replication mechanism, researchers postulate the existence of a maintenance methyltransferase (Holliday and Pugh 1975, Riggs 1975). In mammals, DNA methylation pattern is maintained by the maintenance DNA methyltransferase DNMT1, which is preferentially recruited to hemimethylated DNA (referring to DNA methylate on one of the double strands) by its obligate partner, the ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) (Gruenbaum, Cedar et al. 1982, Bestor, Laudano et al. 1988, Li, Bestor et al. 1992, Bostick, Kim et al. 2007, Sharif, Muto et al. 2007). New DNA methylation patterns are established by *de novo* DNA methyltransferases DNMT3A and DNMT3B, and their catalytic activity can be stimulated by enzymatically inactive family member DNMT3L (Okano, Xie et al. 1998, Gowher, Liebert et al. 2005). However, the division of labor between maintenance DNA methyltransferase and *de novo* DNA methyltransferases is not absolute. Consistent with the previous study (Jeong, Liang et al. 2009), Li *et al.* demonstrated that DNMT3A/3B could methylate daughter stands, DNMT1 and DNMT3A/3B work complementarily and simultaneously to retain methylation patterns (Li,

Dai et al. 2015). More recently, a new *de novo* DNA methyltransferases DNMT3C was identified in mice to methylate and silence the evolutionarily young retrotransposons in the male germ line (Barau, Teissandier et al. 2016).

In contrast to the prevalent non-CpG methylation (CpH, H=A, T, C) in plants, DNA methylation in mammals occurs predominantly in the CpG context, displaying a bimodal distribution: hypermethylation and hypomethylation (He, Chen et al. 2011). In human, 60-80% of ~28 million CpGs among ~1 billion cytosines are methylated in somatic cells (Lister, Mukamel et al. 2013). Notably, in the vertebrate genome, CpG sites are not evenly distributed. Less than 10% of CpGs tends to cluster in CpG islands (CGIs) (referring to CpG-rich regions with 0.5-2kb in length) which generally remain unmethylated, and about half of all CGIs reside in approximately 70% of annotated gene promoters, which includes almost all housekeeping genes, some tissue-specific genes and developmental regulator genes (Deaton and Bird 2011). The other half that is termed "orphan CGIs" are located equally between intergenic and intragenic regions. Contrary to the general unmethylated patterns of promoter-associated CGIs, orphan CGIs exhibit variable methylation patterns with the prevalence of tissue-specific CpG island methylation, suggesting they may serve broader biological roles (Maunakea, Nagarajan et al. 2010). Cumulating evidence suggests orphan CGIs are actually involved in gene regulation, including serving as alternative promoters located upstream of genes or intragenically, regulating gene imprinting, functioning as miRNA promoters, and emerging role as a novel class of highly active enhancers (Maunakea, Nagarajan et al. 2010, Bell and Vertino 2017, Sarda, Das et al. 2017).

Although DNA methylation on cytosine (mC) does not change the Watson–Crick base pairing of C with guanine (G), it introduces methyl moiety into the major groove of the double helix where various DNA-binding proteins contact with. The introduction of this chemical modification in the major groove can attract or repel a variety of DNA-binding proteins, and then modulate the functional state of regulatory regions (Raiber, Hardisty et al. 2017). Generally, DNA methylation in promoters is associated with transcription repression. This repression effect is mediated by directly interfering the binding of transcriptional factors (TFs) to methylated DNA regions or by recruiting proteins that create a refractory chromatin conformation unfavorable for transcription, such as methyl-CpG binding domain proteins (Deaton and Bird 2011).

Many transcription factors recognize GC-rich sequence motifs containing CpG sites. Approximately 25% of characterized TF binding motifs contain at least one CpG site (Blattler and Farnham 2013). Although not all of the CpG-containing binding motifs have the CpG dinucleotides at the position critical for DNA-protein interactions, several TFs with these motifs cannot bind to their recognition motif when the CpG sites are methylated. A well-studied example is CTCF, which functions as an insulator that blocks the interaction between a promoter and an enhancer (Bell, West et al. 1999). CTCF binds to a site within the mouse imprinted IGF2-H19 locus, renders *Igf2* gene silent in maternally derived copy due to its binding that disrupts the enhancer-promoter interaction. By contrast, the methylated CpG-rich CTCF binding sites at the paternal locus prevents the binding of CTCF, and thereby the enhancer interacts with the promoter and activates *Igf2* expression (Bell and Felsenfeld 2000). Besides the disruption of DNA-protein interactions, DNA methylation also exerts its effect on TF binding *in vivo*. Domcke *et al*. found 7000 novel binding sites for NRF1 were created in mouse embryonic stem cells (ESCs) with Dnmt1/Dnmt3a/Dnmt3b triple knockout (TKO) compared to wild-type cells. Restoring *de novo* methyltransferase activity by switching culture medium can remethylate these novel binding sites and outcompetes NRF1 binding, demonstrating that DNA methylation at binding motifs can restrict genomic binding of DNA-methylation-sensitive TFs (Domcke, Bardet et al. 2015).

In fact, the influence between DNA methylation and TF binding is not unidirectional. DNA methylation can dictate the genomic binding of TFs, as mentioned for NRF1 binding. On the other hand, TF binding can also affect the DNA methylation status (Zhu, Wang et al. 2016). This notion is supported by accumulating studies that examine the effects of TF binding on local methylation. Using whole-genome bisulfite sequencing (WGBS) analysis of mouse ES cells and neuronal progenitors (NP), Stadler *et al.* identified low-methylated regions (LMRs) with average methylation of 30%, accounting for 4.1 % of all CpGs, and residing at CpG-poor distal regulatory regions. Intriguingly, the binding of transcription factors within these regions, such as CTCF and REST, are not affected by their methylation status, but rather is necessary and sufficient for LMR creation (Stadler, Murr et al. 2011).

Contrary to the blocking of TF binding on methylated DNA, the second mode of methylation-mediated repression involves the attraction of methyl-CpG binding domain (MBD) proteins. The MBD protein family comprises MBD1, MBD2, MBD3, MBD4 and MeCP2. All the MBD

proteins except MBD3 selectively bind to methylated DNA *via* a common methyl-binding domain, and play a role in epigenetic remodelling and methylation-dependent repression of transcription. Therefore, they are generally described as "reader" of DNA methylation (Hendrich and Bird 1998, Saito and Ishikawa 2002). Protein interaction studies have shown that MBD proteins could interact with chromatin remodelers (i.e. Mi-2), histone deacetylases (i.e. HDAC1, HDAC2), and histone methylases (i.e. SETDB1), guide them to methylated DNA, induce repressive changes to local chromatin, finally lead to transcriptional repression (Du, Luu et al. 2015). A recent report revealed that MBD2 is also associate with DNMT1 and DNMT3A. The depletion of MBD2 in prostate cancer cells causes significant genome-wide hypomethylation, concomitant loss of MBD binding and upregulation of target genes at promoters and enhancer regions. These results support a model that MBD2 can not only serve as a "reader" of the genome methylation, but also can "rewrite" the cancer methylome by recruiting DNMTs to methylate specific CpGs (Stirzaker, Song et al. 2017).

Given the uneven distribution of CpGs along the genome and the structural similarity of MBD proteins, exploring the detailed genomic distribution of individual MBD protein will shed light on the molecular mechanism underlying the cellular readout of DNA methylation. With the aid of protein tagging approaches, Baubec *et al*. profiled the genome-wide distribution of 5 MBD protein: MBD1, MBD2, MBD3, MBD4, and MeCP2. In contrast to the non-apparent enrichment of MBD3 to methylated DNA, other MBD proteins reveal conserved binding, occurs predominantly at genomic regions with relatively high methylation density, such as methylated CGIs, confirming that methylation density is a major determinant of MBD protein binding (Baubec, Ivanek et al. 2013).

1.2 ROLE OF EGR1 IN BRAIN DEVELOPMENT

EGR1 (Early growth response gene-1, also known as Zif268, NGFI-A, Krox-24) is an inducible transcriptional factor belonging to the EGR family of C2H2-type zinc finger proteins, which also contains EGR2, EGR3, EGR4 and WT1 (Wilms tumor 1) (Poirier, Cheval et al. 2008). Numerous studies have shown that EGR1 is involved in diverse cellular processes, such as cell growth (Fahmy, Dass et al. 2003), mitogenesis (Zeng, Sun et al. 2005), cell differentiation (Kubosaki, Tomaru et al. 2009) and synaptic activation (Mokin and Keifer 2005). As one member of the immediate early gene family, EGR1 can be rapidly and transiently induced by

diverse stimulants (growth factors, hormones, and neurotransmitters) (Kubosaki, Tomaru et al. 2009), and is extensively expressed among brain cells, including neurons, microglia, and astrocytes (Zhang, Chen et al. 2014). In rat brains, *Egr1* mRNA is found in the neocortex, primary olfactory bulb, entorhinal cortices, amygdaloid nuclei, nucleus accumbens, striatum, cerebellar cortex, and the hippocampus (Davis, Bozon et al. 2003). The re-analysis of mRNA-seq data (Lister, Mukamel et al. 2013) from frontal cortex shows that *Egr1* expression is largely absent in human and mouse embryonic brain, but rapidly increases during the period of 2 years for human and 2 weeks for mouse, with stable expression level thereafter, consistent with the developmental expression patterns of EGR1 obtained from Allen Brain Atlas.

In the mammalian brain, EGR1 plays an important role in neuron maturation (Veyrac, Gros et al. 2013), synaptic plasticity, long-term memory formation (Penke, Morice et al. 2014) and learning (Maroteaux, Valjent et al. 2014). *Egr1* can be rapidly induced in the hippocampus in learning, recalling of several forms of hippocampal-dependent memories (Davis, Bozon et al. 2003), and exposure to a water maze (Guzowski, Setlow et al. 2001) or contextual fear retrieval tasks (Hall, Thomas et al. 2001). Its inactivation severely compromises the formation of several types of memories (Bozon, Davis et al. 2003, Lee, Everitt et al. 2004). Egr1 knockout (KO) mice show impaired long-term memory formation in both spatial and non-spatial learning tasks, but not in the early phase of long-term potentiation or long-term depression (Wei, Xu et al. 2000, Jones, Errington et al. 2001). Also, antisense knockdown of *Egr-1* in the lateral amygdala impairs long-term memory formation but not affect acquisition or short-term memory formation (Maddox, Monsey et al. 2011).

EGR1 contains three conversed C2H2 zinc finger motifs that recognize 9-base-pair segments in GC rich regions with the consensus sequence in rat and mouse: 5'-GCGGGGGCG-3', in human: 5'-GCGTGGGCG-3' (Pavletich and Pabo 1991, Mora-Lopez, Pedreno-Horrillo et al. 2008). The existence of two CG dinucleotides in the consensus sequence suggests the potential effect of DNA methylation on the binding of EGR1 to its target sites, which is confirmed in a seminal study which demonstrated the increased methylation status of the promoter of glucocorticoid receptor caused by maternal care reduced the binding of EGR1 to the promoter, thus abrogated the transcriptional activity of EGR1, leading to the attenuation of glucocorticoid receptor gene expression (Weaver, Cervoni et al. 2004). Subsequent studies in human also demonstrated that

high DNA methylation level of the target site negatively affected the binding affinity of EGR1 (McGowan, Sasaki et al. 2009).

In fact, the two CpG sites in the consensus sequence 5'-GCGGGGGCG-3‘ have a differential effect on EGR1 binding. The methylation of cytosine within 5’CpG dinucleotides reduces EGR1 binding to a level comparable to the methylation at both CpG sites, in contrast, the methylation within 3’CpG dinucleotides only partially reduces its binding. However, the effect of the mutation within these two CpGs is different from their methylation effect. Compared with the complete abolishment of binding of EGR1 caused by mutation of the 3’CpG dinucleotides, the mutation of the cytosine within 5’CpG only slightly reduces the binding. These findings suggests that the 5’CpG dinucleotides function as a “DNA methylation sensor”, and 3’CpG is crucial for EGR1 binding to the consensus sequence (Weaver, D'Alessio et al. 2007). This study also demonstrated that EGR1 recruits cAMP response element binding protein (CREB) binding protein to the promoter and induces histone acetylation leading to an open chromatin structure on the promoter. Meanwhile, EGR1 induced 50% demethylation of methylated exon I7 GR promoter. These results suggest EGR1 might direct demethylation activity to specific promoters by recuiting HATs (histone acetyltransferases).

On the contrary to the general rule that DNA methylation attenuates the binding affinity of transcription factors, Hashimoto *et al*. (Hashimoto, Olanrewaju et al. 2014) showed both human EGR1 and WT1 exhibited approximate twofold higher binding affinity to methylated censuses sequence 5’-GCG(T/G)GGGCG-3’ compared with the un-methylated sequence *in vitro* experiments, and this higher binding is achieved by their flexibility in conformational adaptation. This discrepancy between the *in vitro* and *in vivo* data could be due to the differences of EGR1 binding property in rat and human, or due to competition with proteins that specifically bind methylated CpGs, such as MBD proteins (Zandarashvili, White et al. 2015). A recent report revealed that MeCP2 could block EGR1 binding with its methylated target sites, supporting the notion that EGR1 binding is under combinatorial control of epigenetic modifications and other factors in cellular context (Kemme, Marquez et al. 2017).

1.3 SUMMARY AND OVERALL EXPERIMENTAL DESIGNS

1.3.1 Summary

The importance of DNA methylation has been well-established for brain development, neuronal differentiation, synaptic plasticity, learning and memory (Tognini, Napoli et al. 2015). Mammalian brain comprises millions to billions of neurons and glia, with various functionally distinct subtypes (Azevedo, Carvalho et al. 2009, Poulin, Tasic et al. 2016). Remarkable variations in DNA methylation are well documented across brain regions, between neurons and glia, and among limited types of neurons (Kozlenkov, Roussos et al. 2014, Mo, Mukamel et al. 2015, Kessler, Van Baak et al. 2016, Kozlenkov, Wang et al. 2016). However, how extensive epigenetic heterogeneity among brain cells remain less understood.

In Chapter 2, we implemented a computational approach to infer cell-subtype specific methylated (CSM) loci from the methylomes of human and mouse frontal cortices at different developmental stages. With the aid of genome-wide hairpin bisulfite sequencing approach, we demonstrated that the majority of predicted CSM (pCSM) loci likely resulted from the varied methylation among brain cells rather than from asymmetric DNA methylation between DNA double strands. Importantly, we found the number of pCSM loci increased dramatically during early stages of mammalian brain development. Although the production and migration of neurons are largely completed at birth, the brain exhibits the most pronounced advances in behavioral abilities (e.g., motor, cognitive and perceptual abilities), undergoing rapid synaptogenesis during early postnatal brain development. This fact implies the dramatic changes in functions and complexities of the brain may be accompanied by the rapid change in epigenetic heterogeneity, which is under the control of epigenetic regulatory networks. This assumption prompted us to explore epigenetic regulatory mechanisms during brain development in Chapter 3.

Accumulating studies reported the widespread methylome reconfiguration occurring in the mammalian frontal cortices during brain development. However, the mechanism underlying this dynamic methylation change on a genome-wide scale during the early stage of postnatal brain development remains elusive (Lister, Mukamel et al. 2013). Several lines of evidence indicate that transcriptional factors can work with TET (ten-eleven translocation) proteins in mammals to activate their expression by demethylation (Zhu, Wang et al. 2016). This study shows that a significant number of EGR1 binding sites gradually lost methylation during early postnatal brain

development. The *in vitro* studies demonstrated human EGR1 exhibited similar high affinity for methylated recognition sequence (Hashimoto, Olanrewaju et al. 2014). The Chapter 3 provided key evidence to support the hypothesis **that EGR1 recruits TET proteins to shape the methylome during early postnatal brain development**.

1.3.2 Overall experimental design

Specific Aim 1 (Chapter 2). To examine the dynamics of the cell-subtype specific methylation during mammalian brain development.

Mammalian brains are composed of millions to billions of neurons and glia, which can be subdivided into diverse, functionally distinct cell subtypes. Cell diversity is essential for brain function, and the understanding of neural diversity at the epigenetic level will gain insight into the underlying regulatory mechanism for neural identity, plasticity, and function (Telese, Gamliel et al. 2013). To this end, a number of studies explored cell-subtype specific DNA methylation by adopting different techniques for the isolation of targeted cell population, which includes manual sorting, laser capture microdissection, fluorescence-activated cell sorting (FACS) and isolation of nuclei tagged in specific cell types (INTACT) (Mo, Mukamel et al. 2015). However, profiling DNA methylation landscapes of each cell type in the brain remains a big challenge, even with the emerging single-cell genome-wide bisulfite sequencing (sc-BS-seq) technique, which shows ~5% genome coverage (Guo, Zhu et al. 2013, Smallwood, Lee et al. 2014). Therefore, we developed a computational approach to infer CSM loci from the methylomes of unsorted brain tissues (refer to the workflow in Fig 1a in Chapter 2).

Previous studies showed the dynamic DNA methylation change during mammalian brain development (Lister, Mukamel et al. 2013, Tognini, Napoli et al. 2015). Based on this finding, we reasoned that epigenetic heterogeneity among brain cells increased during mammalian brain development. To test this hypothesis, we explored the dynamics of cell-subtype specific methylation during mammalian brain development.

The experiments are designed to: 1) Determine the methylation fidelity in mammalian brain; 2) Infer the CSM loci from the methylomes of human and mouse frontal cortices; 3) Investigate the dynamics of cell-subtype specific methylation during mammalian brain development.

Specific Aim 2 (Chapter 3). To explore the epigenetic mechanism underlying the methylation change of EGR1 binding sites during brain development.

EGR1 is an important transcriptional factor with well-established roles in neuron maturation, synaptic plasticity, long-term memory formation and learning (Veyrac, Gros et al. 2013, Penke, Morice et al. 2014). We found that a significant number of EGR1 binding sites gradually lose methylation during early postnatal brain development. The *in vitro* studies demonstrated human EGR1 exhibited similar high affinity for methylated recognition sequence (Hashimoto, Olanrewaju et al. 2014); overexpression of rat EGR1 in HEK293 could provoke active demethylation in the exonI7 glucocorticoid receptor promoter (Weaver, D'Alessio et al. 2007). Noticeably, a recent study demonstrated WT1, a member of the EGR family whose binding motif shows significant similarity to that of EGR1, could recruit TET2 to its target genes to activate their expression by demethylation (Wang, Xiao et al. 2015). Taken together, we hypothesized that, EGR1, which is activated by environmental stimuli during early brain development, recruits Tet proteins to activate or prime EGR1 binding sites by demethylation (refer to the proposed model in supplementary Fig. 20 in Chapter 3).

The experiments are designed to: 1) Identify EGR1 binding sites by ChIP-seq in mouse frontal cortices; 2) Examine the methylation dynamics of EGR1 binding sites during brain development; 3) Identify specific TET protein that physically interacts with Egr1; 4) Investigate epigenetic regulation of target genes mediated by TET1 and EGR1; 5) Explore the DNA methylation and gene expression profiles in frontal cortices of Egr1KO and Tet1KO mice.

1.4 REFERENCES

Azevedo, F. A., L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent and S. Herculano-Houzel (2009). "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain." J Comp Neurol **513**(5): 532-541.

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou and D. Bourc'his (2016). "The DNA methyltransferase DNMT3C protects male germ cells from transposon activity." Science **354**(6314): 909-912.

Baubec, T., R. Ivanek, F. Lienert and D. Schubeler (2013). "Methylation-dependent and -independent genomic targeting principles of the MBD protein family." Cell **153**(2): 480-492.

Bell, A. C. and G. Felsenfeld (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene." Nature **405**(6785): 482-485.

Bell, A. C., A. G. West and G. Felsenfeld (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." Cell **98**(3): 387-396.

Bell, J. S. K. and P. M. Vertino (2017). "Orphan CpG islands define a novel class of highly active enhancers." Epigenetics **12**(6): 449-464.

Bestor, T., A. Laudano, R. Mattaliano and V. Ingram (1988). "Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases." J Mol Biol **203**(4): 971-983.

Blattler, A. and P. J. Farnham (2013). "Cross-talk between Site-specific Transcription Factors and DNA Methylation States." Journal of Biological Chemistry **288**(48): 34287-34294.

Bogdanovic, O. and R. Lister (2017). "DNA methylation and the preservation of cell identity." Curr Opin Genet Dev **46**: 9-14.

Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan and S. E. Jacobsen (2007). "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." Science **317**(5845): 1760-1764.

Bozon, B., S. Davis and S. Laroche (2003). "A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval." Neuron **40**(4): 695-701.

Davis, S., B. Bozon and S. Laroche (2003). "How necessary is the activation of the immediate early gene zif268 in synaptic plasticity and learning?" Behav Brain Res **142**(1-2): 17-30.

Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes Dev **25**(10): 1010-1022.

Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes & Development **25**(10): 1010-1022.

Domcke, S., A. F. Bardet, P. Adrian Ginno, D. Hartl, L. Burger and D. Schübeler (2015). "Competition between DNA methylation and transcription factors determines binding of NRF1." Nature **528**(7583): 575-579.

Du, Q., P. L. Luu, C. Stirzaker and S. J. Clark (2015). "Methyl-CpG-binding domain proteins: readers of the epigenome." Epigenomics **7**(6): 1051-1073.

Fahmy, R. G., C. R. Dass, L. Q. Sun, C. N. Chesterman and L. M. Khachigian (2003). "Transcription factor Egr-1 supports FGF-dependent angiogenesis during neovascularization and tumor growth." Nat Med **9**(8): 1026-1032.

Gowher, H., K. Liebert, A. Hermann, G. Xu and A. Jeltsch (2005). "Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L." J Biol Chem **280**(14): 13341-13348.

Gruenbaum, Y., H. Cedar and A. Razin (1982). "Substrate and sequence specificity of a eukaryotic DNA methylase." Nature **295**(5850): 620-622.

Guo, H., P. Zhu, X. Wu, X. Li, L. Wen and F. Tang (2013). "Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing." Genome Res **23**(12): 2126-2135.

Guzowski, J. F., B. Setlow, E. K. Wagner and J. L. McGaugh (2001). "Experience-dependent gene expression in the rat hippocampus after spatial learning: a comparison of the immediate-early genes Arc, c-fos, and zif268." J Neurosci **21**(14): 5089-5098.

Hall, J., K. L. Thomas and B. J. Everitt (2001). "Cellular imaging of zif268 expression in the hippocampus and amygdala during contextual and cued fear memory retrieval: selective activation of hippocampal CA1 neurons during the recall of contextual memories." J Neurosci **21**(6): 2186-2193.

Hashimoto, H., Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang and X. Cheng (2014). "Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence." Genes & Development **28**(20): 2304-2313.

He, X.-J., T. Chen and J.-K. Zhu (2011). "Regulation and function of DNA methylation in plants and animals." Cell Research **21**(3): 442-465.

Hendrich, B. and A. Bird (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." Mol Cell Biol **18**(11): 6538-6547.

Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-232.

Hotchkiss, R. D. (1948). "The quantitative separation of purines, pyrimidines, and nucleosides by paper

chromatography." J Biol Chem **175**(1): 315-332.
Jeong, S., G. Liang, S. Sharma, J. C. Lin, S. H. Choi, H. Han, C. B. Yoo, G. Egger, A. S. Yang and P. A. Jones (2009). "Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA." Mol Cell Biol **29**(19): 5366-5376.
Johnson, T. B. and R. D. Coghill (1925). "Researches on pyrimidines. CIII. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus." Journal of the American Chemical Society **47**: 2838-2844.
Jones, M. W., M. L. Errington, P. J. French, A. Fine, T. V. Bliss, S. Garel, P. Charnay, B. Bozon, S. Laroche and S. Davis (2001). "A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories." Nat Neurosci **4**(3): 289-296.
Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nat Rev Genet **13**(7): 484-492.
Kemme, C. A., R. Marquez, R. H. Luu and J. Iwahara (2017). "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research **45**(13): 7751-7759.
Kessler, N. J., T. E. Van Baak, M. S. Baker, E. Laritsky, C. Coarfa and R. A. Waterland (2016). "CpG Methylation Differences Between Neurons and Glia are Highly Conserved from Mouse to Human." Faseb Journal **30**.
Kozlenkov, A., P. Roussos, A. Timashpolsky, M. Barbu, S. Rudchenko, M. Bibikova, B. Klotzle, W. Byne, R. Lyddon, A. F. Di Narzo, Y. L. Hurd, E. V. Koonin and S. Dracheva (2014). "Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites." Nucleic Acids Research **42**(1): 109-127.
Kozlenkov, A., M. H. Wang, P. Roussos, S. Rudchenko, M. Barbu, M. Bibikova, B. Klotzle, A. J. Dwork, B. Zhang, Y. L. Hurd, E. V. Koonin, M. Wegner and S. Dracheva (2016). "Substantial DNA methylation differences between two major neuronal subtypes in human brain." Nucleic Acids Research **44**(6): 2593-2612.
Kubosaki, A., Y. Tomaru, M. Tagami, E. Arner, H. Miura, T. Suzuki, M. Suzuki, H. Suzuki and Y. Hayashizaki (2009). "Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation." Genome Biol **10**(4): R41.
Lee, J. L., B. J. Everitt and K. L. Thomas (2004). "Independent cellular processes for hippocampal memory consolidation and reconsolidation." Science **304**(5672): 839-843.
Lee, T.-f., J. Zhai and B. C. Meyers (2010). "Conservation and divergence in eukaryotic DNA methylation." Proceedings of the National Academy of Sciences of the United States of America **107**(20): 9027-9028.
Li, E., T. H. Bestor and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.
Li, E. and Y. Zhang (2014). "DNA methylation in mammals." Cold Spring Harb Perspect Biol **6**(5): a019133.
Li, Z., H. Dai, S. N. Martos, B. Xu, Y. Gao, T. Li, G. Zhu, D. E. Schones and Z. Wang (2015). "Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells." Genome Biology **16**(1): 115.
Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens and J. R. Ecker (2013). "Global epigenomic reconfiguration during mammalian brain development." Science **341**(6146): 1237905.
Maddox, S. A., M. S. Monsey and G. E. Schafe (2011). "Early growth response gene 1 (Egr-1) is required for new and reactivated fear memories in the lateral amygdala." Learn Mem **18**(1): 24-38.
Maroteaux, M., E. Valjent, S. Longueville, P. Topilko, J. A. Girault and D. Herve (2014). "Role of the plasticity-associated transcription factor zif268 in the early phase of instrumental learning." PLoS One **9**(1): e81868.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang and J. F. Costello (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature **466**(7303): 253-257.

McGowan, P. O., A. Sasaki, A. C. D'Alessio, S. Dymov, B. Labonte, M. Szyf, G. Turecki and M. J. Meaney (2009). "Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse." Nat Neurosci **12**(3): 342-348.

Mo, A., E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J. Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker and J. Nathans (2015). "Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain." Neuron **86**(6): 1369-1384.

Mokin, M. and J. Keifer (2005). "Expression of the immediate-early gene-encoded protein Egr-1 (zif268) during in vitro classical conditioning." Learn Mem **12**(2): 144-149.

Mora-Lopez, F., N. Pedreno-Horrillo, L. Delgado-Perez, J. A. Brieva and A. Campos-Caro (2008). "Transcription of PRDM1, the master regulator for plasma cell differentiation, depends on an SP1/SP3/EGR-1 GC-box." Eur J Immunol **38**(8): 2316-2324.

Okano, M., S. Xie and E. Li (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nat Genet **19**(3): 219-220.

Pavletich, N. P. and C. O. Pabo (1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A." Science **252**(5007): 809-817.

Penke, Z., E. Morice, A. Veyrac, A. Gros, C. Chagneau, P. LeBlanc, N. Samson, K. Baumgartel, I. M. Mansuy, S. Davis and S. Laroche (2014). "Zif268/Egr1 gain of function facilitates hippocampal synaptic plasticity and long-term spatial recognition memory." Philos Trans R Soc Lond B Biol Sci **369**(1633): 20130159.

Poirier, R., H. Cheval, C. Mailhes, S. Garel, P. Charnay, S. Davis and S. Laroche (2008). "Distinct functions of egr gene family members in cognitive processes." Front Neurosci **2**(1): 47-55.

Poulin, J. F., B. Tasic, J. Hjerling-Leffler, J. M. Trimarchi and R. Awatramani (2016). "Disentangling neural cell diversity using single-cell transcriptomics." Nat Neurosci **19**(9): 1131-1141.

Raiber, E. A., R. Hardisty, P. van Delft and S. Balasubramanian (2017). "Mapping and elucidating the function of modified bases in DNA." Nature Reviews Chemistry **1**(9).

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet **14**(1): 9-25.

Saito, M. and F. Ishikawa (2002). "The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2." J Biol Chem **277**(38): 35434-35439.

Sarda, S., A. Das, C. Vinson and S. Hannenhalli (2017). "Distal CpG islands can serve as alternative promoters to transcribe genes with silenced proximal promoters." Genome Res **27**(4): 553-566.

Sharif, J., M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano and H. Koseki (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." Nature **450**(7171): 908-912.

Smallwood, S. A., H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik and G. Kelsey (2014). "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity." Nat Methods **11**(8): 817-820.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." Nature **480**(7378): 490-495.

Stirzaker, C., J. Z. Song, W. Ng, Q. Du, N. J. Armstrong, W. J. Locke, A. L. Statham, H. French, R. Pidsley, F. Valdes-Mora, E. Zotenko and S. J. Clark (2017). "Methyl-CpG-binding protein MBD2 plays a key role in maintenance and spread of DNA methylation at CpG islands and shores in cancer." Oncogene **36**(10): 1328-1338.

Telese, F., A. Gamliel, D. Skowronska-Krawczyk, I. Garcia-Bassets and Michael G. Rosenfeld (2013). "“Seq-ing” Insights into the Epigenetics of Neuronal Gene Regulation." Neuron **77**(4): 606-623.

Tognini, P., D. Napoli and T. Pizzorusso (2015). "Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity." Front Cell Neurosci **9**: 331.

Veyrac, A., A. Gros, E. Bruel-Jungerman, C. Rochefort, F. B. Kleine Borgmann, S. Jessberger and S. Laroche (2013). "Zif268/egr1 gene controls the selection, maturation and functional integration of adult hippocampal newborn neurons by learning." Proc Natl Acad Sci U S A **110**(17): 7062-7067.

Wang, Y., M. Xiao, X. Chen, L. Chen, Y. Xu, L. Lv, P. Wang, H. Yang, S. Ma, H. Lin, B. Jiao, R. Ren, D. Ye, K. L. Guan and Y. Xiong (2015). "WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation." Mol Cell **57**(4): 662-673.

Weaver, I. C., N. Cervoni, F. A. Champagne, A. C. D'Alessio, S. Sharma, J. R. Seckl, S. Dymov, M. Szyf and M. J. Meaney (2004). "Epigenetic programming by maternal behavior." Nat Neurosci **7**(8): 847-854.

Weaver, I. C., A. C. D'Alessio, S. E. Brown, I. C. Hellstrom, S. Dymov, S. Sharma, M. Szyf and M. J. Meaney (2007). "The transcription factor nerve growth factor-inducible protein a mediates epigenetic programming: altering epigenetic marks by immediate-early genes." J Neurosci **27**(7): 1756-1768.

Wei, F., Z. C. Xu, Z. Qu, J. Milbrandt and M. Zhuo (2000). "Role of EGR1 in hippocampal synaptic enhancement induced by tetanic stimulation and amputation." J Cell Biol **149**(7): 1325-1334.

Wyatt, G. R. (1951). "Recognition and estimation of 5-methylcytosine in nucleic acids." Biochemical Journal **48**(5): 581-584.

Zandarashvili, L., M. A. White, A. Esadze and J. Iwahara (2015). "Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1." Febs Letters **589**(15): 1748-1753.

Zeng, X. R., Y. Sun, L. Wenger and H. S. Cheung (2005). "Basic calcium phosphate crystal-induced Egr-1 expression stimulates mitogenesis in human fibroblasts." Biochem Biophys Res Commun **330**(3): 658-664.

Zhang, Y., K. Chen, S. A. Sloan, M. L. Bennett, A. R. Scholze, S. O'Keeffe, H. P. Phatnani, P. Guarnieri, C. Caneda, N. Ruderisch, S. Deng, S. A. Liddelow, C. Zhang, R. Daneman, T. Maniatis, B. A. Barres and J. Q. Wu (2014). "An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex." J Neurosci **34**(36): 11929-11947.

Zhu, H., G. H. Wang and J. Qian (2016). "Transcription factors as readers and effectors of DNA methylation." Nature Reviews Genetics **17**(9): 551-565.

Chapter 2 - Mammalian Brain Development is Accompanied by a Dramatic Increase in Bipolar DNA Methylation

2.1 ABSTRACT

DNA methylation is an epigenetic mechanism critical for tissue development and cell specification. Mammalian brains consist of many different types of cells with assumedly distinct DNA methylation profiles, and thus some genomic loci may demonstrate bipolar DNA methylation pattern, i.e., hypermethylated in one cell subset but hypomethylated in others. Currently, how extensive methylation patterns vary among brain cells is unknown and bipolar methylated genomic loci remain largely unexplored. In this study, we implemented a procedure to infer CSM loci from the methylomes of human and mouse frontal cortices at different developmental stages. With the genome-scale hairpin bisulfite sequencing approach, we demonstrated that the majority of CSM loci predicted likely resulted from the methylation differences among brain cells rather than from asymmetric DNA methylation between DNA double strands. Correlated with enhancer-associated histone modifications, putative CSM loci increased dramatically during early stages of brain development and were enriched for GWAS variants associated with neurological disorder-related diseases/traits. Altogether, this study provides a procedure to identify genomic regions showing methylation differences in a mixed cell population and our results suggest that a set of cis-regulatory elements are primed in early postnatal life whose functions may be compromised in human neurological disorders.

2.2 INTRODUCTION

DNA methylation is the most common covalent modification known to occur to mammalian genomic DNA. The importance of DNA methylation has been firmly established for neuronal differentiation, neural plasticity and function throughout the lifespan (Martinowich, Hattori et al. 2003, Ballas, Grunseich et al. 2005, Mohn, Weber et al. 2008). During early neuronal differentiation, *de novo* DNA methylation occurs at the promoters of germ line-specific genes to repress pluripotency in progenitor cells, while the methylation loss at other promoters activates neuron-specific genes (Mohn, Weber et al. 2008). After birth, neuronal methylation profiles continue to evolve in parallel with developmental plastic changes (Lister, Mukamel et al. 2013).

In mature brains, substantial DNA methylation changes can result from neuronal activity, for example within hours after electroconvulsive stimulation (Guo, Ma et al. 2011). The methylation dynamics in neurons have been recognized to be critical for activity-dependent plasticity underlying brain functions including learning and memory (Weaver, Cervoni et al. 2004, Day and Sweatt 2010, Day, Childs et al. 2013).

Remarkable heterogeneity in DNA methylation has been observed within mammalian brains, which are comprised of functionally distinct cell subsets (Ladd-Acosta, Pevsner et al. 2007). Variations in DNA methylation across brain regions are widespread in human methylomes and consistent within individual brains (Ladd-Acosta, Pevsner et al. 2007). In the mouse brain, unique epigenetic landscapes distinguish different brain regions and account for region-specific functional specialization (Sanchez-Mut, Aso et al. 2013). To explore CSM, several studies have been conducted to compare methylation profiles of neuronal and non-neuronal cells using a neuron-specific antibody (NeuN) and fluorescence-activated cell sorting (Iwamoto, Bundo et al. 2011, Lister, Mukamel et al. 2013, Montano, Irizarry et al. 2013, Kozlenkov, Roussos et al. 2014). Compared with those of non-neuronal cells, neuronal methylomes show distinctive DNA methylation signatures with low global DNA methylation and high inter-individual variations (Iwamoto, Bundo et al. 2011). The differentially methylated regions between neuronal and non-neuronal cells are enriched in CpG island shores, enhancers and gene bodies of neuron-specific genes (Kozlenkov, Roussos et al. 2014). Despite these advances, current understanding of brain methylation heterogeneity is still very limited, and the epigenetic regulatory mechanisms associated with brain cell specification are largely unexplored.

Because the classification of cell types in the brain remains a work in progress (Fishell and Heintz 2013), even with the advance of single-cell methylome sequencing technique, the identification of epigenetic marks for each brain cell subset is a daunting task. This prompted us to explore alternative ways to decode the brain methylome derived from unsorted cells. In normal somatic tissues, DNA methylation usually displays a bimodal distribution and the methylation levels between neighboring CpG dinucleotides are strongly correlated (Eckhardt, Lewin et al. 2006). Thus, genomic DNA may be partitioned into two fractions: hypermethylated and hypomethylated (Deaton and Bird 2011). Within heterogeneous tissues, there exist so-called CSM loci, which show bipolar methylation patterns, i.e. hypermethylated in one cell subset but hypomethylated in others. However, bipolar DNA methylation patterns may also result from

epigenetic phenomena unrelated to CSM: allele-specific DNA methylation (ASM) and asymmetric DNA methylation. Fortunately, such bipolar methylation loci can be distinguished from the CSM loci. Recently, a mouse ASM map has been generated with brain tissues derived from reciprocal crosses between two distantly related mouse strains (Xie, Barr et al. 2012). A total of 1,952 CG dinucleotides in 55 discrete genomic loci in the mouse have been identified as imprinted. The number of human imprinted regions were found to be very limited as well (John and Lefebvre 2011) and 51 ASM loci were identified (Court, Tayama et al. 2014). Additionally, asymmetric DNA methylation may be detected with hairpin bisulfite sequencing technique (Zhao, Sun et al. 2014), which generates methylation data for two complementary DNA strands simultaneously. Thus, the distinction of CSM from other types of bipolar methylation patterns (i.e. ASM and asymmetric DNA methylation) may provide an alternative and complementary solution to the cell-sorting-based method to dissect brain epigenetic heterogeneity.

Here, we first developed an analytical procedure to infer CSM loci and applied it to human and mouse brain methylomes (Lister, Mukamel et al. 2013). We next used the genome-scale hairpin bisulfite sequencing technique to explore the symmetry of methylation on DNA double strands in human fetal and adolescent brain tissues. We found that, compared with ESCs (Zhao, Sun et al. 2014), brains exhibit exceedingly higher levels of symmetrical DNA methylation. We further explored the functional relevance of the predicted brain CSM loci *via* integrative "omics" analysis with disease/trait-associated genetic variants and ChIP-seq data for histone modifications. The integrative analysis suggested that putative brain CSM loci are critical elements which may be associated with epigenetic aberrations in human brain diseases.

2.3 METHODS

2.3.1 Accession codes.

The hairpin bisulfite sequencing data and ChIP-seq data generated in this study, including relevant processed data files, have been deposited in NCBI Gene Expression Omnibus (GEO) under accession number GSE67482. Additional data used in this manuscript were summarized in **Supplementary Table 6**.

2.3.2 Collection and dissection of post-mortem human brain tissues.

De-identified postmortem human brain tissues were acquired from autopsies performed at Children's Medical Center, University of Texas Southwestern, Dallas (IRB-exempted research). Gray and white matters from frontal cortices were obtained from a 22 weeks gestation male fetus and a 17-year-old male, who died of non-neurological diseases. Hematoxylin and eosin stained sections of formalin fixed paraffin embedded (FFPE) brain tissue from standard sections of the left frontal cortex (watershed area, adjacent to the frontal horn of the lateral ventricle) were reviewed. Representative cortical gray matter and white matter areas were demarcated, and their corresponding areas were mapped on the tissue block. Three to six 1mm tissue cores were taken from these areas and labeled as a cortical gray matter or white matter. For each tissue sample, deparaffinization was performed using 100% xylene with gentle shaking at room temperature (RT) for 15 min, repeated twice and then subjected to 100% ethanol wash, 3 times. The sections were dried on the bench for 1h, and then 1ml lysis buffer (50mM Tris, 25mM EDTA, 100mM NaCl, 0.5% Tween-20/SDS, pH 8.0) was added with 40µl proteinase K (Ambion). After gently shaking overnight at 55°C, heat-inactivation at 90°C for 1h was performed and samples were cooled down gradually, RNA digestion was performed using 5µl RNase cocktail (Ambion) at 37°C for 2h. Finally, genomic DNA was isolated with phenol/chloroform extraction followed by ethanol precipitation.

2.3.3 Genome-scale hairpin bisulfite sequencing.

Hairpin bisulfite-seq library construction was performed according to the previously described protocol (Zhao, Sun et al. 2014) with slight modifications. Briefly, 10µg genomic DNA of each sample was spiked with 0.02% unmethylated Lambda DNA (Promega) and sonicated to 200bp fragments with Covaris. After MseI and MluCI digestion (NEB), end repair and dA tailing, genomic DNA fragments were ligated to the Biotin-modified hairpin adapter (5'P-GGCCAGCTGCAAG/iBiodT/GAAGCAGCTGGCCT-3', IDT). After captured with Dynabeads® MyOne™ Streptavidin C1 beads (Invitrogen), genomic DNA fragments were subjected to bisulfite conversion using the EpiTect Bisulphite Kit (Qiagen), PCR, and pair-end sequenced using Illumina MiSeq and HiSeq 2000. Illumina Sequencing services were performed at the genomic core of Virginia Bioinformatics Institute.

2.3.4 Hairpin bisulfite sequencing data analysis.

The hairpin bisulfite sequencing reads were processed and aligned similarly to a previous study (Zhao, Sun et al. 2014). For each read, adaptor and hairpin sequences were searched with cross_match. Additional searches on the 3'-end of sequence reads were conducted to eliminate any sub-string derived from a hairpin sequence adaptor. Then, the processed read pairs were globally aligned using Needleman-Wunsch algorithm. After trimming the overhangs of the aligned sequences, original sequences were recovered for read pairs with at least 90% identity between two arms. Finally, the recovered original sequences were mapped to human reference (hg19) using Bowtie 2 (Langmead and Salzberg 2012) with parameters (-N1-L22), and only uniquely mapped reads were retained. Methylation level and fidelity were calculated according to a previous study (Zhao, Sun et al. 2014). Statistics including the number of total uniquely mapped reads, genome coverage, CpG coverage and the average sequencing depth were summarized in **Supplementary Table 4**.

2.3.5 MethylC-Seq data analysis to infer brain pCSM loci.

MethylC-Seq data for human and mouse frontal cortex (Lister, Mukamel et al. 2013) were retrieved from NCBI Sequence Read Archive (SRA) with accession SRP026048. Due to low sequencing depth, several samples (Hs 55yr tissue, Hs 64yr, Mm 6wk NeuN+ and Mm 6wk NeuN-) were excluded from pCSM prediction. Read processing was performed as previously described (Lister, Mukamel et al. 2013). The processed reads were aligned to the corresponding human (hg19) or mouse (mm10) reference genomes using Bismark (Krueger and Andrews 2011) with parameters –n2 –l50. After PCR duplicates were removed, methylation callings were extracted. Basic statistics can be found in **Supplementary Table 1.**

To normalize sequencing depth across methylomes, a "down-sampling" strategy is applied at each 4-CpG segment. For the methylomes analyzed, we first obtained a set of common 4-CpG segments, which were covered by at least ten reads in all samples. For each 4-CpG segment, we determined the minimum sequencing depth $D_{\min}$ in these samples. Then, for a given sample, if its sequencing depth was bigger than D_{min}, reads were randomly discarded from the corresponding data set until reaching $D_{\min}$. For each sample, the "down-sampling" procedure was repeated 100 times for each segment, and the possibility to detect CSM pattern was calculated. Such "down-sampling" strategy was adopted to compare brain methylomes classified into the following four

groups: human developmental stages, human cell types (NeuN+ vs. NeuN-), mouse developmental stages, mouse cell types (NeuN+ vs. NeuN-).

Nonparametric Bayesian clustering (Wu, Sun et al. 2015) was used to identify the 4-CpG segments with bipolar methylation patterns. We first scanned all possible segments with four neighboring CpGs within a sequence read. The 4-CpG segments covered with at least ten reads were used for pCSM prediction. CSM segments in chromosome X, Y and known imprinted regions for mouse (Xie, Barr et al. 2012) and human (Court, Tayama et al. 2014) were excluded from further analysis. The overlapped pCSM segments were further merged to pCSM regions or so-called pCSM loci.

2.3.6. Genome annotation, gene ontology and KEGG pathway analysis.

The annotations for genomic regions, including transcripts and CpG islands, were downloaded from UCSC genome browser (Kuhn, Haussler et al. 2013). Promoters were arbitrarily defined as regions 2 kb upstream of each TSS (transcription start site). Promoters were further classified as CGI and non-CGI groups based on whether they overlapped with CGIs. pCSM associated genes were defined as those with at least one pCSM segment identified within 10kb from their TSSs. The background gene list was determined for each sample to include genes with at least one 4-CpG segment within 10kb from their TSSs. GO enrichment analysis and KEGG pathway analysis were performed using DAVID functional annotation tools (Dennis, Sherman et al. 2003).

2.3.7. Integrative "Omic" data analysis.

ChIP-seq data for histone modifications (**Supplementary Table 6**) were collected from previous publications. The histone modifications include H3K4me1, H3K4me3 and H3K27ac for mouse brain (Shen, Yue et al. 2012), and H3K4me1, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3 and H3K36me3 for the human brain (Bernstein, Stamatoyannopoulos et al. 2010). The coordinates of histone peaks were converted to mm 10 by using UCSC liftOver. Based on histone modifications including H3K4me1, H3K4me3 and H3K27ac, several types of important regulatory elements could be annotated (Nord, Blow et al. 2013, Zhu, Adli et al. 2013). First, all overlapped peaks for these three histone modifications were merged to form the merged-peak regions. By examining the histone modification occupancy of each merged-peak region, regulatory elements were annotated as: 1) Active promoters: with H3K4me3. They were further classified into two groups based on their overlapping status with CGIs. 2) Active enhancers: with

H3K27ac but without H3K4me3. 3) Poised enhancers: with H3K4me1 but without H3K27ac or H3K4me3. The microarray and ISH data for human brain used to validate our result were adopted from Allen Brain Atlas (Hawrylycz, Lein et al. 2012, Zeng, Shen et al. 2012).

GWAS variants information was collected from the NHGRI GWAS catalog (http://www.genome.gov/GWAStudies, downloaded on Sept 17, 2014). For each disease/trait, its associated SNPs were collapsed into one unique set, then extended by propagating disease/trait associations to proxy SNPs using the SNAP search tool (Johnson, Handsaker et al. 2008) (http://www.broadinstitute.org/mpg/snap) based on linkage disequilibrium (r^2>0.8) between SNPs (within 250kb) in any of the three populations in the 1000 genomes project pilot data (Abecasis, Altshuler et al. 2010). With this approach, the total 15,578 SNP-disease associations were extended to a set of 379,762 proxy SNPs corresponding to 1,127 diseases/traits. Next, for each of the eleven human brain methylomes, locations of the identified pCSM regions were obtained after a 100bp extension on both up and down-streams. Enrichment analysis of GWAS variants in pCSM regions for each disease/trait was done by testing whether the overlapping frequency of the disease-associated SNPs was significantly higher than expected. To calculate enrichment p-values, a permutation test was performed by randomly shuffling (shifting along the circulated genome) the locations of pCSM regions for 10,000 iterations. The number of SNPs together with their proxy SNPs within 100bp from pCSM regions was counted in each iteration to generate a null distribution. Finally, the total 1,127 diseases/traits were ranked in descending order by counting their significant overlapping p-values shown in the eleven human methylomes. This recurrent enrichment in multiple methylomes provides strong evidence for the functional importance of pCSM regions.

2.4 RESULTS

2.4.1 CSM dynamics during mammalian brain development.

To investigate brain CSM, we designed a reverse engineering approach to analyze DNA methylation patterns embedded in bisulfite sequencing reads (**Fig. 1a**). During this procedure, we first excluded potential PCR duplicates and progressively scanned each sequence to extract methylation patterns for genomic segments containing four neighboring CpG dinucleotides. The 4-CpG segments mapped to the same locus were clustered together, and the clusters with at least 10Xs read coverage were selected for further analysis. To increase the certainty of bipolar

detection, each selected cluster was required to include at least one completely methylated and one completely unmethylated read concurrently. We further analyzed the distributions of methylation patterns for each cluster to identify bipolar methylated ones using nonparametric Bayesian clustering approach (Wu, Sun et al. 2015).

For both human and mouse, we re-analyzed eleven methylomes[4] derived from the frontal cortices from seven developmental stages along with two pairs of fluorescence activated cell sorted NeuN+ and NeuN- cells. After the filtering of imprinted loci, we inferred the rest of bipolar methylated segments as putative cell-subset specific methylated loci, which were denoted as pCSM loci. We identified 5,636 to 96,033 pCSM loci in human autosomes and 3,343 to 57,001 pCSM loci in mouse autosomes (**Supplementary Table 1**). A recent comparison of neuronal and stem-cell methylomes revealed that cell-subset-specific regulation is associated with a set of low-methylated regions with an average methylation level of 30% (Stadler, Murr et al. 2011). In contrast, we found that the majority of pCSM loci were hypermethylated. For these pCSM loci, completely unmethylated patterns may exist in a small percentage of cells within bulk brain tissues. For instance, a number of bipolar methylated loci were adjacent to the transcription start sites (TSSs) of genes expressed in a specific cell type, such as CAMK2A, which is known to be selectively expressed in excitatory neurons (Ochiishi, Yamauchi et al. 1998) (**Fig. 1b,c**).

We next determined the frequency of pCSM for each methylome, which was defined as the percentage of pCSM segments predicted from all 4-CpG segments with at least 10Xs read coverage. The pCSM frequencies ranged from 1.2% in human fetal frontal cortex to 4.6% in 55-year-old NeuN+ cells. To enable a quantitative comparison across developmental stages, we focused on the 4-CpG segments with at least 10Xs read coverage in all stages (n = 171,178 for human; n = 35,605 for mouse). Since the number of pCSM loci identified from methylomes highly depends on the sequencing depth, we performed down-sampling normalization to minimize the bias that results from uneven sequencing depth. As previously reported[4], the levels of global DNA methylation remain constant during brain development (**Fig. 1d**). Interestingly, we observed a dramatic increase in pCSM frequency during early stages of brain development in both human (**Fig. 1d**) and mouse (**Supplementary Fig. 2**). The pCSM frequency was 1.7% for human fetal frontal cortex, increasing to 4.4% in two-year-old, and remained stable at later stages. This result suggests that a growing number of genomic loci are involved in brain cell specification at early stages and reach a plateau in adult brains.

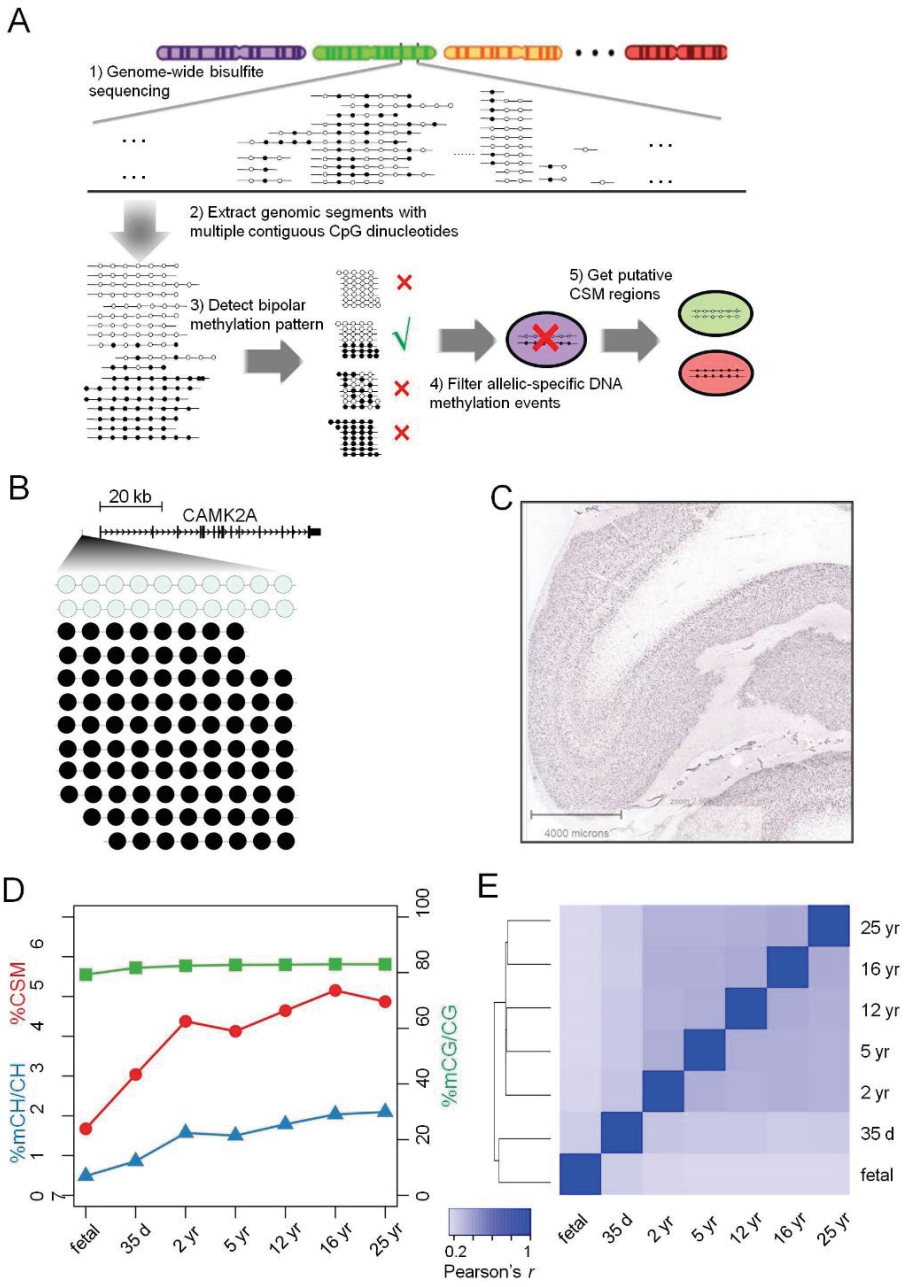


Figure 1. Identification of pCSM loci in human brain methylomes.

(a) Workflow for the computational inference of pCSM loci. (b) Methylation pattern of a predicted CSM loci (Chr5:149675541-149675604) at 6,138 bp upstream of CAMK2A gene in 25 yr brain methylome. (c) In situ hybridization for CAMK2A in human prefrontal cortex showing positive staining in the excitatory neurons in layers II-VI with absent staining in the glial cells (modified from Allen Brain Atlas; http://human.brain-map.org/ish/specimen/show/80936541?gene=811). (d) Changes of mC level in CG and CH context and the percentage of CSM segments after down-sampling normalization during human brain development. (e) Hierarchical clustering based on Pearson's correlations of pCSM statuses predicted for 4-CpG segments in different human brain methylomes.

We further performed pair-wise comparisons of pCSM profiles and determined their correlations among developmental stages. In coherence with the swiftly established CSM status during early developmental stages, the fetal and 35-day brains were clustered together and maturing brains gave rise to another cluster (**Fig. 1e**). For human and mouse brains, the closer the two developmental stages are, the higher the correlation in pCSM profiles (**Fig. 1e, Supplementary Fig. 2b** and **Supplementary Table 2**). Compared with non-neuronal cells, both human and mouse neurons exhibited higher frequencies of pCSM (**Supplementary Fig. 3a, b**), and the pCSM profiles were conserved for the same cell types (**Supplementary Fig. 3c, d).** This suggests that more genomic loci are needed to encode highly diverse neuronal cells than non-neuronal cells. Such a high epigenetic heterogeneity within NeuN+ cell population may explain the higher inter-individual methylation variations observed for NeuN+ cells in a recent study (Iwamoto, Bundo et al. 2011). In addition, for both human and mouse brain methylomes, the percentages of pCSMs shared within a cell type are consistently higher than those between different cell types (**Supplementary Table 3** and **Supplementary Fig. 4**).

It has been reported that the levels of non-CpG methylation (mCH) and 5-hydroxymethylcytosine (5hmC) increase during neuronal development (Jin, Wu et al. 2011, Szulwach, Li et al. 2011, Hahn, Qiu et al. 2013, Lister, Mukamel et al. 2013). This prompted us to investigate whether they were related to the establishment of CSM loci. Similar to the previous report (Lister, Mukamel et al. 2013), an increase in the mCH frequency was observed during brain development; while the mCG levels remained relatively stable (**Fig. 1d** and **Supplementary Fig. 2a**). We next examined the levels of mCH within pCSM loci and controls (all 4-CpG segments with at least 10Xs coverage but with CSM loci excluded) in fetal and adult brains, NeuN+ and NeuN- cells (**Supplementary Fig. 5**). For both CSM and control 4-CpG segments, the level of mCH was positively correlated with the level of mCG (Pearson's *r* ranges from 0.05 to 0.41) in all samples examined. We further compared pCSM segments with the controls at the same interval of mCG level and found that pCSM loci tend to have a higher level of mCH, especially in adult brain and neurons. In addition, we observed that hmC was also enriched in pCSM loci (**Supplementary Fig. 6**). This suggested that the establishment of pCSM loci may have mechanistic links with the increasing 5hmC and mCH levels during neuronal development.

2.4.2 Human frontal cortex exhibits high methylation fidelity.

As mentioned previously, the predicted brain CSM loci may result from the methylation difference within a DNA molecule, i.e. asymmetric DNA methylation. To assess the levels of asymmetric DNA methylation in the brain, we performed genome-scale hairpin bisulfite sequencing with gray and white matters from human fetal and adolescent frontal cortices (**Supplementary Table 4**). For all samples, bisulfite conversion rates were determined to be over 99.7% with the spike-in lambda DNA control. Despite the variation in the numbers of sequence reads generated for these samples, we found that the genome-wide average of methylation level and fidelity could be denoted with data from as low as 10^4 randomly generated read pairs (**Supplementary Fig. 7**). All four brain tissues demonstrated exceedingly high methylation fidelity, which was defined as the percentage of CpG dyads with symmetrical methylation patterns. We recently determined that in undifferentiated and differentiating mouse ESCs, the average methylation fidelities were 88.5% and 91.9% respectively (Zhao, Sun et al. 2014). Here we found that in the human fetal brain, the average methylation fidelities were 94.9% and 95.0% for cortical grey and white matters, respectively. These numbers increase to 95.6% for grey matter and 95.8% for white matter in the adolescent brain (**Supplementary Table 4**).

Next, we examined the genomic distribution of asymmetrically methylated CpG dyads. Similar to previous observation made with mouse ESCs (Zhao, Sun et al. 2014), we found that the average methylation levels decrease to approximately 20% but the average methylation fidelity increase slightly approaching the TSSs (**Fig. 2a**). By comparing the methylation fidelity for CpG dyads within +/−1 kb from TSS against all other CpG dyads, we found that CpG dyads near TSSs could show small (~2%) yet significant increase of methylation fidelity (p-value between 0.02 and $p < 2.2e\text{-}16$ for different samples; Wilcoxon Rank Sum Test). The methylation levels of exons showed high variance across samples and were lower than those of introns from the same samples. However, the methylation fidelity remained constantly high along the entire gene body in all samples. Furthermore, compared with genic regions, intergenic regions and repetitive elements (in particular SINE) have 3–10% higher methylation levels (**Fig. 2b**) and their methylation fidelities were above 94% in all four brain samples (**Fig. 2c**). We next questioned whether asymmetrically methylated CpG dyads tend to cluster together and thus result in bipolar DNA methylation patterns. We applied sliding windows with one to four adjacent CpG dyads to all sequence reads and compared the methylation patterns between two complementary DNA

strands. DNA methylation pattern, herein, is defined as the combination of methylation statuses of adjacent CpG dinucleotides on the same DNA strand. We focused on CpG dinucleotides

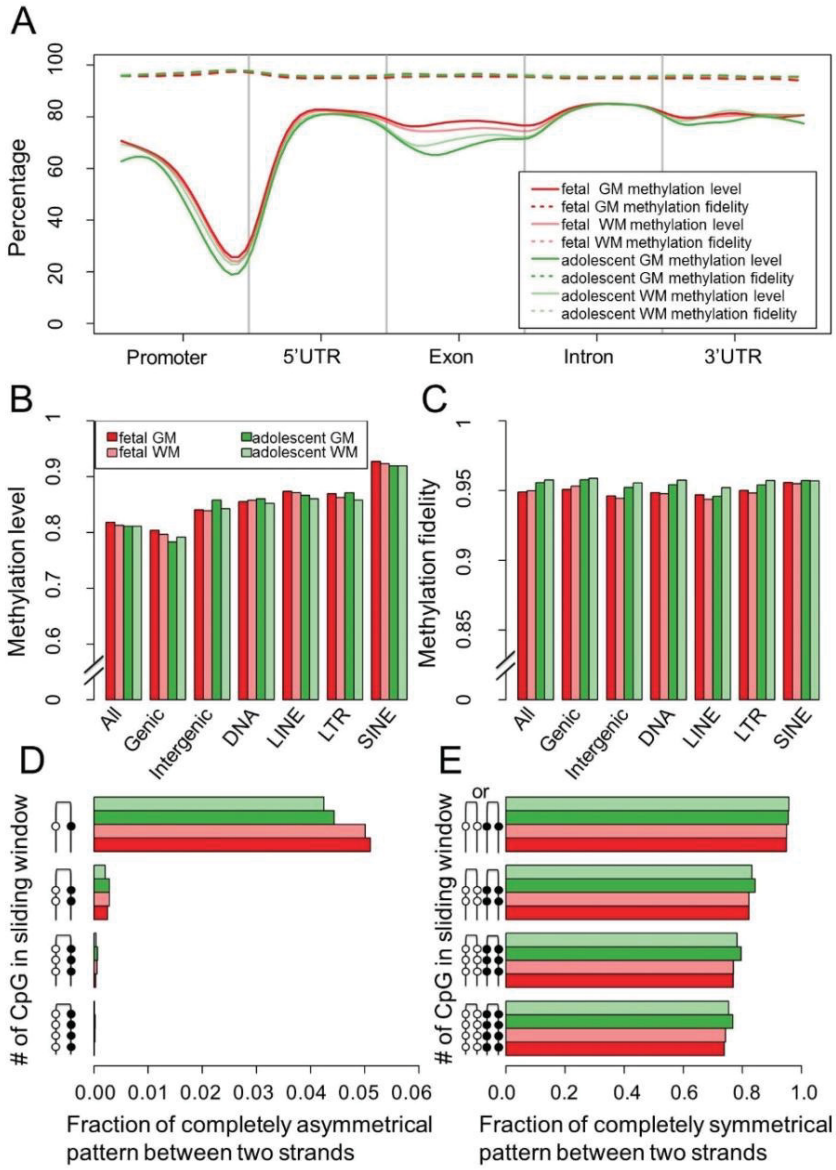


Figure 2. Hairpin bisulfite sequencing for human fetal and adolescent brains.

(a) Methylation level and fidelity along different gene-associated regions. Each genomic region was divided into 20 equal-sized bins. The smoothed lines represent the mean methylation level (solid lines) and methylation fidelity (dashed lines). (b) Methylation level in different genomic regions. (c) Methylation fidelity in different genomic regions. (d) Fraction of bipolar methylation pattern using different sliding windows. (e) Fraction of symmetrical (un)methylation pattern using different sliding windows. GM and WM represent gray matter and white matter, respectively.

showing the same methylation pattern (all methylated or all unmethylated) in a given sliding window. In all four brain samples, the frequency of asymmetrically methylated CpG dyads decreased more than 200 times from approximately 5% for a single CpG dyad to 0.02% for a sliding window of four CpG dyads (**Fig. 2d**). In contrast, the frequency of symmetrical methylation patterns decreased from approximately 95% for a single CpG dyad to 74% for a sliding window of four CpG dyads (**Fig. 2e**). Thus, in both fetal and adolescent brain methylomes, genomic segments with four neighboring CpG dinucleotides showing completely methylated or unmethylated patterns are more likely to have symmetrical (74%) rather than asymmetrical (0.02%) methylation patterns. Therefore, high methylation fidelity observed in the brain indicated that the predicted CSM loci result from methylation differences among brain cells instead of asymmetric DNA methylation.

2.4.3 CSM regions are enriched in non-CGI promoters and enhancers.

To evaluate the functional relevance of pCSM segments, we examined the occurrence of pCSM loci in promoters and identified their associated genes. Based on the presence of CGIs, gene promoters may be classified into CGI promoters and non-CGI promoters. Most CGI promoters are associated with housekeeping genes, while non-CGI promoters are usually for genes of non-ubiquitous (tissue or cell-subset specific) expression (Forrest, Kawaji et al. 2014). We observed that pCSM regions were depleted from CGI promoters but enriched in non-CGI promoters (**Fig. 3**). This finding is in agreement with previous studies, which reported that differentially methylated regions between human neuronal and glial cells are depleted from CGI-promoters (Kessler, Van Baak et al. 2016).

Various histone modifications have been used to generate genome-wide maps of chromatin state and to annotate regulatory elements (Zhu, Adli et al. 2013). Active histone marks such as H3K4me3 and H3K9ac are signatures of active promoters, while H3K4me1 marks enhancers in active or poised state and H3K27ac identifies active enhancers (Creyghton, Cheng et al. 2010). Based on the chromatin maps including H3K4me1, H3K4me3 and H3K27ac modifications in the adult human brain (Bernstein, Stamatoyannopoulos et al. 2010), we annotated genomic regions for active promoters, poised enhancers and active enhancers. Subsequently, the pCSM occurrence surrounding these regulatory elements was examined. We found that pCSM loci were overrepresented in both active non-CGI promoters and active/poised enhancers (**Supplementary**

Fig. 8). In contrast, no significant change on pCSM frequency was observed in surrounding genomic regions enriched for several other histone marks (Bernstein, Stamatoyannopoulos et al. 2010) including H3K9ac, H3K9me3, H3K27me3 and H3K36me3. This is in consistent with the fact that the enhancers are usually highly cell-subset specific (Creyghton, Cheng et al. 2010, Zhu, Adli et al. 2013).

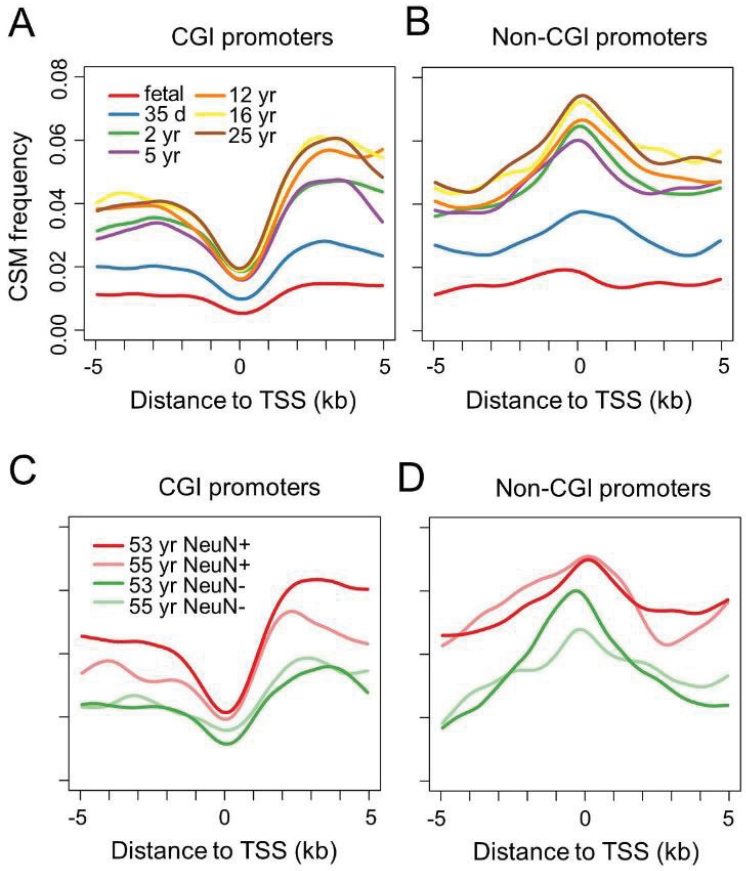


Figure 3. Aggregate plots of pCSM frequency surrounding TSSs.

pCSM frequency was calculated as the percentage of 4-CpG segments predicted to be pCSM. pCSM segments are depleted from CGI promoters (a,c), while over-represented in non-CGI promoters (b,d).

2.4.4 CSM regions are associated with brain functions and rich in brain disease/trait-associated SNPs.

For genes with pCSM loci within 10 kb of their TSSs, we performed GO enrichment analysis using DAVID functional annotation tools (Dennis, Sherman et al. 2003). Not surprisingly, we found that these genes were involved in functions including neuronal differentiation, cell morphogenesis, transcription factor activity and cell projection (**Fig. 4a** and **Supplementary**

Table 5). This result indicates that the identified pCSM segments may play important roles in the epigenetic regulation of brain cell specification and morphogenesis. Interestingly, remarkable differences in GO terms enriched were observed among different developmental stages and between cell types. GO terms including "neuron differentiation" and "cell morphogenesis" were enriched for NeuN+ cells, while "hemophilic cell adhesion" and "plasma membrane" were enriched for NeuN- cells. In addition, GO terms such as "neuron differentiation" and "transcription regulation" were highly enriched in maturing brains, but not in fetal brains. These results imply that the pCSM profiles in brains are functionally important for brain development and cell-subset specificity.

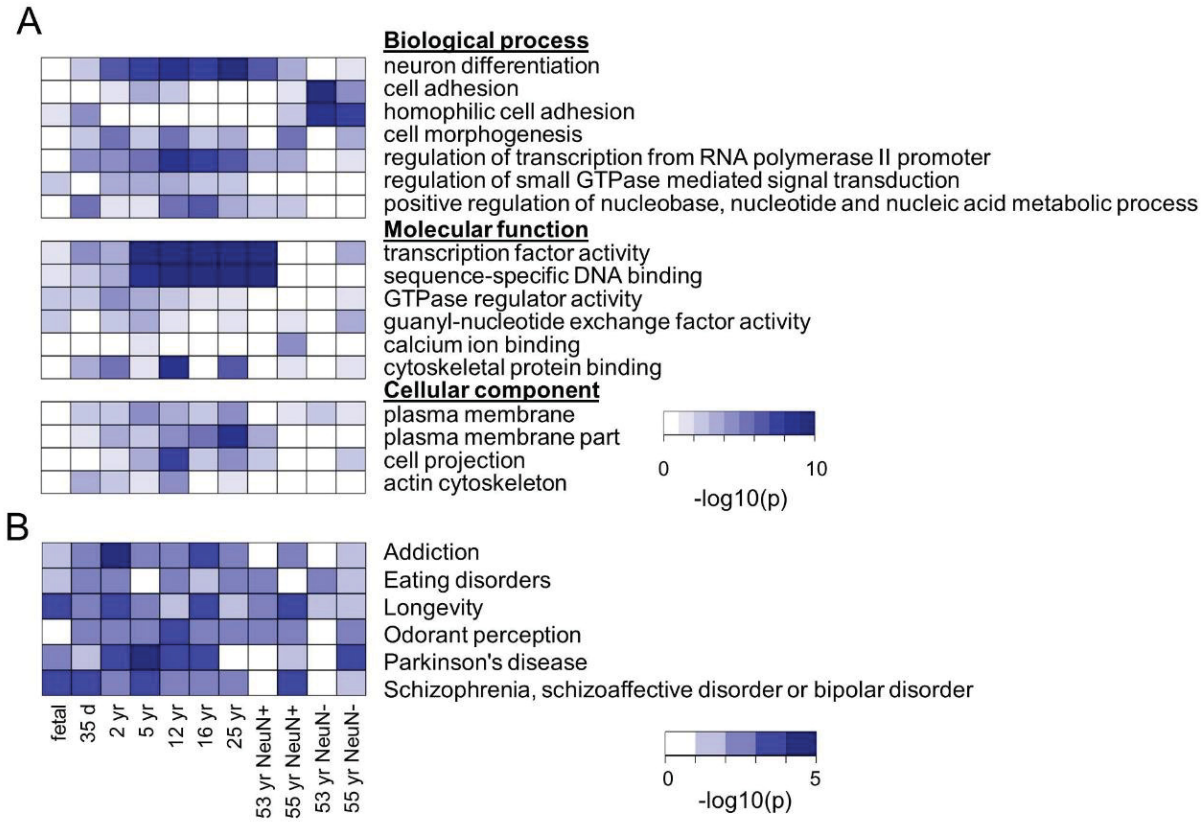


Figure 4. Functional annotation analyses for pCSM loci.

(a) GO enrichment analysis for genes associated with pCSM loci during human brain development and cell specification. Genes associated with pCSM were defined as those with pCSM loci within 10 kb from their TSSs. The color represents p-values after Bonferroni correction. The GO terms were grouped as "biological process", "molecular function" and "cellular component" (for the full result, see Supplementary Table 6). (b) Diseases/traits associated with pCSM loci. The color represents p-values determined by permutation tests with 10,000 iterations.

To explore whether the identified pCSM regions can provide insight into SNP variants associated with disease phenotypes, we performed statistical analysis on the overlap between disease/trait-associated SNP variants and pCSM regions. With the disease/trait-associated SNPs documented in NHGRI genome-wide association studies (GWAS) catalog (Hindorff, Sethupathy et al. 2009) and their proxy SNPs (Johnson, Handsaker et al. 2008) in strong linkage disequilibrium, we observed significant enrichment of SNP variants associated with major brain diseases/traits in the pCSM regions, including odorant perception, Schizophrenia, addiction, eating disorders, and Parkinson's disease (**Fig. 4b**). The SNPs associated with these diseases/traits were recurrently enriched in pCSM regions identified in the eleven human methylomes. These results not only suggest functional importance of the pCSM regions but also provide a basis for better understanding the underlying epigenetic mechanism of the 'common disease-common variant' association. Further studies are needed to determine whether the brain disease/trait-associated SNPs correlate with methylation alterations at pCSM regions within specific cell-subsets and at specific developmental stages.

2.5 DISCUSSION

The high degree of cellular complexity within the human brain has been well recognized. However, little is known of its epigenetic heterogeneity. To our knowledge, this study is the first attempt to systematically exploit the dynamics of epigenetic heterogeneity associated with mammalian brain development. The ideal approach to identify cell-subset specific methylation is to determine the methylomes of single cells. Recently, three labs (Guo, Zhu et al. 2013, Smallwood, Lee et al. 2014, Farlik, Sheffield et al. 2015) have generated methylomes at the single-cell level for embryonic stem cells. Unfortunately, the genome coverage of single-cell methylome data is frequently lower than 5%. This greatly limits the comparison of methylomes derived from different cells. Notably, single-cell methylation data alone cannot rule out methylation variations within a cell, i.e. allelic-specific methylation and asymmetric DNA methylation.

With hairpin bisulfite sequencing technique, we were able to estimate the contribution of asymmetric DNA methylation to bipolar methylation patterns observed in methylomes derived from brain tissues. We observed that human fetal and adolescent brain methylomes are with high methylation fidelity. This probably results from the fact that DNA methyltransferases in the brain

postmitotic cells have sufficient time to faithfully replicate parental DNA methylation patterns. Such high methylation fidelity in brain tissues leads to an extremely low frequency of asymmetrical DNA methylation. In both fetal and adolescent brain methylomes, as low as 0.02% of genomic segments with four neighboring CpG dinucleotides showing completely methylated or unmethylated patterns are with asymmetrical DNA methylation patterns. Currently, the catalogue of brain cell types is under debate[15] and it remains impossible to clearly define every cell subtype in the brain with existing methods. However, the high methylation fidelity of brain methylomes observed in this study indicates that assessing large populations of methylation patterns might aid in this endeavor. In light of this finding, we developed a unique approach to dissect brain methylomes and identified genomic loci potentially associated with brain cell specification. Such a tool provides an alternative and complementary solution to the cell-sorting-based approach to dissect brain epigenetic heterogeneity. There are several limitations in our study. Due to the still prohibitive cost of sequencing, most current methylomes were generated at around 10Xs coverage. Considering the numerous cell types in mammalian brains, we are aware that this study may not capture all the methylation variations among brain cells. Recently, forty-nine monoallelic methylated loci have been identified in genic regions (Steyaert, Van Criekinge et al. 2014). Although the majority of bipolar methylated loci in brain are unlikely caused by random monoallelic methylation, we cannot completely rule out such a possibility. For instance, the mosaic methylation patterns at the promoters of protocadherin-α gene cluster may help the monoallelic and combinatorial expression of variable isoforms in individual Purkinje cells (Kawaguchi, Toyama et al. 2008). In addition, our analytical procedure may be applied to regular bisulfite sequencing data for other normal tissues, but not for tumors or fast-dividing cells with low methylation fidelity, in which prevalent asymmetric DNA methylation may contribute substantially to bipolar DNA methylation (Zhao, Sun et al. 2014).

In this study, we observed a dramatic increase in the frequency of pCSM loci during early postnatal brain development. This intriguing result indicates that early postnatal stages are critical for mammalian brain development to create the diversity at the epigenetic level although neurogenesis has largely completed at birth. During the first few years for human and weeks for mouse, we found that brain cells gain significant variations in DNA methylation patterns. The established pCSM loci were highly enriched in regulatory elements controlling gene expression, i.e., enhancers and non-CGI promoters. In addition, the functional annotation analyses indicate

that brain pCSM loci are associated with genes relevant to brain development and neuron-specific functions. By exploring the NHGRI GWAS catalog, we further observed that the strong links between brain pCSM loci and the genomic loci associated with human neurological disorders, including addiction, Parkinson's diseases, schizophrenia, etc. Thus, the brain pCSM loci predicted, and the tool we provided in this study could inform the candidate genomic loci associated with epigenetic aberrations in neurological disorders.

Lastly, the pCSM loci identified in this study may serve as epigenetic markers for specific cell lineages and associated gene networks possibly involved in epigenetic regulatory processes determining cell fate. During cellular differentiation, the establishment of cell-subset specific methylation patterns enables cells with same genetic composition to stably silence specified genes and exhibit distinct phenotypes (Mohn and Schubeler 2009). The "on/off" combinations of epigenetic switches may be used to classify cells within the frontal cortex into distinct subsets. Such information on brain developmental trajectory will help design a probe set for monitoring developmental stages of specific cells, which will be useful in pre-clinical models including neural stem cell differentiation and/or reprogramming. The information about cell-subset specific methylated loci may be integrated with gene expression profile, morphologic, neurochemical and electrophysiological properties of specific cell-subsets. Such efforts will extend our understanding of brain cell identity and may enable reprogramming strategies for specific neuronal cell types with parameters at the epigenetic level for standardized quantification.

2.6 ACKNOWLEDGEMENTS

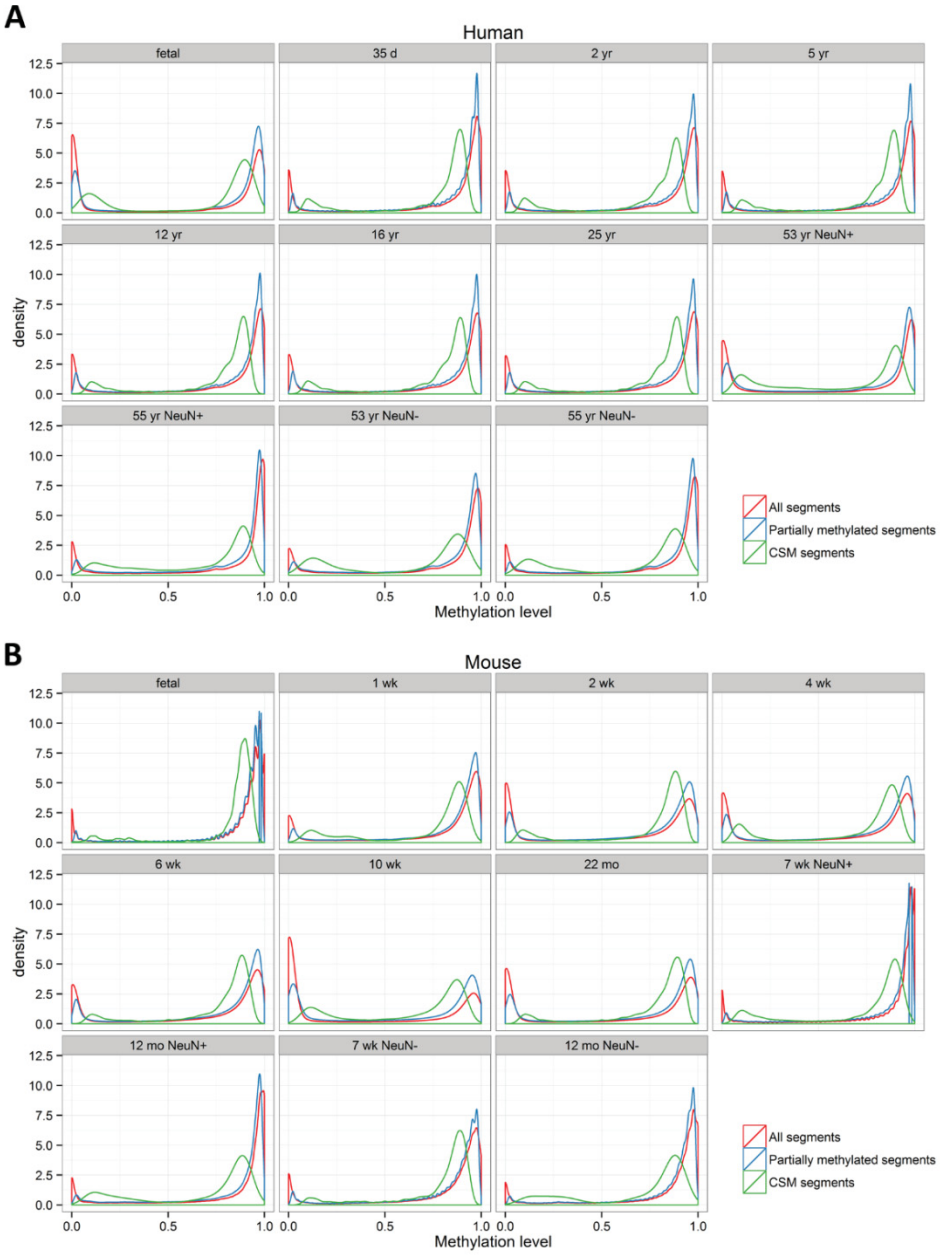
The authors thank Dr. Alexei Morozov for critical reading of the manuscript, Joseph, R. Ecker, Ryan Lister and Eran A. Mukamel for sharing brain methylome data, Bing Ren and Bradley E Bernstein for sharing ChIP-seq data, and the labs contributing to ENCODE project and NHGRI GWAS catalog. This work was supported by NIH grant NS094574, new faculty startup fund from the Biocomplexity Institute of Virginia Tech and Virginia Tech's Open Access Subvention Fund.

2.7 AUTHOR CONTRIBUTIONS

M.-a.S. performed the methylome data analysis, Z.S. and J.L. carried out hairpin bisulfite sequencing, X.W. and H.Z. performed GWAS associated study, V.R. processed human brain tissues, D.K. participated in the bisulfite sequence alignment, H.X. conceived and designed the

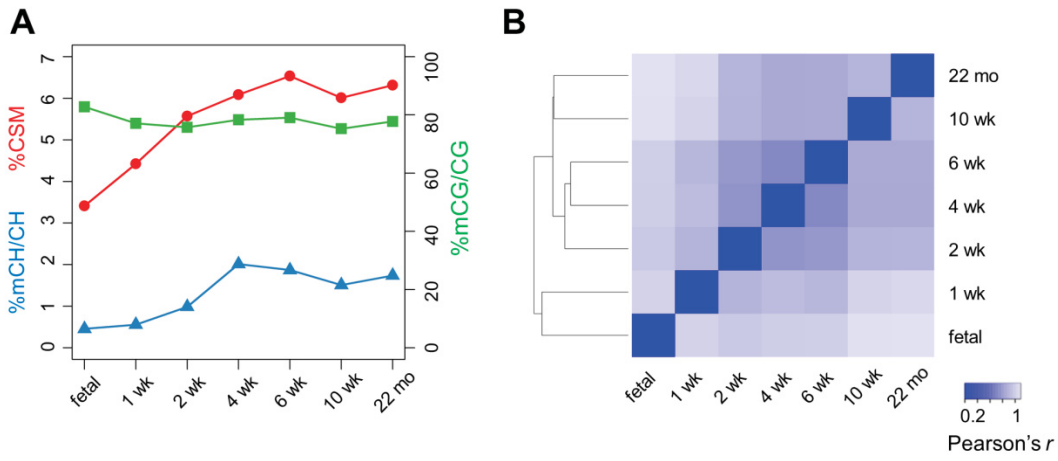
study. M.-a.S. and Z.S. contributed equally to this work. All authors drafted, read and approved the final manuscript.

2.8 SUPPLEMENTARY DATA



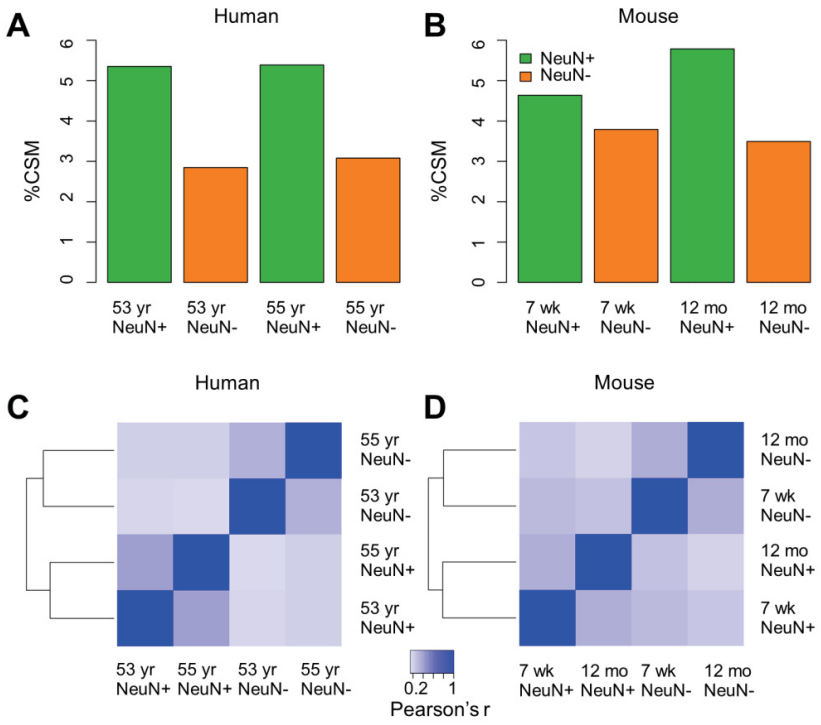
Supplementary Figure 1. The methylation level distribution for 4-CpG segments in human (a) and mouse (b).

For each methylome, the distributions of methylation levels were shown for all 4-CpG segments, partially methylated 4-CpG segments and CSM 4-CpG segments.



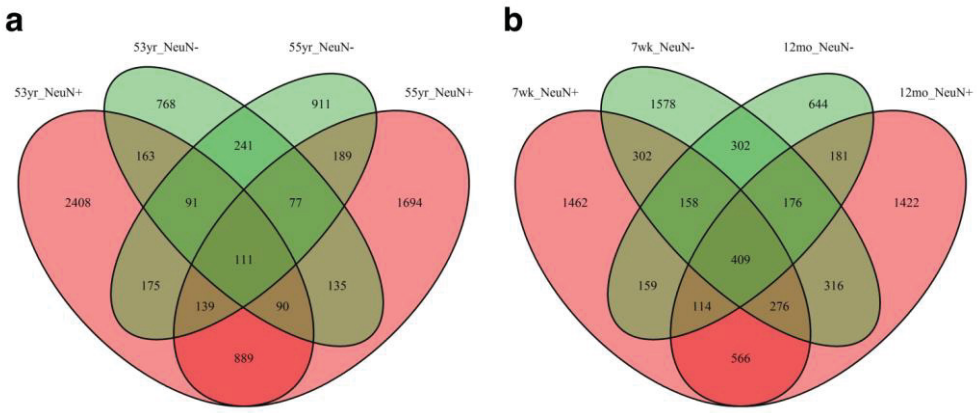
Supplementary Figure 2. Dynamic of CSM during mouse brain development.

(**a**) Changes of mC levels in CG and CH context and the percentage of CSM segments during mouse brain development. (**b**) Hierarchical clustering based on Pearson's correlations of CSM statuses predicted for all 4-CpG segments in different mouse brain methylomes.



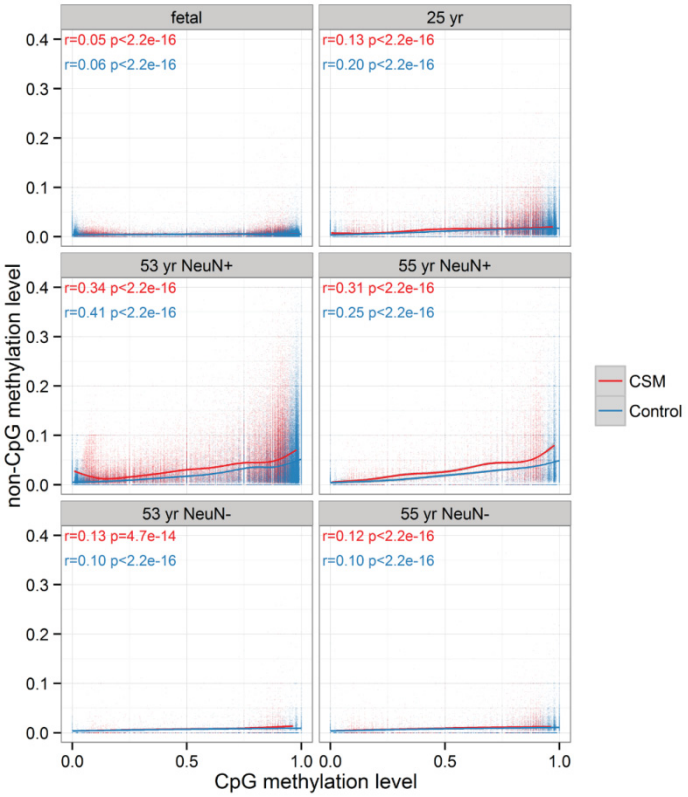
Supplementary Figure 3. Global CSM profiles in NeuN+ and NeuN- cells.

(a-b) The percentage of CSM 4-CpG segment in the methylomes of NeuN+ and NeuN- cells for human and mouse, respectively. **(c-d**) Hierarchical clustering based on Pearson's correlations of CSM statuses predicted for all 4-CpG segments in the methylomes of NeuN+ and NeuN- cells for human and mouse, respectively.



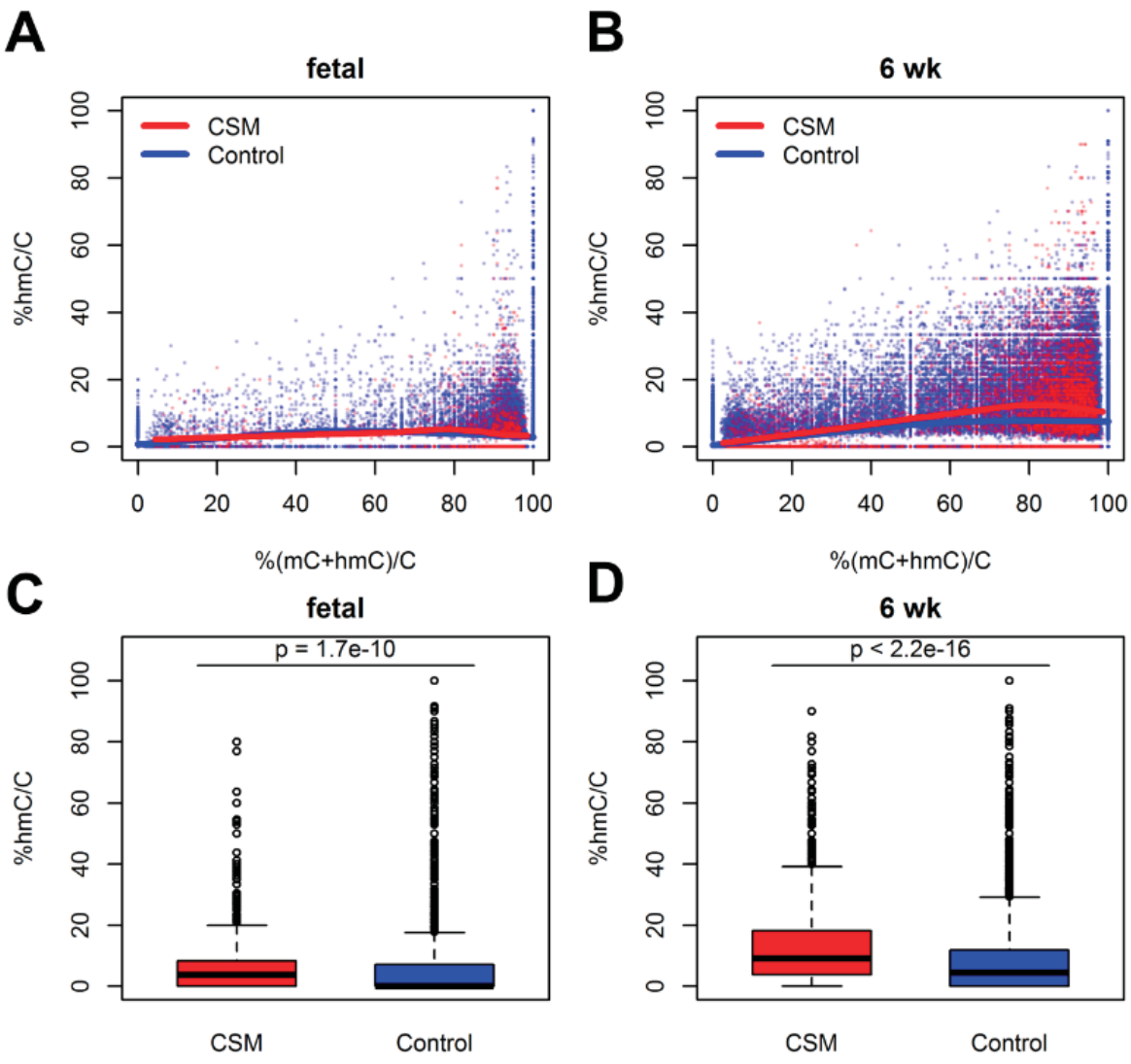
Supplementary Figure 4. Venn diagrams show the overlapping of pCSM segments between the NeuN+ and NeuN- cells.

The figures show the overlapping of pCSM segments between the purified NeuN+ and NeuN- samples for human (a) and mouse (b), respectively.



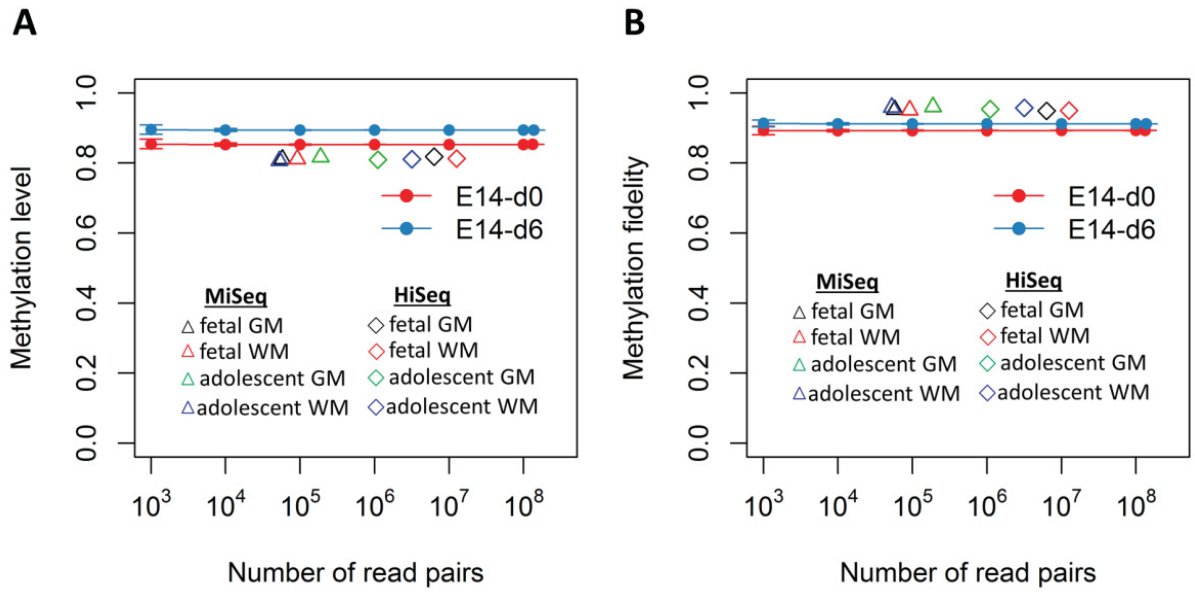
Supplementary Figure 5.Correlation between CpG and non-CpG methylation levels within CSM and control regions

The x-axis and y-axis show the mC levels in CpG and non-CpG context for each 4-CpG segment, respectively. The curves show the smoothed spline. Pearson's r and p-values are shown.



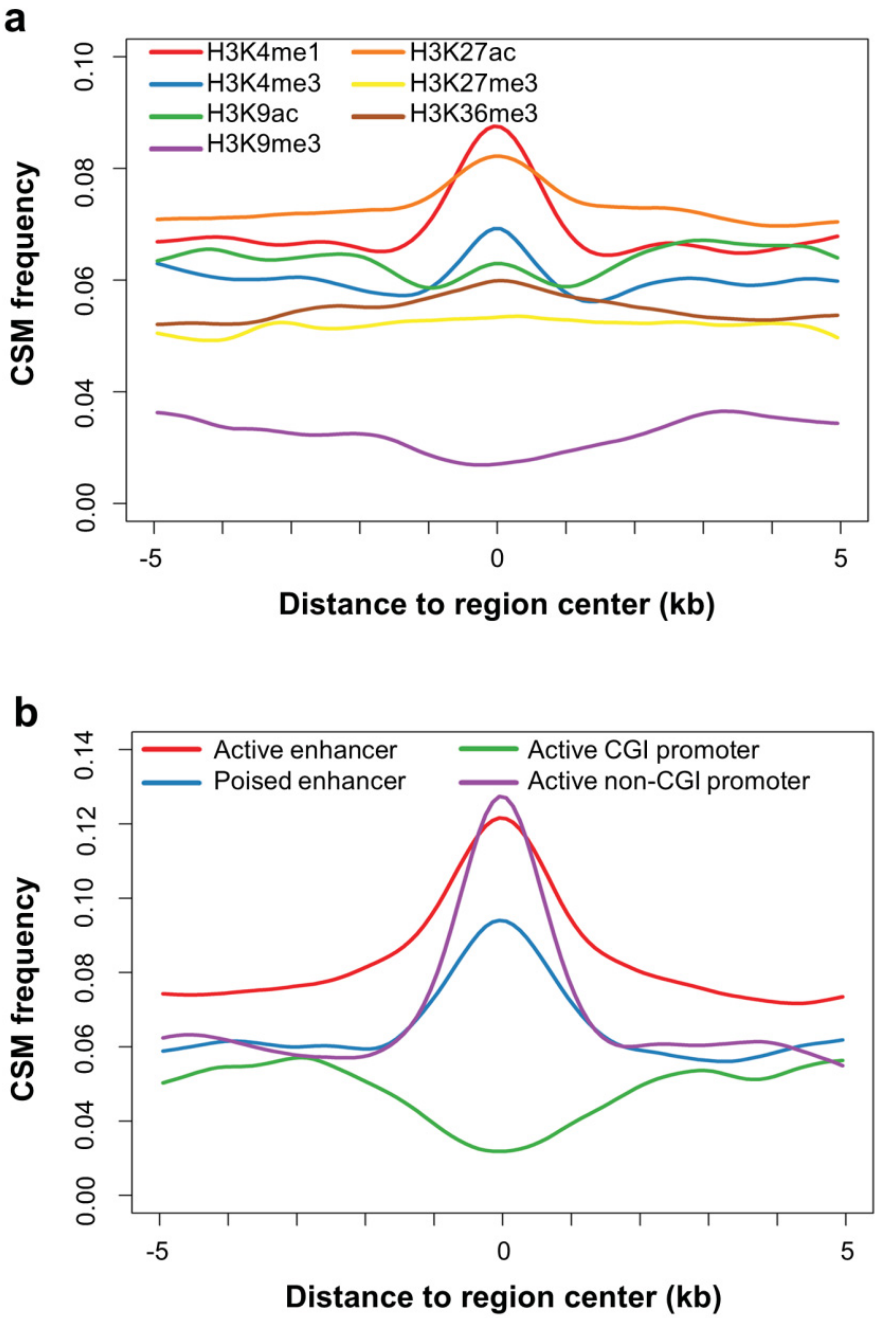
Supplementary Figure 6. 5hmC occurrence in CSM and control regions

(a-b) Scatter plots show the relationship between 5mC and 5hmC in CpG context within CSM or control regions of fetal (**a**) and 6 wk (**b**) mouse brain methylomes, respectively. The lines are the smoothed curve computed using loess method. **(c-d)** Boxplots show the distributions of 5hmC level within CSM or control regions of fetal (**c**) and 6 wk (**d**) mouse brain methylomes, respectively.



Supplementary Figure 7. Methylation level (a) and fidelity (b) estimated with hairpin bisulfite sequencing data at different sequencing depths.

Various numbers of read pairs were selected randomly from the two previously published data sets generated for mouse embryonic stem cell E14 with genome-scale hairpin bisulfite sequencing technique (Court, Tayama et al. 2014). The error bars show the standard deviation estimated from 100 times of simulations. The filled circles represent methylation level (**a**) or fidelity (**b**) determined with hairpin bisulfite sequencing data for mouse ESCs (E14-d0: mouse ESCs; E14-d6: mouse ESCs at day 6 with LIF withdrawn). In each line, the first six circles represent simulation results and the last circle shows the result using all reads generated. The open triangles and diamonds represent hairpin bisulfite sequencing data generated for human brain tissues in this study using MiSeq and HiSeq, respectively.
"fetal GM" represents human grey matter from a fetal frontal cortex.
"fetal WM" represents human white matter from a fetal frontal cortex.
"Adolescent GM" represents human grey matter from an adolescent frontal cortex.
"Adolescent WM" represents human white matter from an adolescent frontal cortex.



Supplementary Figure 8. Aggregate plots of CSM segments surrounding different histone modifications (A) and regulatory elements(B) in the adult human brain.

The regulatory elements were annotated based on the combinations of histone modifications and the presence of CGI. Active promoters are with H3K4me3 mark, Active enhancers are with H3K27ac but without H3K4me3 mark, and poised enhancers are with H3K4me1 but without H3K27ac or H3K4me3.

Supplemental Table 1. Statistics for CSM predicted from mammalian brain methylomes.

Sample	Species	#mapped reads	#4CG segment (>=10X, noXY, noASM)	#predicted CSM segments	%CSM frequency (unnormalized)	#CSM segments after merging	#CpGs covered in CSM segments	Total length of CSM segments (bp)
Hs fetal	human	718,715,354	3,575,021	41,369	1.20%	26,960	124,217	1,260,254
Hs 35 do	human	754,934,101	1,014,430	25,030	2.50%	18,375	80,778	761,814
Hs 2 yr	human	769,797,789	1,255,212	47,202	3.80%	32,822	147,195	1,413,735
Hs 5 yr	human	749,639,389	1,105,550	40,181	3.60%	28,236	126,072	1,198,411
Hs 12 yr	human	757,884,985	1,133,642	46,538	4.10%	32,216	144,736	1,377,415
Hs 16 yr	human	746,971,503	1,289,687	57,503	4.50%	39,051	176,787	1,687,696
Hs 25 yr	human	725,663,402	1,133,157	50,024	4.40%	34,366	154,856	1,455,879
Hs 53 yrNeuN+	human	841,935,270	2,210,113	96,033	4.30%	61,121	284,305	2,891,375
Hs 53 yrNeuN-	human	476,665,175	217,362	5,636	2.60%	4,276	18,583	159,715
Hs 55 yrNeuN+	human	558,556,927	471,011	21,566	4.60%	15,516	68,659	619,658
Hs 55 yrNeuN-	human	539,415,159	362,534	9,925	2.70%	7,369	32,229	282,263
Mm fetal	mouse	553,383,395	162,533	5,218	3.20%	3,821	16,771	140,672
Mm 1 wk	mouse	385,489,360	235,332	8,455	3.60%	5,441	24,994	187,504
Mm 2 wk	mouse	618,425,417	1,439,393	55,568	3.90%	39,091	174,828	1,784,255
Mm 4 wk	mouse	672,968,248	1,197,542	57,001	4.80%	40,094	179,238	1,847,952
Mm 6 wk	mouse	667,130,394	805,885	40,088	5.00%	29,524	129,877	1,310,969
Mm 10 wk	mouse	448,962,696	797,566	23,794	3.00%	17,580	77,231	714,279
Mm 22 mo	mouse	682,097,075	944,270	51,795	5.50%	33,086	152,841	1,249,124
Mm 7 wkNeuN+	mouse	633,803,213	357,027	15,569	4.40%	9,864	45,567	383,268
Mm 7 wkNeuN-	mouse	676,970,294	587,176	20,815	3.50%	13,900	63,013	552,001
Mm 12 moNeuN+	mouse	436,577,319	194,533	9,765	5.00%	6,951	30,825	253,737
Mm 12 moNeuN-	mouse	332,614,011	102,534	3,343	3.30%	2,166	9,927	70,782

Supplementary Table 2. Pearson's correlations of CSM statuses predicted for all 4-CpG segments in the methylomes of different samples.

a. human developmental stages (related to Figure 1E).

	fetal	35d	2yr	5yr	12yr	16yr	25yr
fetal		0.20	0.16	0.15	0.15	0.15	0.15
35d	0.20		0.23	0.21	0.21	0.20	0.20
2y	0.16	0.23		0.31	0.30	0.31	0.30
5y	0.15	0.21	0.31		0.31	0.31	0.30
12y	0.15	0.21	0.30	0.31		0.32	0.31
16y	0.15	0.20	0.31	0.31	0.32		0.33
25y	0.15	0.20	0.30	0.30	0.31	0.33	

b. mouse developmental stages (related to Supplemental Figure 2B).

	fetal	1wk	2wk	4wk	6wk	10wk	22mo
fetal		0.18	0.21	0.19	0.19	0.12	0.11
1wk	0.18		0.29	0.26	0.28	0.17	0.15
2wk	0.21	0.29		0.41	0.39	0.29	0.29
4wk	0.19	0.26	0.41		0.45	0.34	0.33
6wk	0.19	0.28	0.39	0.45		0.33	0.33
10wk	0.12	0.17	0.29	0.34	0.33		0.29
22mo	0.11	0.15	0.29	0.33	0.33	0.29	

c. human cell types (related to Supplemental Figure 3C).

	53yr NeuN+	55yr NeuN+	53yr NeuN-	55yr NeuN-
53y NeuN+		0.37	0.16	0.19
55y NeuN+	0.37		0.15	0.19
53y NeuN-	0.16	0.15		0.30
55y NeuN-	0.19	0.19	0.30	

d. mouse cell types (related to Supplemental Figure 3D).

	7wk NeuN+	12mo NeuN+	7wk NeuN-	12mo NeuN-
7wk NeuN+		0.31	0.27	0.22
12mo NeuN+	0.31		0.24	0.18
7wk NeuN-	0.27	0.24		0.32
12mo NeuN-	0.22	0.18	0.32	

Supplemental Table 3a. Summary of CSM predicted for human and mouse purified cells.

Human (54,276 common 10X 4CG segment)		**Mouse (36,692 common 10X 4CG segment)**	
Sample	#pCSM segment	Sample	#pCSM segment
53yr NeuN+	4066	7wk NeuN+	3446
55yr NeuN+	3324	12mo NeuN+	3460
53yr NeuN-	1676	7wk NeuN-	3517
55yr NeuN-	1934	12mo NeuN-	2143

Supplemental Table 3b. Number of shared pCSM segments between different samples.

Human				
	53y_ NeuN+	55y_ NeuN+	53y_NeuN-	55y_NeuN-
53y_ NeuN+		1229	455	516
55y_ NeuN+	37%		413	516
53y_NeuN-	27%	25%		520
55y_NeuN-	27%	27%	31%	

Mouse				
	7wk_ NeuN+	12mo_ NeuN+	7wk_NeuN-	12mo_NeuN-
7wk_ NeuN+		1365	1145	840
12mo_ NeuN+	40%		1177	880
7wk_NeuN-	33%	34%		1045
12mo_NeuN-	39%	41%	49%	

Supplemental Table 4. Statistics of the hairpin-bisulfite sequencing data.

Sample	Age	Tissue/cell type	Uniquely mapped read pairs (M)	Bisulfite conversion rate (%)	Genome covered (%)	CpG covered (%)	Genome sequencing depth for genome covered (X)	Average methylation level (%)	Average methylation fidelity (%)
fetal CG	22 weeks gestation	Cortical grey matter	6.4	99.8	8.2	9.7	10	81.8	94.9
fetal CW	22 weeks gestation	Cortical white matter	12.7	99.8	15	17.7	10.9	81.3	95
17y CG	17 years	Cortical grey matter	1.3	99.8	2.6	3.1	6.5	81.1	95.6
17y CW	17years	Cortical white matter	3.2	99.7	4.7	5.2	8.9	81.1	95.8

Supplemental Table 5a. GO analysis for CSM associated genes in human brain methylomes. Bonferroni adjusted p-values are indicated.

Category	Term	fetal	35d	2y	5y	12y	16y	25y	53y NeuN+	55y NeuN+	53y NeuN-	55y NeuN-
GOTERM_BP_FAT	GO:0030182~neuron differentiation	1	0.003668	1.27E-07	3.04E-08	6.12E-09	1.43E-08	2.66E-11	1.17E-07	0.000145	1	0.060787
GOTERM_BP_FAT	GO:0007155~cell adhesion	0.999798	0.818608	0.081253	0.000769	0.001069	0.984857	0.672019	1	0.040015	9.34E-10	2.88E-05
GOTERM_BP_FAT	GO:0022610~biological adhesion	0.999888	0.839425	0.092534	0.00009	0.001258	0.990978	0.716109	1	0.040015	9.34E-10	3.16E-05
GOTERM_BP_FAT	GO:0006357~regulation of transcription from RNA polymerase II promoter	0.912495	6.3E-05	1.29E-05	8.05E-06	1.75E-09	2.8E-08	4.92E-07	0.000228	0.000668	0.466876	0.012903
GOTERM_BP_FAT	GO:0048598~embryonic morphogenesis	0.985051	0.227799	0.002688	0.504055	2.03E-09	0.526055	0.039627	0.034303	1	1	0.442114
GOTERM_BP_FAT	GO:0007156~homophilic cell adhesion	0.042016	1.53E-05	1	1	0.291043	1	1	1	0.005455	5E-09	7.58E-08
GOTERM_BP_FAT	GO:0048562~embryonic organ morphogenesis	0.393794	0.034805	0.67924	0.376796	1.63E-08	0.995656	0.126355	0.637947	1	1	0.986188
GOTERM_BP_FAT	GO:0016337~cell-cell adhesion	0.067533	0.002648	1	0.980075	0.996	1	1	1	0.623224	5.16E-08	6.54E-06
GOTERM_BP_FAT	GO:0031328~positive regulation of cellular biosynthetic process	1	3.1E-05	0.001015	0.003819	2.18E-06	1.37E-07	2.96E-05	0.00106	0.004871	1	0.315097
GOTERM_BP_FAT	GO:0051173~positive regulation of nitrogen compound metabolic process	1	8.43E-06	0.003772	0.003302	1.33E-05	2.37E-07	5.29E-05	0.00785	0.017502	1	0.273181
GOTERM_BP_FAT	GO:0009891~positive regulation of biosynthetic process	1	4.77E-05	0.00136	0.006909	4.71E-06	2.49E-07	4.49E-05	0.000641	0.003989	1	0.423249
GOTERM_BP_FAT	GO:0010557~positive regulation of macromolecule biosynthetic process	1	0.000169	0.000679	0.009218	1.31E-06	3.42E-07	0.000295	0.000539	0.002003	1	0.606859
GOTERM_BP_FAT	GO:0045941~positive regulation of transcription	1	0.000108	0.004825	0.047159	3.71E-06	5.48E-07	0.00101	0.005086	0.001141	1	0.081331
GOTERM_BP_FAT	GO:0032989~cellular component morphogenesis	1	0.002398	7.16E-06	0.020753	6.15E-07	0.066436	0.000633	0.028985	3.12E-05	1	0.002392
GOTERM_BP_FAT	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1	5.53E-06	0.01073	0.015573	4.58E-06	7.08E-07	0.000469	0.00367	0.007402	1	0.227363
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	1	0.028116	0.018119	6.73E-05	8.88E-06	8.06E-07	0.001852	6.22E-06	0.512889	0.994076	0.898678
GOTERM_BP_FAT	GO:0048568~embryonic organ development	0.976844	0.001386	0.116507	0.085035	9.11E-07	0.835345	0.195324	0.803198	1	1	0.967088
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	1	0.087079	0.015647	4.1E-05	9.55E-06	1E-06	0.005964	1.28E-05	0.594936	0.999951	0.923844
GOTERM_BP_FAT	GO:0010628~positive regulation of gene expression	1	0.000104	0.005188	0.034948	6.65E-06	1.06E-06	0.000823	0.004624	0.002501	1	0.048922
GOTERM_BP_FAT	GO:0051254~positive regulation of RNA metabolic process	1	0.00131	0.000971	0.003745	1.16E-06	1.72E-05	0.000921	0.00192	0.019521	1	0.013012
GOTERM_BP_FAT	GO:0000902~cell morphogenesis	1	0.001105	3.36E-06	0.003574	1.91E-06	0.008185	0.000013	0.168907	2.54E-06	1	0.000901
GOTERM_BP_FAT	GO:0030030~cell projection organization	1	0.141269	1.98E-06	0.020916	3.72E-05	3.75E-05	9.26E-06	0.097241	0.000307	1	7.79E-05
GOTERM_BP_FAT	GO:0007389~pattern specification process	1	0.111928	0.366446	0.061265	2.02E-06	0.803125	0.080633	0.021019	1	0.998786	0.291476
GOTERM_BP_FAT	GO:0048666~neuron development	1	0.344752	6.86E-05	6.99E-05	0.000056	2.52E-06	7.33E-06	0.005593	0.008879	1	0.592635
GOTERM_BP_FAT	GO:0045893~positive regulation of transcription, DNA-dependent	1	0.000714	0.000724	0.00798	2.96E-06	1.09E-05	0.000657	0.001717	0.012023	1	0.03165
GOTERM_BP_FAT	GO:0031175~neuron projection development	1	0.270487	3.96E-06	0.017897	0.009288	3.01E-06	1.76E-05	0.054603	0.00221	1	0.003548
GOTERM_BP_FAT	GO:0009792~embryonic development ending in birth or egg hatching	0.943173	0.006801	0.022865	0.008463	6.4E-06	0.060819	0.008248	0.01928	1	1	1
GOTERM_BP_FAT	GO:0032990~cell part morphogenesis	1	0.015874	3.8E-05	0.395832	7.98E-06	0.000042	0.000236	0.693036	0.001824	1	0.000693
GOTERM_BP_FAT	GO:0043009~chordate embryonic development	0.987562	0.004152	0.034837	0.007981	1.17E-05	0.051892	0.007126	0.0228	1	1	1
GOTERM_BP_FAT	GO:0010604~positive regulation of macromolecule metabolic process	1	0.000464	0.000203	0.019161	1.37E-05	0.000116	0.001424	0.001186	0.094979	1	0.37009
GOTERM_BP_FAT	GO:0048858~cell projection morphogenesis	1	0.007714	2.21E-05	0.149817	1.43E-05	0.000378	0.000259	0.317057	0.000192	1	0.000333
GOTERM_BP_FAT	GO:0003002~regionalization	1	0.138338	0.387085	0.737723	1.67E-05	0.995808	0.095828	0.083656	1	1	0.76314
GOTERM_BP_FAT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	1	0.001514	0.015778	0.098865	2.1E-05	0.000392	0.000171	0.027135	0.074538	1	0.061108
GOTERM_BP_FAT	GO:0048812~neuron projection morphogenesis	1	0.009419	3.68E-05	0.01581	5E-05	4.49E-05	2.45E-05	0.140662	0.001417	1	0.000473
GOTERM_BP_FAT	GO:0030900~forebrain development	0.450084	0.000669	0.004344	0.008036	0.007696	0.002875	3.45E-05	0.110036	0.062851	1	0.99363
GOTERM_BP_FAT	GO:0043583~ear development	1	1	1	0.226407	3.5E-05	0.997806	1	1	1	1	1

GOTERM_BP_FAT	GO:0000904~cell morphogenesis involved in differentiation	1	0.001468	0.000705	0.016276	5.02E-05	0.000556	0.000802	0.749313	6.6E-05	1	0.010497
GOTERM_BP_FAT	GO:0048667~cell morphogenesis involved in neuron differentiation	1	0.002832	0.000471	0.009275	7.15E-05	0.000106	6.11E-05	0.129362	0.000912	1	0.006057
GOTERM_BP_FAT	GO:0045892~negative regulation of transcription, DNA-dependent	1	0.961864	0.000901	0.153299	0.001134	6.24E-05	0.055493	0.023795	0.338082	0.966484	0.999147
GOTERM_BP_FAT	GO:0051253~negative regulation of RNA metabolic process	1	0.982697	0.000064	0.163806	0.002214	0.000184	0.110547	0.046687	0.394358	0.990738	0.998345
GOTERM_BP_FAT	GO:0051056~regulation of small GTPase mediated signal transduction	0.003939	0.15937	0.000026	0.000198	0.000469	0.008817	0.003747	1	1	1	0.300085
GOTERM_BP_FAT	GO:0030036~actin cytoskeleton organization	1	0.999988	0.000705	0.027215	0.002195	0.411791	0.037415	0.000254	0.267967	1	0.004486
GOTERM_BP_FAT	GO:0007409~axonogenesis	1	0.010579	0.000742	0.083189	0.001446	0.001023	0.000826	0.674077	0.000384	1	0.001752
GOTERM_BP_FAT	GO:0010629~negative regulation of gene expression	1	0.438533	0.000671	0.684573	0.276782	0.000458	0.109975	0.032333	0.743879	0.999681	1
GOTERM_BP_FAT	GO:0016481~negative regulation of transcription	1	0.726167	0.025783	0.509642	0.464579	0.000611	0.119101	0.228696	0.798685	0.926267	1
GOTERM_BP_FAT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	1	1	0.022635	0.636415	0.000213	0.000629	0.274623	0.250272	0.493022	0.937503	0.046084
GOTERM_BP_FAT	GO:0007242~intracellular signaling cascade	1	1	0.001076	0.079131	0.001115	0.000667	0.657656	0.473567	1	1	1
GOTERM_BP_FAT	GO:0030029~actin filament-based process	1	0.998402	0.007078	0.359975	0.001712	0.535124	0.253275	0.001015	0.166299	1	0.032172
GOTERM_BP_FAT	GO:0048839~inner ear development	1	1	0.998821	0.200932	0.001298	0.999997	0.999995	1	0.999656	1	1
GOTERM_BP_FAT	GO:0045449~regulation of transcription	0.995496	0.053711	0.003659	0.018811	0.029527	0.001871	0.001499	0.001639	0.529952	0.989526	0.887139
GOTERM_BP_FAT	GO:0035023~regulation of Rho protein signal transduction	0.134168	1	0.207915	0.034279	0.99668	0.332817	0.002138	1	1	1	1
GOTERM_BP_FAT	GO:0006928~cell motion	1	0.012873	0.608543	0.842254	0.002486	0.224588	0.143479	0.999984	0.25527	1	1
GOTERM_BP_FAT	GO:0045934~negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1	0.980871	0.01106	0.985741	0.463076	0.002539	0.657019	0.251729	0.976668	0.985902	1
GOTERM_BP_FAT	GO:0051172~negative regulation of nitrogen compound metabolic process	1	0.950698	0.006828	0.997458	0.388247	0.002767	0.449347	0.371167	0.996049	0.957469	1
GOTERM_BP_FAT	GO:0045165~cell fate commitment	1	0.015777	0.090509	0.999885	0.003007	0.379813	0.136377	0.024735	0.017291	0.848959	0.831491
GOTERM_BP_FAT	GO:0007010~cytoskeleton organization	1	1	0.005981	0.037914	0.007613	1	0.01636	0.123489	0.661477	1	0.029837
GOTERM_BP_FAT	GO:0046578~regulation of Ras protein signal transduction	0.033289	0.94371	0.046164	0.006002	0.703566	0.826555	0.14585	1	1	1	0.999986
GOTERM_BP_FAT	GO:0031327~negative regulation of cellular biosynthetic process	1	0.88103	0.006336	0.99712	0.034573	0.033416	0.277297	0.745081	0.997844	0.997645	1
GOTERM_BP_FAT	GO:0007423~sensory organ development	1	1	1	0.836565	0.00933	1	0.99374	1	1	1	1
GOTERM_MF_FAT	GO:0043565~sequence-specific DNA binding	0.092255	0.001069	0.00017	3.98E-09	4.3E-11	2.28E-14	2.81E-11	2.99E-11	0.927374	0.351749	0.106717
GOTERM_MF_FAT	GO:0003700~transcription factor activity	0.01161	1.62E-05	0.000388	2.07E-10	9.66E-10	9.44E-10	2.78E-10	5.46E-10	0.609878	0.418755	0.000328
GOTERM_MF_FAT	GO:0008092~cytoskeletal protein binding	0.999998	0.000415	8.18E-06	0.015441	7.01E-09	0.988795	4.33E-07	0.147963	0.029578	1	0.071346
GOTERM_MF_FAT	GO:0030528~transcription regulator activity	0.28533	6.32E-05	5.37E-05	3.54E-06	2.5E-08	3.03E-06	3.82E-06	2.29E-05	0.219894	0.441981	0.495927
GOTERM_MF_FAT	GO:0003779~actin binding	0.999937	0.022381	0.000102	0.913165	1.48E-05	0.972776	7.61E-06	0.253425	0.047669	1	0.951775
GOTERM_MF_FAT	GO:0005509~calcium ion binding	1	0.299189	0.105932	0.088565	0.507222	0.17591	0.54918	0.652358	4.19E-05	0.229797	0.99696
GOTERM_MF_FAT	GO:0030695~GTPase regulator activity	0.003005	0.001195	9.67E-05	0.000728	0.009869	0.010608	0.016399	1	1	0.993429	0.081512
GOTERM_MF_FAT	GO:0005085~guanyl-nucleotide exchange factor activity	0.006858	0.378217	0.002575	0.000307	0.071024	0.154262	0.042852	1	0.040861	0.999647	0.000196
GOTERM_MF_FAT	GO:0005083~small GTPase regulator activity	0.070062	0.572844	0.004009	0.000214	0.007726	0.553969	0.752381	1	1	1	0.396679
GOTERM_MF_FAT	GO:0060589~nucleoside-triphosphatase regulator activity	0.005161	0.000838	0.000427	0.001466	0.027194	0.013565	0.014268	1	1	0.990343	0.154296
GOTERM_MF_FAT	GO:0005088~Ras guanyl-nucleotide exchange factor activity	0.007285	0.906855	0.09469	0.000665	0.6359	0.24303	0.398484	1	0.99947	1	0.006872
GOTERM_MF_FAT	GO:0046872~metal ion binding	1	1	1	0.997909	0.983092	0.566292	1	0.004117	0.736402	1	1
GOTERM_MF_FAT	GO:0051020~GTPase binding	0.998474	1	0.004357	0.048597	0.999147	1	0.970068	1	1	1	0.335056
GOTERM_MF_FAT	GO:0043169~cation binding	1	1	1	0.999998	0.993691	0.886842	1	0.005232	0.777497	1	1
GOTERM_MF_FAT	GO:0005089~Rho guanyl-nucleotide exchange factor activity	0.010085	1	0.715279	0.005719	0.969842	0.845939	0.126195	1	1	1	0.157347
GOTERM_MF_FAT	GO:0003677~DNA binding	0.612701	0.994854	0.941214	0.360967	0.918315	0.048891	0.081667	0.00875	1	1	0.720504

GOTERM_MF_FAT	GO:0017016~Ras GTPase binding	1	1	0.000 9115	0.36 572	1	1	0.99 9999	1	1	1	0.999 983
GOTERM_MF_FAT	GO:0031267~small GTPase binding	1	1	0.00 9932	0.02 2576	0.99 9398	1	0.99 8768	1	1	1	0.665 716
GOTERM_CC_FAT	GO:0044459~plasma membrane part	0.40 7777	0.04 1658	0.00 0481	0.00 1517	1.16 E-05	6.42 E-06	1.4E-09	0.000 136	0.108 267	0.943 267	0.994 262
GOTERM_CC_FAT	GO:0042995~cell projection	0.99 8505	0.94 7856	0.02 5037	0.00 0458	2.04 E-08	0.00 1371	7.37 E-05	0.008 612	0.473 695	1	0.007 126
GOTERM_CC_FAT	GO:0005886~plasma membrane	1	0.00 1812	0.00 1227	5.02 E-05	0.00 0576	0.00 3839	1.1E-05	1	0.013 067	0.001 934	0.016 96
GOTERM_CC_FAT	GO:0031252~cell leading edge	1	1	0.01 2673	0.01 342	1.11 E-05	0.27 0394	0.00 0332	1	0.999 937	0.923 035	0.010 321
GOTERM_CC_FAT	GO:0015629~actin cytoskeleton	0.97 4026	0.00 0113	0.00 3151	0.01 2589	1.15 E-05	0.99 9343	0.06 6103	0.199 906	0.991 886	0.999 718	0.610 729
GOTERM_CC_FAT	GO:0070161~anchoring junction	0.99 9999	1	0.62 6428	0.23 2271	1.61 E-05	1	0.01 5806	0.999 985	1	1	0.329 768
GOTERM_CC_FAT	GO:0005912~adherens junction	0.99 9994	0.99 9916	0.05 7315	0.43 2369	5.56 E-05	1	0.00 0597	0.985 495	1	0.997 607	0.704 59
GOTERM_CC_FAT	GO:0043005~neuron projection	1	1	0.99 4538	0.00 6105	0.00 0106	0.00 0476	0.00 0161	0.034 353	0.559 676	1	1
GOTERM_CC_FAT	GO:0005856~cytoskeleton	0.55 9133	0.00 0211	0.04 8487	0.12 6774	0.00 4282	1	0.01 0426	0.016 756	1	1	0.637 98
GOTERM_CC_FAT	GO:0031226~intrinsic to plasma membrane	0.99 9996	0.99 9915	0.86 1834	1	0.99 9946	0.00 4372	0.00 9343	0.990 691	0.999 39	1	1
GOTERM_CC_FAT	GO:0001726~ruffle	1	1	0.04 9609	0.08 4498	0.00 6542	0.99 5862	0.16 3973	1	0.943 787	1	0.023 915
GOTERM_CC_FAT	GO:0030054~cell junction	0.16 0824	1	0.43 2586	0.22 904	0.28 3948	1	0.00 8732	0.147 74	1	1	0.999 985
GOTERM_CC_FAT	GO:0005887~integral to plasma membrane	1	0.99 9968	0.85 9517	1	1	0.00 913	0.03 1307	0.998 587	0.996 976	1	1

Supplemental Table 5b. GO analysis for CSM associated genes in mouse brain methylomes. Bonferroni adjusted p-values are indicated.

Category	Term	fetal	1wk	2wk	4wk	6wk	10wk	22mo	7wk NeuN+	12mo NeuN+	7wk NeuN-	12mo NeuN-
GOTERM_BP_FAT	GO:0006357~regulation of transcription from RNA polymerase II promoter	1	1	0.000 0136	3.51E-12	0.000 00191	0.037 30834	2.77E-09	0.034 69611	1	1	0.6968 7388
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	1	1	0.048 23688	6.47E-09	0.664 5714	0.013 7196	0.000 0288	0.995 71048	1	0.999 99995	0.9996 4
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	1	1	0.029 24705	2.74E-08	0.614 00667	0.018 35816	0.000 0188	0.993 49774	1	0.999 99949	0.9962 5616
GOTERM_BP_FAT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	1	1	0.000 00982	1.54E-07	0.002 34425	0.371 65561	0.012 73511	0.999 29161	1	1	1
GOTERM_BP_FAT	GO:0030030~cell projection organization	1	1	0.000 16	1.85E-07	0.004 40408	0.000 00344	0.003 11638	0.000 188	1	0.898 37613	1
GOTERM_BP_FAT	GO:0031328~positive regulation of cellular biosynthetic process	1	1	0.000 00574	2.29E-07	0.115 3873	0.999 96499	0.040 30429	1	1	1	0.9999 9598
GOTERM_BP_FAT	GO:0010628~positive regulation of gene expression	1	1	0.000 0255	2.75E-07	0.070 52883	0.769 34305	0.010 46978	1	1	1	1
GOTERM_BP_FAT	GO:0048568~embryonic organ development	1	1	0.000 0375	0.001 62536	0.010 10744	0.001 78811	3.77E-07	1	1	0.999 87055	1
GOTERM_BP_FAT	GO:0045941~positive regulation of transcription	1	1	0.000 01	4.07E-07	0.129 37878	0.973 81298	0.024 92542	1	1	1	1
GOTERM_BP_FAT	GO:0048598~embryonic morphogenesis	1	1	5.26E-07	0.000 00728	0.152 34801	0.006 11266	0.000 0268	1	1	0.878 66169	1
GOTERM_BP_FAT	GO:0048562~embryonic organ morphogenesis	1	1	0.051 96454	0.000 735	0.966 26782	0.039 48178	6.44E-07	1	1	1	1
GOTERM_BP_FAT	GO:0030182~neuron differentiation	1	1	0.000 0952	0.000 00088	0.000 355	6.84E-07	0.000 0406	0.024 20354	1	0.998 4906	1
GOTERM_BP_FAT	GO:0045165~cell fate commitment	1	1	0.008 7582	0.000 00434	1	0.000 853	7.91E-07	1	1	1	1
GOTERM_BP_FAT	GO:0009891~positive regulation of biosynthetic process	1	1	0.000 0104	9.38E-07	0.143 93803	0.999 98115	0.083 21277	1	1	1	0.9999 9921
GOTERM_BP_FAT	GO:0010557~positive regulation of macromolecule biosynthetic process	1	1	0.000 00585	0.000 00214	0.050 31232	0.998 09526	0.132 33503	1	1	1	0.9999 9995
GOTERM_BP_FAT	GO:0035023~regulation of Rho protein signal transduction	1	1	0.000 0495	0.000 00253	0.000 122	0.107 86495	0.266 49996	0.859 34817	1	1	1
GOTERM_BP_FAT	GO:0051254~positive regulation of RNA metabolic process	1	1	0.000 0578	0.000 00387	0.023 08443	0.554 91714	0.020 43549	1	1	1	1
GOTERM_BP_FAT	GO:0051173~positive regulation of nitrogen compound metabolic process	1	1	0.000 0097	0.000 00396	0.273 1382	0.999 99996	0.043 28349	1	1	1	1
GOTERM_BP_FAT	GO:0030900~forebrain development	1	1	0.147 84522	0.018 66507	0.237 43616	0.000 00446	1	0.760 77788	1	0.186 19935	1

GOTERM_BP_FAT	GO:0051056~regulation of small GTPase mediated signal transduction	1	1	0.000133	0.00000478	0.00000862	0.000789	0.001 31495	0.778 51662	0.5650 0361	0.736 10634	1
GOTERM_BP_FAT	GO:0045449~regulation of transcription	1	1	0.278 76429	0.00000566	1	0.999 77327	0.035 83804	1	1	1	0.9999 9995
GOTERM_BP_FAT	GO:0000902~cell morphogenesis	1	1	0.000 0321	0.000033	0.001 19395	0.000 00705	0.000 497	0.551 72766	1	0.881 91868	1
GOTERM_BP_FAT	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1	1	0.000 00949	0.00000794	0.206 02657	0.998 64411	0.036 21397	1	1	1	1
GOTERM_BP_FAT	GO:0031175~neuron projection development	1	1	0.007 97294	0.018 78602	0.358 80559	0.000 00837	0.051 71247	0.007 88448	1	0.935 83751	1
GOTERM_BP_FAT	GO:0045893~positive regulation of transcription, DNA-dependent	1	1	0.000 0492	0.00000101	0.037 74497	0.596 68151	0.019 22309	1	1	1	1
GOTERM_BP_FAT	GO:0051253~negative regulation of RNA metabolic process	1	1	1	0.001 21664	0.020 85781	0.512 61821	0.000 0162	0.999 99846	1	1	1
GOTERM_BP_FAT	GO:0010604~positive regulation of macromolecule metabolic process	1	1	0.000 0165	0.000374	0.025 52884	0.959 20803	0.430 79814	1	1	1	1
GOTERM_BP_FAT	GO:0046578~regulation of Ras protein signal transduction	1	1	0.000 84	0.00000186	0.000287	0.015 34605	0.188 80303	0.502 59221	0.8138 7829	0.999 98952	1
GOTERM_BP_FAT	GO:0045892~negative regulation of transcription, DNA-dependent	1	1	1	0.002 27831	0.014 14306	0.603 43917	0.000 0186	1	1	1	1
GOTERM_BP_FAT	GO:0032989~cellular component morphogenesis	1	1	0.000 239	0.000151	0.030 91756	0.000019	0.000 378	0.551 01108	1	0.999 91037	1
GOTERM_BP_FAT	GO:0042127~regulation of cell proliferation	1	1	0.000 0267	0.001 67989	0.255 27466	0.099 22981	0.003 74376	1	1	1	1
GOTERM_BP_FAT	GO:0001501~skeletal system development	1	1	0.872 33541	0.001 0373	0.278 52268	0.996 62023	0.000 0309	1	1	0.978 20595	1
GOTERM_BP_FAT	GO:0051172~negative regulation of nitrogen compound metabolic process	1	1	1	0.066 66748	0.920 37855	0.998 9017	0.000 0385	0.999 99975	1	1	1
GOTERM_BP_FAT	GO:0006928~cell motion	1	0.999 99576	0.000 117	0.00000396	0.001 24733	0.020 19604	0.945 35178	0.841 43151	1	0.999 99992	1
GOTERM_BP_FAT	GO:0045934~negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1	1	1	0.043 57484	0.860 38844	0.999 99469	0.000 0441	0.999 99863	1	1	1
GOTERM_BP_FAT	GO:0048666~neuron development	1	1	0.000 731	0.000528	0.001 25403	0.000 0504	0.010 89479	0.018 07112	1	0.999 98438	1
GOTERM_BP_FAT	GO:0030036~actin cytoskeleton organization	0.923 81054	1	0.000 0545	0.796 97011	0.964 44501	0.086 39962	0.006 89933	1	1	0.999 99999	1
GOTERM_BP_FAT	GO:0007242~intracellular signaling cascade	1	1	0.217 78307	0.000109	0.028 40076	0.222 63034	0.000 0712	0.999 99708	1	1	1
GOTERM_BP_FAT	GO:0031327~negative regulation of cellular biosynthetic process	1	1	1	0.023 17665	0.185 07593	0.999 95488	0.000 0754	1	1	1	1
GOTERM_BP_FAT	GO:0007423~sensory organ development	1	1	0.324 36847	0.00000921	0.278 84263	0.000245	0.000 246	0.944 93599	1	0.851 82166	1
GOTERM_BP_FAT	GO:0016481~negative regulation of transcription	1	1	1	0.029 87363	0.733 23293	0.999 71905	0.000 0936	1	1	1	1
GOTERM_BP_FAT	GO:0022610~biological adhesion	1	0.019 77393	0.178 56808	0.002 30698	0.000108	0.718 60104	0.005 27361	1	0.6744 597	0.086 54606	1
GOTERM_BP_FAT	GO:0007155~cell adhesion	1	0.019 77393	0.178 56808	0.002 30698	0.000108	0.718 60104	0.005 27361	1	0.6744 597	0.086 54606	1
GOTERM_BP_FAT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	1	1	1	0.006 50442	0.033 35184	0.999 96922	0.000 14	0.993 79179	1	1	1
GOTERM_BP_FAT	GO:0030029~actin filament-based process	0.990 06564	1	0.000 151	0.997 04018	0.999 99948	0.330 86506	0.022 30148	0.999 99998	1	1	1
GOTERM_BP_FAT	GO:0009890~negative regulation of biosynthetic process	1	1	1	0.008 50239	0.301 34378	0.999 99926	0.000 184	1	1	1	1
GOTERM_BP_FAT	GO:0048663~neuron fate commitment	1	1	1	0.000259	1	0.999 9242	0.462 64028	1	1	1	1
GOTERM_BP_FAT	GO:0001525~angiogenesis	1	1	0.000 266	0.287 6649	0.006 88587	0.999 96318	0.043 11551	1	1	1	1
GOTERM_BP_FAT	GO:0048858~cell projection morphogenesis	1	1	0.014 56906	0.000314	0.345 28961	0.000838	0.338 89904	0.044 58724	1	0.953 88207	1
GOTERM_BP_FAT	GO:0048514~blood vessel morphogenesis	1	1	0.001 37776	0.000321	0.015 84411	0.060 84972	0.018 19585	1	1	1	1
GOTERM_BP_FAT	GO:0007389~pattern specification process	1	1	0.001 4123	0.000369	0.051 58117	0.004 35286	0.000 661	0.999 85417	1	1	1
GOTERM_BP_FAT	GO:0010558~negative regulation of macromolecule biosynthetic process	1	1	1	0.013 26021	0.743 9336	1	0.000 369	1	1	1	1
GOTERM_BP_FAT	GO:0003002~regionalization	1	1	0.013 59872	0.000374	0.906 29482	0.084 85772	0.008 04197	1	1	0.995 6671	1
GOTERM_BP_FAT	GO:0000904~cell morphogenesis involved in differentiation	1	1	0.001 50134	0.018 90318	0.178 94506	0.000413	0.020 83409	0.034 85549	1	0.998 02699	1
GOTERM_BP_FAT	GO:0051270~regulation of cell motion	1	1	0.999 7025	0.000491	0.057 44856	0.999 7832	0.814 34136	1	1	1	1
GOTERM_BP_FAT	GO:0030334~regulation of cell migration	1	1	0.909 23674	0.000647	0.251 21682	0.999 99161	0.605 60116	1	1	1	1
GOTERM_BP_FAT	GO:0010629~negative regulation of gene expression	1	1	1	0.023 19303	0.996 662	0.999 66069	0.000 676	1	1	1	1
GOTERM_BP_FAT	GO:0032990~cell part morphogenesis	1	1	0.123 58777	0.000747	0.812 73969	0.008 02495	0.700 00794	0.030 51798	1	0.999 94175	1

GOTERM_BP_FAT	GO:0001944~vasculature development	1	1	0.00575696	0.000806	0.045879	0.108191	0.002679	1	1	1	1
GOTERM_BP_FAT	GO:0048812~neuron projection morphogenesis	1	1	0.019711	0.004134	0.647485	0.000869	0.218429	0.028934	1	0.577283	1
GOTERM_BP_FAT	GO:0001568~blood vessel development	1	1	0.003894	0.001625	0.024019	0.075615	0.004033	1	1	1	1
GOTERM_BP_FAT	GO:0043009~chordate embryonic development	1	1	0.165854	0.002096	0.011914	0.002660	0.004066	1	1	1	1
GOTERM_BP_FAT	GO:0009792~embryonic development ending in birth or egg hatching	1	1	0.180277	0.002612	0.007106	0.002341	0.004362	1	1	1	1
GOTERM_BP_FAT	GO:0045596~negative regulation of cell differentiation	1	1	0.999999	0.006443	0.999999	0.002729	0.002427	1	1	1	1
GOTERM_BP_FAT	GO:0007507~heart development	1	1	0.307429	0.178014	0.687288	0.002955	0.003083	1	1	1	1
GOTERM_BP_FAT	GO:0048667~cell morphogenesis involved in neuron differentiation	1	1	0.042553	0.027708	0.984363	0.004943	0.743372	0.027934	1	0.900730	1
GOTERM_BP_FAT	GO:0016477~cell migration	1	1	0.005065	0.008874	0.063059	0.330943	1	1	1	1	1
GOTERM_BP_FAT	GO:0007409~axonogenesis	1	1	0.121052	0.038567	0.992858	0.005427	0.821669	0.018668	1	0.549568	1
GOTERM_BP_FAT	GO:0035239~tube morphogenesis	1	1	0.050394	0.006242	0.793805	0.968802	0.890339	1	1	1	1
GOTERM_BP_FAT	GO:0008284~positive regulation of cell proliferation	1	1	0.006360	0.107490	1	0.998281	0.201370	1	1	0.999988	1
GOTERM_BP_FAT	GO:0051094~positive regulation of developmental process	1	1	0.797252	0.006743	0.908741	0.511980	0.186690	1	1	1	1
GOTERM_BP_FAT	GO:0035295~tube development	1	1	0.038023	0.009932	0.667558	0.245417	0.074194	1	1	1	1
GOTERM_MF_FAT	GO:0003700~transcription factor activity	1	1	9.3E-09	3.93E-19	0.000373	0.000001	2.61E-10	0.032969	1	0.015968	0.722625
GOTERM_MF_FAT	GO:0043565~sequence-specific DNA binding	1	1	0.000029	1.68E-17	0.000043	0.005789	3.07E-09	0.001009	1	0.174009	0.863169
GOTERM_MF_FAT	GO:0030528~transcription regulator activity	1	1	0.000005	3.58E-11	0.071488	0.002680	6E-09	0.827480	1	0.870625	0.876275
GOTERM_MF_FAT	GO:0005088~Ras guanyl-nucleotide exchange factor activity	1	0.999967	0.000273	1.29E-08	0.000169	0.088399	0.136293	0.998046	1	0.712139	1
GOTERM_MF_FAT	GO:0005085~guanyl-nucleotide exchange factor activity	1	1	0.011550	3.48E-08	0.000186	0.505195	0.713595	0.976261	1	0.298457	1
GOTERM_MF_FAT	GO:0005089~Rho guanyl-nucleotide exchange factor activity	1	0.999961	0.000023	2.17E-07	0.000206	0.604802	0.279479	0.945700	1	0.999947	1
GOTERM_MF_FAT	GO:0003677~DNA binding	1	0.977442	0.020774	5.21E-07	1	0.975948	0.053545	0.999932	1	0.999999	0.999721
GOTERM_MF_FAT	GO:0005083~small GTPase regulator activity	1	1	0.000019	0.000465	8.84E-07	0.114862	0.003772	0.508703	0.795213	0.426871	1
GOTERM_MF_FAT	GO:0030695~GTPase regulator activity	1	1	0.000007	0.000002	0.000001	0.006724	0.003566	0.589750	0.020030	0.053843	1
GOTERM_MF_FAT	GO:0060589~nucleoside-triphosphatase regulator activity	1	1	0.000040	0.000002	0.000003	0.016107	0.011874	0.742828	0.032862	0.100711	1
GOTERM_MF_FAT	GO:0003779~actin binding	0.308731	1	0.000010	0.001285	0.122208	0.015289	0.001890	0.253951	0.992704	0.685824	1
GOTERM_MF_FAT	GO:0008092~cytoskeletal protein binding	0.841721	0.999659	0.000044	0.052360	0.996225	0.021661	0.000048	0.292931	1	0.601161	1
GOTERM_MF_FAT	GO:0003705~RNA polymerase II transcription factor activity, enhancer binding	1	1	0.999461	0.001623	0.999942	1	1	0.999617	1	1	1
GOTERM_CC_FAT	GO:0005886~plasma membrane	1	0.002127	0.000003	3.07E-11	1.48E-10	2.85E-12	3E-13	0.922687	0.000012	0.000011	0.138114
GOTERM_CC_FAT	GO:0005856~cytoskeleton	0.242034	0.954994	0.000014	0.000047	0.000093	1.45E-09	0.000007	0.004607	0.149719	0.000049	0.998613
GOTERM_CC_FAT	GO:0044459~plasma membrane part	0.999991	0.994617	0.037404	0.000030	0.011657	1.95E-07	9.74E-08	0.863945	0.002220	0.097946	0.276763
GOTERM_CC_FAT	GO:0030054~cell junction	1	0.037269	0.040947	0.095421	0.001353	0.001284	0.005846	0.238365	0.000060	0.015631	1
GOTERM_CC_FAT	GO:0044430~cytoskeletal part	0.923290	1	0.025373	0.024843	0.037981	0.000012	0.022709	0.981672	1	0.018568	1
GOTERM_CC_FAT	GO:0042995~cell projection	0.958490	1	0.018720	0.009391	0.002445	0.000727	0.000874	0.999999	1	0.004237	0.999999
GOTERM_CC_FAT	GO:0015629~actin cytoskeleton	0.041718	1	0.001028	0.002117	0.029351	0.013084	0.015486	0.984644	0.999958	0.990498	0.962434
GOTERM_CC_FAT	GO:0070161~anchoring junction	1	0.758404	0.995599	0.003931	0.074776	0.076851	0.024468	1	0.787310	0.999810	1
GOTERM_CC_FAT	GO:0031252~cell leading edge	1	0.775221	0.160792	0.009244	0.138654	0.044656	0.006415	0.004651	0.800450	0.999899	1

Supplemental Table 6. Summary of "omics" data used in this study.

GEO accession	Organism	Tissue/cell type	Age	Data type	Brief description	Reference
GSE67482	Human	brain cortex	fetal (22wk of gestation) and 17y	hairpin bisulfite-Seq	Hairpin bisulfite-Seq data for human brain	This study
GSE47966	Human	brain frontal cortex, neuron, glia	fetal, adolescent and adult	methylC-Seq	Methylomes for several developmental stages and two cell types (neuron and glia) of human and mouse brain	(Lister, Mukamel et al. 2013)
	Mouse					
GSE29184	Mouse	brain cortex	8wk	ChIP-Seq	Chromatin maps for H3K4me1, H3K4me3 and H3K27ac	(Shen, Yue et al. 2012)
GSE16368	Human	brain	75y/81y	ChIP-Seq	Chromatin maps for H3K4me1, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3 and H3K36me3	(Bernstein, Stamatoyannopoulos et al. 2010)
NHGRI GWAS catalog	Human	Various	Various	GWAS	NHGRI GWAS catalog	(Welter, MacArthur et al. 2014)

2.9 REFERENCES

Abecasis, G. R., D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Ballas, N., C. Grunseich, D. D. Lu, J. C. Speh and G. Mandel (2005). "REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis." Cell **121**(4): 645-657.

Bernstein, B. E., J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen and J. A. Thomson (2010). "The NIH Roadmap Epigenomics Mapping Consortium." Nat Biotechnol **28**(10): 1045-1048.

Court, F., C. Tayama, V. Romanelli, A. Martin-Trujillo, I. Iglesias-Platas, K. Okamura, N. Sugahara, C. Simon, H. Moore, J. V. Harness, H. Keirstead, J. V. Sanchez-Mut, E. Kaneki, P. Lapunzina, H. Soejima, N. Wake, M. Esteller, T. Ogata, K. Hata, K. Nakabayashi and D. Monk (2014). "Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment." Genome Res **24**(4): 554-569.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Day, J. J., D. Childs, M. C. Guzman-Karlsson, M. Kibe, J. Moulden, E. Song, A. Tahir and J. D. Sweatt

(2013). "DNA methylation regulates associative reward learning." Nat Neurosci **16**(10): 1445-1452.

Day, J. J. and J. D. Sweatt (2010). "DNA methylation and memory formation." Nat Neurosci **13**(11): 1319-1323.

Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes Dev **25**(10): 1010-1022.

Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biol **4**(5): P3.

Eckhardt, F., J. Lewin, R. Cortese, V. K. Rakyan, J. Attwood, M. Burger, J. Burton, T. V. Cox, R. Davies, T. A. Down, C. Haefliger, R. Horton, K. Howe, D. K. Jackson, J. Kunde, C. Koenig, J. Liddle, D. Niblett, T. Otto, R. Pettett, S. Seemann, C. Thompson, T. West, J. Rogers, A. Olek, K. Berlin and S. Beck (2006). "DNA methylation profiling of human chromosomes 6, 20 and 22." Nat Genet **38**(12): 1378-1385.

Farlik, M., N. C. Sheffield, A. Nuzzo, P. Datlinger, A. Schonegger, J. Klughammer and C. Bock (2015). "Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics." Cell Rep **10**(8): 1386-1397.

Fishell, G. and N. Heintz (2013). "The neuron identity problem: form meets function." Neuron **80**(3): 602-612.

Forrest, A. R., H. Kawaji, M. Rehli, J. K. Baillie, M. J. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jorgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drablos, A. S. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furino, J. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-Sim, R. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, S. Noma, T. Noazaki, S. Ogishima, N. Ohkura, H. Ohimiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. Prendergast, O. J. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. t Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyodo, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verado, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S.

Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci and Y. Hayashizaki (2014). "A promoter-level mammalian expression atlas." Nature **507**(7493): 462-470.

Guo, H., P. Zhu, X. Wu, X. Li, L. Wen and F. Tang (2013). "Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing." Genome Res **23**(12): 2126-2135.

Guo, J. U., D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao and H. Song (2011). "Neuronal activity modifies the DNA methylation landscape in the adult brain." Nat Neurosci **14**(10): 1345-1351.

Hahn, M. A., R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer and Q. Lu (2013). "Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis." Cell Rep **3**(2): 291-300.

Hawrylycz, M. J., E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. D. Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Vawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. Grant and A. R. Jones (2012). "An anatomically comprehensive atlas of the adult human brain transcriptome." Nature **489**(7416): 391-399.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins and T. A. Manolio (2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits." Proc Natl Acad Sci U S A **106**(23): 9362-9367.

Iwamoto, K., M. Bundo, J. Ueda, M. C. Oldham, W. Ukai, E. Hashimoto, T. Saito, D. H. Geschwind and T. Kato (2011). "Neurons show distinctive DNA methylation profile and higher interindividual variations compared with non-neurons." Genome Res **21**(5): 688-696.

Jin, S. G., X. Wu, A. X. Li and G. P. Pfeifer (2011). "Genomic mapping of 5-hydroxymethylcytosine in the human brain." Nucleic Acids Res **39**(12): 5015-5024.

John, R. M. and L. Lefebvre (2011). "Developmental regulation of somatic imprints." Differentiation **81**(5): 270-280.

Johnson, A. D., R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell and P. I. de Bakker (2008). "SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap." Bioinformatics **24**(24): 2938-2939.

Kawaguchi, M., T. Toyama, R. Kaneko, T. Hirayama, Y. Kawamura and T. Yagi (2008). "Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster." J Biol Chem **283**(18): 12064-12075.

Kessler, N. J., T. E. Van Baak, M. S. Baker, E. Laritsky, C. Coarfa and R. A. Waterland (2016). "CpG Methylation Differences Between Neurons and Glia are Highly Conserved from Mouse to Human." Faseb Journal **30**.

Kozlenkov, A., P. Roussos, A. Timashpolsky, M. Barbu, S. Rudchenko, M. Bibikova, B. Klotzle, W. Byne, R. Lyddon, A. F. Di Narzo, Y. L. Hurd, E. V. Koonin and S. Dracheva (2014). "Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites." Nucleic Acids Res **42**(1): 109-127.

Krueger, F. and S. R. Andrews (2011). "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." Bioinformatics **27**(11): 1571-1572.

Kuhn, R. M., D. Haussler and W. J. Kent (2013). "The UCSC genome browser and associated tools." Brief Bioinform **14**(2): 144-161.

Ladd-Acosta, C., J. Pevsner, S. Sabunciyan, R. H. Yolken, M. J. Webster, T. Dinkins, P. A. Callinan, J. B. Fan, J. B. Potash and A. P. Feinberg (2007). "DNA methylation signatures within the human brain." Am J Hum Genet **81**(6): 1304-1315.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens and J. R. Ecker (2013). "Global epigenomic reconfiguration during mammalian brain development." Science **341**(6146): 1237905.

Martinowich, K., D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan and Y. E. Sun (2003). "DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation." Science **302**(5646): 890-893.

Mohn, F. and D. Schubeler (2009). "Genetics and epigenetics: stability and plasticity during cellular differentiation." Trends Genet **25**(3): 129-136.

Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel and D. Schubeler (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-766.

Montano, C. M., R. A. Irizarry, W. E. Kaufmann, K. Talbot, R. E. Gur, A. P. Feinberg and M. A. Taub (2013). "Measuring cell-type specific differential methylation in human brain tissue." Genome Biol **14**(8): R94.

Nord, A. S., M. J. Blow, C. Attanasio, J. A. Akiyama, A. Holt, R. Hosseini, S. Phouanenavong, I. Plajzer-Frick, M. Shoukry, V. Afzal, J. L. Rubenstein, E. M. Rubin, L. A. Pennacchio and A. Visel (2013). "Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development." Cell **155**(7): 1521-1531.

Ochiishi, T., T. Yamauchi and T. Terashima (1998). "Regional differences between the immunohistochemical distribution of Ca2+/calmodulin-dependent protein kinase II alpha and beta isoforms in the brainstem of the rat." Brain Res **790**(1-2): 129-140.

Sanchez-Mut, J. V., E. Aso, N. Panayotis, I. Lott, M. Dierssen, A. Rabano, R. G. Urdinguio, A. F. Fernandez, A. Astudillo, J. I. Martin-Subero, B. Balint, M. F. Fraga, A. Gomez, C. Gurnot, J. C. Roux, J. Avila, T. K. Hensch, I. Ferrer and M. Esteller (2013). "DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease." Brain **136**(Pt 10): 3018-3027.

Shen, Y., F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov and B. Ren (2012). "A map of the cis-regulatory sequences in the mouse genome." Nature **488**(7409): 116-120.

Smallwood, S. A., H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik and G. Kelsey (2014). "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity." Nat Methods **11**(8): 817-820.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." Nature **480**(7378): 490-495.

Steyaert, S., W. Van Criekinge, A. De Paepe, S. Denil, K. Mensaert, K. Vandepitte, W. Vanden Berghe, G. Trooskens and T. De Meyer (2014). "SNP-guided identification of monoallelic DNA-methylation events from enrichment-based sequencing data." Nucleic Acids Res **42**(20): e157.

Szulwach, K. E., X. Li, Y. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He and P. Jin (2011). "5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging." Nat Neurosci **14**(12): 1607-1616.

Weaver, I. C., N. Cervoni, F. A. Champagne, A. C. D'Alessio, S. Sharma, J. R. Seckl, S. Dymov, M. Szyf

and M. J. Meaney (2004). "Epigenetic programming by maternal behavior." Nat Neurosci **7**(8): 847-854.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson (2014). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations." Nucleic Acids Res **42**(Database issue): D1001-1006.

Wu, X., M. A. Sun, H. Zhu and H. Xie (2015). "Nonparametric Bayesian clustering to detect bipolar methylated genomic loci." BMC Bioinformatics **16**(1): 11.

Xie, W., C. L. Barr, A. Kim, F. Yue, A. Y. Lee, J. Eubanks, E. L. Dempster and B. Ren (2012). "Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome." Cell **148**(4): 816-831.

Zeng, H., E. H. Shen, J. G. Hohmann, S. W. Oh, A. Bernard, J. J. Royall, K. J. Glattfelder, S. M. Sunkin, J. A. Morris, A. L. Guillozet-Bongaarts, K. A. Smith, A. J. Ebbert, B. Swanson, L. Kuan, D. T. Page, C. C. Overly, E. S. Lein, M. J. Hawrylycz, P. R. Hof, T. M. Hyde, J. E. Kleinman and A. R. Jones (2012). "Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures." Cell **149**(2): 483-496.

Zhao, L., M. A. Sun, Z. Li, X. Bai, M. Yu, M. Wang, L. Liang, X. Shao, S. Arnovitz, Q. Wang, C. He, X. Lu, J. Chen and H. Xie (2014). "The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation." Genome Res **24**(8): 1296-1307.

Zhu, J., M. Adli, J. Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P. L. De Jager, D. A. Bennett, J. A. Houmard, D. M. Muoio, T. T. Onder, R. Camahort, C. A. Cowan, A. Meissner, C. B. Epstein, N. Shoresh and B. E. Bernstein (2013). "Genome-wide chromatin state transitions associated with developmental and environmental cues." Cell **152**(3): 642-654.

Chapter 3 - EGR1 Recruits TET1 to Shape the Brain Methylome during Development and upon Neuronal Activity

3.1 ABSTRACT

Life experience can leave lasting marks on brain cells, such as epigenetic changes in neurons induced by neuronal activity. How life experience or neuronal activity is translated into storable epigenetic information remains largely unknown. With unbiased data-driven approaches, we predicted that *Egr1*, a transcription factor important for memory formation, plays an essential role in brain epigenetic programming. We performed EGR1 ChIP-seq and validated that thousands of EGR1 binding sites are with cell-type specific methylation patterns established during postnatal frontal cortex development. More specifically, the CpG dinucleotides within these EGR1 binding sites become hypomethylated in mature neurons but remain heavily methylated in glia. We further demonstrated that EGR1 recruits a DNA demethylase TET1 to remove the methylation marks at EGR1 binding sites and activate downstream genes. In addition, we found that the frontal cortices from the knockout mice lacking *Egr1* or *Tet1* share strikingly similar profiles in both gene expression and DNA methylation. In summary, our study reveals EGR1 programs brain methylome together with TET1 during postnatal development and provides a new insight into how life experience and neuronal activity may shape the brain methylome.

3.2 INTRODUCTION

It has been well acknowledged that early postnatal experience is critical for brain development and may induce long-lasting epigenetic changes in postmitotic neurons (Weaver, Cervoni et al. 2004, Murgatroyd, Patchev et al. 2009). Growing evidence indicates that learning and memory are highly dependent on the functions of epigenetic machinery such as DNA methyltransferases (DNMTs) (Feng, Zhou et al. 2010, LaPlant, Vialou et al. 2010, Stroud, Su et al. 2017) and DNA demethylases (Kaas, Zhong et al. 2013, Rudenko, Dawlaty et al. 2013, Yu, Su et al. 2015, Zhu, Girardo et al. 2016), the Ten-Eleven Translocation (Tet) proteins including TET1, TET2, and TET3. The double knockout of DNMT1 and DNMT3a leads to abnormal expression of genes contributing to synaptic plasticity and the deficits in learning and memory (Feng, Zhou et al.

2010). Genetic deletion or knock-down of each TET enzyme results in a unique set of phenotypes (Dawlaty, Ganz et al. 2011, Koh, Yabuuchi et al. 2011). *Tet1* is involved in neural progenitor cell proliferation (Zhang, Cui et al. 2013) and neuronal activity-induced active DNA demethylation in the dentate gyrus of the adult mouse brain (Guo, Su et al. 2011). *Tet1* knockout mice exhibited impaired hippocampal neurogenesis, significant deficiency in short-term memory retention (Zhang, Cui et al. 2013), abnormal long-term depression and impaired memory extinction (Rudenko, Dawlaty et al. 2013). The deletion of *Tet3* leads to neonatal lethality (Gu, Guo et al. 2011) and neural progenitor cells induced from *Tet3* knockout ES cells undergo apoptosis rapidly with reduced terminal differentiation of neurons (Li, Yang et al. 2014). Significant impairment in fear extinction memory was observed in mice with *Tet3* knockdown *via* shRNA (Mathelier, Zhao et al. 2014). Although little is known about the role of *Tet2* in neuronal differentiation or function, *Tet2* knockout mice show abnormal hyper-methylation in frontal cortex (Lister, Mukamel et al. 2013). Despite the known needs of DNMTs and TETs for learning and memory, how these enzymes are directed to specific genomic loci in neurons remains elusive.

Neuronal activity-induced DNA methylation changes may occur within hours after electroconvulsive stimulation (Guo, Ma et al. 2011). This suggests that neurons can react to environmental stimuli and guide the epigenetic machinery to desired genomic loci swiftly. As an immediate early gene, *Egr1* (*Egr1* in mice, *EGR1* in humans, also known as Zif268, Krox-24, and NGFI-A) can be rapidly and transiently induced by neuronal activity (Li, Carter et al. 2005, Veyrac, Besnard et al. 2014). *Egr1* is a critical transcriptional regulator involved in brain development, learning, and long-term neuronal plasticity (Wei, Xu et al. 2000, Mataga, Fujishima et al. 2001, Renaudineau, Poucet et al. 2009, Veyrac, Gros et al. 2013). With a rapid increase in expression during the first few weeks after birth, *Egr1* controls the selection, maturation and functional integration of newborn neurons (Veyrac, Gros et al. 2013). A seminal study has established a link between maternal care and methylation programming during early postnatal brain development, and *Egr1* was proposed to be an epigenetic regulator of the glucocorticoid receptor (Weaver, Cervoni et al. 2004). More interestingly, EGR1 has a binding motif containing CpG dinucleotides (5'- GCGTGGGCG-3') (Mora-Lopez, Pedreno-Horrillo et al. 2008) and the binding of EGR1 to target DNA is insensitive to methylation (Hashimoto,

Olanrewaju et al. 2014, Zandarashvili, White et al. 2015). However, whether EGR1 can direct epigenetic machinery to its target sites upon neuronal activation is unknown.

Recently, we have implemented a nonparametric Bayesian clustering approach (Wu, Sun et al. 2015) to identify genomic loci with bipolar DNA methylation patterns: the presence of both hypo-methylated and hyper-methylated patterns within a mixed cell population. With this approach, we observed the number of bipolar methylated loci increased dramatically during early stages of brain development and brain bipolar methylated loci were enriched for GWAS variants associated with neurological disorder-related diseases/traits (Sun, Sun et al. 2016). Interestingly, genes associated with brain bipolar methylated loci are involved in neuronal differentiation, cell migration, and cell morphogenesis. In this study, we explored the epigenetic regulatory mechanism underlying the birth of bipolar methylated loci and identified EGR1 as a key mediator involving in brain epigenome programming during postnatal development. Our study provides the first compelling data demonstrating EGR1 recruits TET1 to demethylate EGR1 binding sites. Our results implicate the interaction between transcription factors and epigenetic machinery as a general mechanism to achieve locus-specific epigenetic regulation upon neuronal activation.

3.3 METHODS

3.3.1 Accession codes.

ChIP-seq, RNA-seq and methylome data generated in this study have been deposited in NCBI GEO under accession number GSE108768. Publicly available brain "omics" data used in this manuscript were summarized in **Supplementary Table 5**.

3.3.2 Animal.

All animal experiments were performed according to guidelines of the Institutional Animal Care and Use Committee at Virginia Tech (Blacksburg, VA, USA). The *Egr1* heterogeneous mouse strain (B6N; 129-*Egr1*$^{tm1Jmi/J}$), the *Tet1* heterogeneous mouse strain (B6; 129S4-*Tet1*$^{tm1.1Jae/J}$) were purchased from The Jackson Laboratory. Genomic DNA was isolated from tail biopsies and genotyped by PCR according to The Jackson Laboratory's protocols.

3.3.3 Mouse neuronal stem cell isolation, cell culture and nucleofection.

The subventricular zones (SVZ) of the adult mouse forebrain were dissected and dissociated as reported previously (Theus, Ricard et al. 2012). Briefly, the dissected tissue was subjected to 15 min at 37 ºC in 1× Hank's balanced salt solution (HBSS) containing 1.33 mg/mL trypsin (Fisher), 0.7 mg/mL hyaluronidase (Sigma), and 0.2 mg/mL kynurenic acid (Sigma). After trituration 10 times with 1 ml pipette, cells were subjected to another 15 min incubation at 37 ºC, and then passed through a 70µm strainer (Falcon), followed by centrifugation at 230g for 5 min. The cell pellet was resuspended in 10ml neuronal stem cell medium supplemented with EGF, FGF (PeproTech) and Amphotericin B (Sigma), and then transferred to flasks. After 11-day culture with the medium changed every two days, neuronal stem cells grew to 90% confluency and ready for sub-culturing. The nucleofection of NSCs was carried out with 3µg shRNA plasmids by using the Amaxa Basic Neuron SCN Nucleofector kit (Lonza) according to the manufacturer's instructions. 48hr after nucleofection, cells were treated with or without 25mM KCl stimulation for 4 h, and then harvested for methylation-sensitive qPCR (MS-qPCR).

HEK293T (ATCC, CRL-11268™) cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Life Technologies) supplemented with 10% heat-inactivated fetal bovine serum (Corning) and 1% penicillin/streptomycin (Gibco). All cells were cultured in a humidified atmosphere of 5% CO_2 incubator at 37 ºC.

3.3.4 Antibodies.

Antibody against Tet1 (Millipore, 09-872), Tet2 (Ambion, ab135087), Tet3 (Abiocode, M1092-4a), HA (Invitrogen, 26183), Flag (Sigma, F1804) were purchased commercially. Egr1 antibody (sc-189), Egr1 antibody (sc-101033), rabbit normal IgG (sc-2027) and mouse normal IgG (sc-2025) were purchased from Santa Cruz Biotechnology. For western blot analysis, goat anti-rabbit horseradish peroxidase-conjugated secondary antibody (Invitrogen, 65-6120) was used at a 5000:1 dilution, goat anti-mouse horseradish peroxidase-conjugated secondary antibody (Sigma, A8924) was used at a 10000:1 dilution.

3.3.5 Plasmid construction.

Flag-tagged mouse EGR1 expression vector was obtained from Addgene (plasmid 11729). Plasmids encoding Flag-tagged EGR1 N-terminal (amino acids 1-318) and Flag-tagged zinc

fingers region (amino acids 318-533) were generated by subcloning of the DNA fragments into EcoRI and XhoI sites of pcDNA3 vector. In order to clone truncated versions of TET1 protein, cDNA was synthesized from mRNA of C57BL/6J mice. Plasmids encoding HA-tagged TET1 FL (amino acids 1-2039), TET1-Mid domain (amino acids 701-1366), TET1 CD domain (amino acids 1367-2039) were generated by subcloning of the DNA fragments into BamHI and XbaI of pcDNA3 vector. Plasmids encoding TET1-CXXC domain (amino acids 1-700) were generated by subcloning of the DNA fragment into AgeI and NotI of pCAG-GFPmG1 vector (plasmid 50444 from Addgene). Egr1 shRNA (sc-35267-SH) and scrambled shRNA (sc-108060) were purchased from Santa Cruz Biotechnology, Tet1 shRNA (shTet1-a+b) (Huang, Jiang et al. 2013) was a kind gift from Dr. Jianjun Chen (City of Hope).

For luciferase constructs, three selected EGR1 binding loci: *Arc*, *Galnt9* and *Npas4* loci were amplified from genomic DNA of C57BL/6 mice using the primers listed in **Supplementary Table 6**. After enzymatic digestion for at least 4 h, PCR-amplified products were cloned into pCpGfree-promoter-Lucia Vector (Invivogen). All inserts were verified by Sanger sequencing.

3.3.6 Luciferase reporter assays.

Luciferase reporter constructs were either mock-treated or methylated *in vitro* with M.SssI methylase (NEB) for at least 4h at 37°C and purified with PureLink PCR Purification Kit (Qiagen). NSCs were seeded at $5x10^4$/well in 24-well plates overnight, then transfected with 0.2µg of reporter constructs and 0.02µg of firefly luciferase control vector pGL 4.13 (Promega) using lipofectamine 3000 (Invitrogen). For each sample, triplicate transfections were carried out. 48h after transfection, cell lysates, and the medium were assayed for luciferase activity by Dual-Luciferase Reporter Assay (Promega). Lucia luciferase activity of individual transfections was normalized to firefly luciferase activity and analyzed relatively to empty pCpG-free promoter vector (Kennedy, Schmidt et al. 2014).

3.3.7 qRT-PCR analysis.

Total RNA was extracted using the RNeasy kit (Qiagen) and cDNA was generated using high-capacity cDNA reverse transcription kit (Applied Biosystems). qRT-PCR experiments were performed using GoTaq® qPCR Master Mix (Promega) on StepOnePlus™ Real-Time PCR Systems. Relative expression levels were determined by a comparative ΔΔCt method with beta-actin as endogenous control.

3.3.8 Quantitative analysis of DNA methylation by MS-qPCR.

NSC or NSC stimulated with 25mM KCl for 4h were collected and lysed in 500µl lysis buffer (10mM Tris, 25mM EDTA, 100mM NaCl, 0.5% SDS, pH 8) supplemented with 10µl proteinase K (Ambion) at 55°C overnight. After heat-inactivation of proteinase K at 80°C for 20 min, the cell lysate was treated with 5ul RNase cocktail (Life technologies) at 37ºC for 1h, followed by phenol-chloroform and ethanol precipitation. 1µg genomic DNA was incubated with 20 units HapII (NEB) or in a mock reaction without HpaII at 37ºC 2h, followed by heat-inactivation at 80ºC for 20 min. Equal amounts from both the HpaII reaction and mock reaction were used in qPCR with primers (**Supplementary Table 6**) flanking the HpaII digestion sites -CCGG-. The methylation level was calculated by $2^{Ct(mock)-Ct(HpaII)}$ x 100%, and typically averaged across 3 technical replicates (Guo, Su et al. 2011).

3.3.9 ChIP-seq, ChIP-qPCR and sequential ChIP-qPCR.

Frontal cortices of adult mice (6-week-old) were dissected on ice, cross-linked with 1% formaldehyde, and then neutralized by 0.125M glycine. Samples were lysed in lysis buffer I (50mM HEPES-KOH, 1mM EDTA, pH 8.0, 140mM NaCl, 0.25% Triton X-100, 0.5% NP-40 and 10% glycerol and halt protease inhibitor cocktail), lysis buffer II (10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 200mM NaCl and 0.5mM EGTA, pH 8.0 and halt protease inhibitor cocktail), respectively. Samples were then sonicated with Covaris M2 (Covaris) into 200-700bp. 10% of pre-cleared chromatin was stored as input material. The rest was incubated overnight at 4°C with 30µl of Dynal protein A/G magnetic beads (Life technologies) that had been pre-incubated with specific antibodies.

For ChIP-seq, frontal cortices from 4-5 male mice at 6-week-old were used for a single ChIP experiment. The genomic DNA fragments were obtained from anti-EGR1 immunoprecipitated chromatin as mentioned above, and the library construction was performed as described previously (Blecher-Gonen, Barnett-Itzhaki et al. 2013). DNA in the range 270-600bp was recovered by Pippin Prep (Sage Science), after size distribution assessment by Agilent Bioanalyzer and quantification by qPCR (Kapa Library quantification kit), libraries were subjected to 100-bp paired-end read sequencing on the Illumina HiSeq 2000 platform.

For ChIP-qPCR, frontal cortices from 2 male mice at 6-week-old were used for a single ChIP experiment. Real-time PCR was performed using GoTaq® qPCR Master Mix (Promega) on

StepOnePlus™ Real-Time PCR Systems. Antibodies used for immunoprecipitation were anti-H3K27ac (ab4729, Abcam), anti-EGR1 (sc-189, Santa Cruz Biotechnology), anti-TET1 (09-872, Millipore) and normal rabbit IgG (sc-2027, Santa Cruz Biotechnology). Primers specific for chosen genomic regions were summarized in **Supplementary Table 6**.

For sequential ChIP-qPCR, frontal cortices from 2 wild-type male mice at 6-week-old were used for a single sequential ChIP experiment. Chromatin was prepared as described above, after chromatin immunoprecipitation with the first antibody, magnetic beads were resuspended in 1% SDS supplemented with 10mM DTT at 37°C for 30 min. The elute was then diluted 20 times with dilution buffer (1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH8, 150mM NaCl) and the second antibody was added for immunoprecipitation overnight. Beads were then washed, crosslinking reversed, and DNA purified. The antibodies used were anti-EGR1 (sc-189, Santa Cruz Biotechnology) anti-TET1 (09-872, Millipore) and normal rabbit IgG (sc-2027, Santa Cruz Biotechnology).

3.3.10 ChIP-seq data analysis.

All reads were first trimmed according to their sequencing qualities, then the trimmed reads were mapped to the mouse reference (mm10) by using Bowtie(Langmead and Salzberg 2012) with parameters "–n 2 –l 50". Peak calling was performed using SPP (Kharchenko, Tolstorukov et al. 2008) with parameters "-npeak=300000 -p=5 -savr -savp –rf". The data reproducibility between biological replicates was examined following irreproducible discovery rate (IDR) framework with parameters "0 F signal.value" (Landt, Marinov et al. 2012). The IDR threshold was set as 2%, which is recommended by ENCODE. The reported peaks were produced by merging highly reproducible peaks of biological replicates. The top 200 most significant peaks were selected for *de novo* motif discovery using MEME Suite (Bailey, Boden et al. 2009). Known motif enrichment analysis was performed using the script findMotifs.pl in HOMER with parameter "–mset vertebrates".

3.3.11 Methylome analysis to identify bipolar methylated loci.

The bipolar DNA methylation inference was performed by pooling all human and mouse brain methylome datasets together, respectively, and then bipolar DNA methylation loci with at least 100Xs read coverage were identified following the procedure described previously (Sun, Sun et

al. 2016). After merging of the overlapped loci, a total of 39,114 and 21,946 bipolar DNA methylation loci were identified for human and mouse brain methylomes, respectively.

3.3.12 Reduced Representation Bisulfite Sequencing (RRBS) library construction and data analysis.

Genomic DNA from mouse frontal cortex was extracted using AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Five microgram mouse genomic DNA was spiked with 0.02% unmethylated cl857 Sam7 Lambda DNA (Promega) and sonicated to 200bp fragments with Covaris M2 (Covaris). After purification (PureLink PCR Purification Kit, Invitrogen), DNA fragments were then subjected to end repair with the end repair enzyme mix (NEB), dA tailing using Klenow 3'-5' exo- (NEB) with purification at each step. Ligation with cytosine-methylated Illumina TruSeq DNA adapters were performed at 16°C using T4 DNA ligase (NEB) overnight. The adapter-ligated DNA was then digested with MseI and MluCI (NEB) at 37°C for 1h. After purification, DNA fragments were subject to bisulfite conversion using the EpiTect Bisulfite Kit (Qiagen). After bisulfite conversion, the single-stranded uracil-containing DNA was subjected to 12 cycles of PCR reaction with Illumina TruSeq PCR primers and 2.5 U Pfu TurboCx Hotstart DNA polymerase (Agilent) to recover enough DNA for sequencing on Hiseq 4000 platform with 75bp paired-end mode (Illumina).

Sequencing bases with low quality in reads were trimmed by a customized Perl script. Adapters in reads were removed by Cutadapt. After trimming, sequencing reads were mapped to mm10 using Bismark with Bowtie2 and duplication reads were removed by a customized Perl script. Fisher Exact test was used to evaluate the significance of differential methylation on CpG site (Lister, Mukamel et al. 2013). In order to control FDR, a sequential permutation method is employed (Bancroft, Du et al. 2013). A total of 1,000 permutations were performed for each CpG site. The number of true null hypotheses (m0) was estimated by a histogram method (Lister, Mukamel et al. 2013). Based on the estimated m0, the adjusted p-value for each CpG site was calculated. Differentially methylated sites (DMSs) were identified with adjusted p-value lower than or equal to 0.05. To determine differentially methylated regions (DMRs), we developed a two-step approach. First, any two adjacent DMSs with at most 500bp distance were merged into a cluster. In each of clusters which include at least 5 CpG sites, at least 80% of DMSs are prone to be methylated or unmethylated in one of conditions. All clusters filled requirements above

will be considered as DMR candidates. Second, at least 80% of CpG sites in a candidate DMR are prone to be methylated or unmethylated in one of conditions and each CpG site was required to have at least 0.1 methylation differences.

3.3.13 RNA-seq library construction and data analysis.

Total RNA from mouse frontal cortex was extracted using All Prep DNA/RNA/miRNA Universal Kit (Qiagen). mRNA-seq libraries were constructed using the TruSeq Stranded mRNA Library Preparation Kit (Illumina) following the manufacturer's instructions. Briefly, the polyA-containing mRNA molecules were enriched from 500ng total RNA *via* two rolls of oligo-dT magnetic beads purification. The resultant mRNA was fragmented and primed into first strand cDNA using reverse transcriptase and random primers, followed by the removal of the RNA template and synthesis of the second strand to generate blunt-ended ds cDNA. Then a single 'A' nucleotide was added to the 3' ends of the blunt fragments and indexing adapter was ligated to the ends of the ds cDNA. Those DNA fragments with adapter molecules on both ends were enriched by PCR amplification for 12 cycles. After Ampure XP beads purification, the PCR product was size-selected with the range from 220bp to 500bp on 2% dye-free agarose gel using pippin recovery system (Sage Science). The recovered libraries were sequenced on Hiseq 4000 platform with 75bp paired-end mode (Illumina). After trimming bases of low quality and removing adapters, reads were mapped to mm10 by RSEM (Li and Dewey 2011) with Bowtie2. The raw counts were employed to identify differentially expression genes by DESeq2 (Love, Huber et al. 2014). The definition of differentially expression genes includes two requirements: (1) the p-value adjusted is less than 0.05, and (2) there are at least 1.5 fold changes. The visualized data normalized to 1 million was generated by Bedtools (Quinlan 2014).

3.3.14 GO analysis.

Gene Ontology (GO) analysis was performed via the Gene Functional Annotation Tool at the DAVID (Huang, Sherman et al. 2007) website (https://david.ncifcrf.gov/, version 6.8). Default parameters were used for the enrichment analysis for biological process (BP), cellular component (CC), and molecular function (MF). The resulting GO terms and the corresponding p-values were then processed by using REVIGO (Supek, Bosnjak et al. 2011) to remove redundancy. The ten most significant BP categories were shown.

3.3.15 Co-immunoprecipitation and western blotting.

HEK 293T cells were transfected with plasmids by Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. Cell lysis, immunoprecipitation, and western blot analysis were performed as previously described (de la Rica, Rodriguez-Ubreva et al. 2013). Briefly, cells were lysed in ice-cold lysis buffer (50mM Tris-HCl (pH 7.4), 150mM NaCl, 1% Triton X-100, 2mM EDTA), supplemented with protease inhibitor cocktail (Thermo scientific). Immunoprecipitation was carried out by incubating specific antibody coupled Dynabeads Protein A or Protein G (Life Technologies) at 4°C overnight. The samples were washed four times with ice-cold lysis buffer, and then suspended in 30µl loading buffer (Life Technologies). After boiling at 95°C for 5 min, the samples were analyzed by western blot with specific antibodies.

3.4 RESULTS

3.4.1 Thousands of EGR1 binding sites gradually lose DNA methylation in neurons during brain development.

To explore epigenetic regulatory mechanisms during brain development, we followed our previous approach (Sun, Sun et al. 2016) (see Methods for details) to re-analyze methylomes for frontal cortices at different development stages and identified a total of 11,178 (human) and 4,692 (mouse) bipolar methylated loci within 10 kb from TSSs. For these bipolar methylated loci, we determined the methylation correlations between all possible pairs (**Supplementary Fig. 1a and 1d**) and identified five major co-methylated modules showing distinct methylation profiles during brain development and neural cell specification (**Supplementary Fig. 1b and 1e**). For instance, in mouse frontal cortices, the bipolar methylated loci in module I and II were hypomethylated in neurons. In contrast, the bipolar methylated loci in module III and IV were found to be hypomethylated in non-neuronal cells, while the bipolar methylated loci in module V tended to show age-related methylation. Using HOMER (Heinz, Benner et al. 2010) , we determined the motifs for TFs enriched in each co-methylated module (**Supplementary Fig. 1c, 1f and Supplementary Table 1**) and identified *Egr1* is associated with module I, the largest module for both human and mouse. More interestingly, the CpG dinucleotides within the EGR1 binding motifs are gradually demethylated during postnatal brain development and the methylation losses are limited in neurons (**Supplementary Fig. 2***).*

To validate such a computational prediction, we performed ChIP-seq for EGR1 in duplicate with mouse frontal cortices at 6 weeks and identified 12,014 high-confidence peaks (**Supplementary Fig. 3**). Independent ChIP-qPCR assays were performed for six loci to confirm the significant enrichment of EGR1 binding in peaks identified (**Supplementary Fig. 4a**). From the sequences of EGR1 peaks, we determined the most significantly enriched motif as "GCGGGGGCGG" (**Fig. 1a, E-value=1.1e^{-252}**), which is similar to the canonical EGR1 response element reported previously (Christy and Nathans 1989). A total of 81.8% of EGR1 peaks localize in gene promoters or within genic regions (**Fig. 1b**), and the frequency of EGR1 binding sites increases when approaching transcription start sites (**Supplementary Fig. 4b**). We further integrated EGR1 peaks with ChIP-seq datasets for histone modifications (Shen, Yue et al. 2012) and observed that active enhancer mark H3K27ac and active promoter mark H3K4me3 are strongly enriched in the vicinity of EGR1 peaks (**Fig. 1c**). More specifically, 46.9% and 44.4% of EGR1 binding sites overlap with H3K27ac peaks and H3K4me3 peaks, respectively. Additional ChIP-qPCR assays validated that the H3K27ac mark is enriched at the six genomic loci with EGR1 binding (**Supplementary Fig. 4c**). These results suggest that EGR1 mainly binds onto active promoters or enhancers to activate the expression of downstream genes in mouse frontal cortex.

To examine methylation dynamics of EGR1 peaks, we made use of embryonic forebrain methylomes recently released by ENCODE to obtain the methylation profiles for 12,014 EGR1 peaks during brain development. 51.0% of EGR1 binding sites show constant hypomethylation (methylation level <= 0.2) throughout all developmental stages with methylomes available from E11.5 to 22 m (**Supplementary Fig. 5**). On the other hand, 34.3% of EGR1 binding sites (n = 4,125) exhibit methylation dynamics during development with the maximum methylation difference between stages greater than or equal to 0.2. The majority of these EGR1 binding sites show decreased methylation changes during development and become hypomethylated in neurons (**Fig. 1d**). We next focused on the comparisons of DNA methylation between d0 and 6-week-old frontal cortices, and between NeuN+ and NeuN- cells at 7 weeks. Out of the 4,125 EGR1 binding sites with methylation variation, 2,106 (51.1%) showed methylation decrease by at least 20% during postnatal development (d0 to 6 week) and 3,451 (83.7%) showed methylation decrease by at least 20% in neurons compared with those in glial cells. In contrast, only 111 (2.7%) and 37 (0.8%) of EGR1 binding sites showed methylation increase by at least 20% during development or in neurons, respectively. We next asked if the methylation changes

on EGR1 binding sites during development and between the two cell types were correlated. We found that 1,925 out of 2,106 (91.4%) EGR1 binding sites hypomethylated in adult frontal cortex were also hypomethylated in neurons. In addition, the methylation changes in d0 vs. 6 week, and glia vs. neuron were significantly positively correlated (**Fig. 1e**, Pearson's R=0.35, p-value=$1.31e^{-122}$).

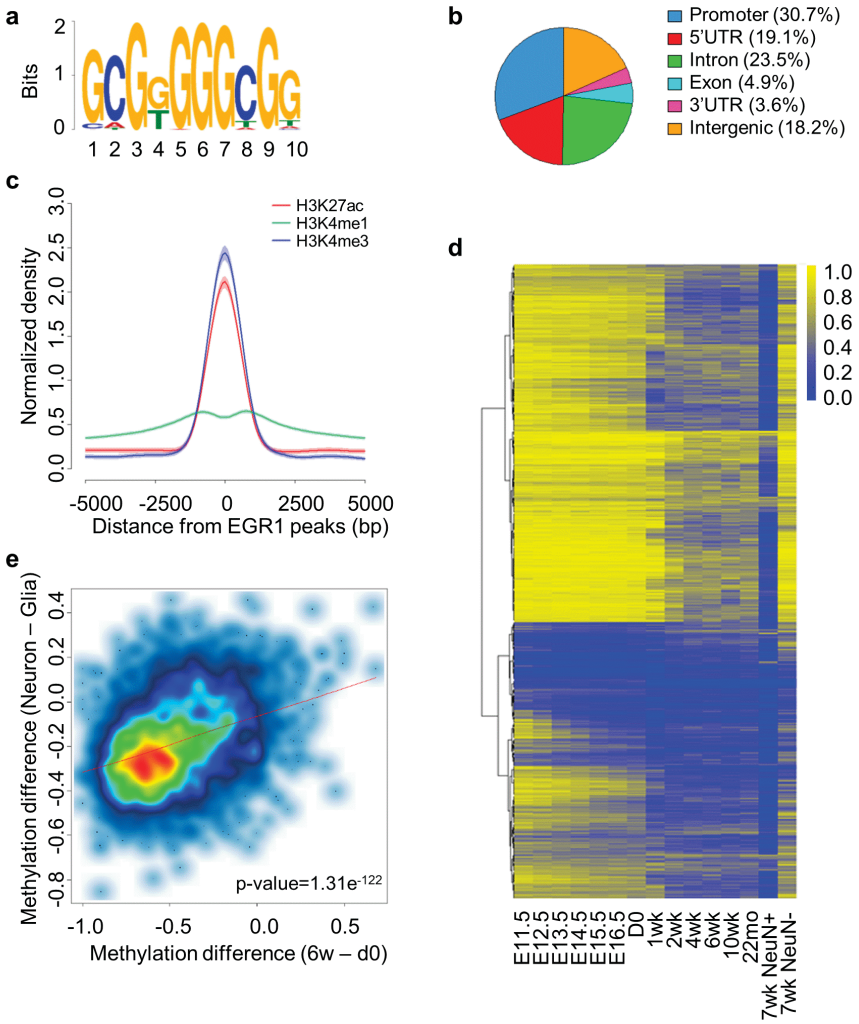


Figure 1. Methylation dynamics of EGR1 binding sites during mouse brain development.

(a) EGR1 binding motif (E value = 1.1e-252) identified from ChIP-seq data generated with mouse adult frontal cortices. (b) Genomic distribution of EGR1 peaks. (c) Distribution of histone marks H3K27ac, H3K4me1 and H3K4me3 surrounding EGR1 peaks. (d) Methylation dynamics of EGR1 binding sites during brain development from embryonic day 11.5 to 22 months and in neurons (NeuN+) and non-neuronal cells (NeuN-). (e) Correlation of methylation changes at EGR1 binding sites during mouse brain development and between cell specification. Only binding sites with at least ten methylation calls in all four methylomes were included. P-values were determined with Wilcoxon Rank Sum Test.

To determine if the 1,925 EGR1 peaks are linked to specific functions, we performed Gene Ontology (GO) enrichment analysis and found that genes with TSS flanking 10 kb of these EGR1 peaks are significantly enriched in the regulation of ion membrane transport, which is important for membrane potential formation and action potential propagation in neurons (**Supplementary Fig. 6 and Supplementary Table 2**).

3.4.2 EGR1 interacts and recruits TET1 to its target binding sites

A recent report indicated that a member of EGR family, WT1 (Wilms tumor suppressor gene 1) may recruit TET2 to demethylate its binding sites in leukemia cells (Wang, Xiao et al. 2015). More specifically, the zinc-finger domain (residues 323-449) of WT1 binds directly to the CD domain (C-terminal region) of TET2 enzyme. WT1 and EGR1 share a similar structure and bind to a same consensus DNA sequence (Nakagama, Heinrich et al. 1995, Ritchie, Yue et al. 2010). Interestingly, the three TET family members also share significant homology (Piccolo, Bagci et al. 2013). These findings raise the possibility that EGR1 may interact with TET enzymes to program brain methylome. To test this hypothesis, we reanalyzed RNA-seq data (Lister, Mukamel et al. 2013) for mouse frontal cortices to examine the expression profiles of *Egr1* and *Tet* gene family (*Tet1-3*) during mouse brain development. *Egr1* transcript in mouse frontal cortices rapidly increased during fetal to 2-week stage, maintaining at a higher level throughout later developmental stages (**Supplementary Fig. 7a**). The levels of *Tet2* and *Tet3* expression gradually decrease during development while *Tet1* shows an increased expression level during the second postnatal week (**Supplementary Fig. 7b**). We further examined the methylation profiles of EGR1 binding sites in *Tet2* knockout mice (Lister, Mukamel et al. 2013). It has been reported that 19.7% of regions hypo-methylated in adult frontal cortex *vs* fetal are with increased methylation in adult Tet2 knockout mice (Lister, Mukamel et al. 2013). Interestingly, we found that EGR1 binding sites show no significant methylation difference (Wilcoxon rank-sum test) between *Tet2-/-* and wild-type mice (**Supplementary Fig. 8**). This indicates that *Tet2* is not required for the demethylation of EGR1 binding sites.

To further determine which of the three TET enzymes may participate in the demethylation of EGR1 binding sites, we performed co-immunoprecipitation assays (Co-IP) using mouse frontal cortices and found that TET1, but not TET2 or TET3, presents in the EGR1 complex (**Fig. 2a**). To determine the binding regions responsible for EGR1-TET1 interaction, we constructed a

series of deletion mutants for both EGR1 and TET1. We first conducted Co-IP in HEK293T cells with Flag-tagged EGR1 co-transfected with HA-tagged TET1-CXXC, TET1-Mid or TET1-CD, respectively. Using anti-EGR1 antibody to probe anti-HA immunoprecipitates, we found that EGR1 binds to TET1-CD (residues 1367-2039), but not TET1-CXXC (residues 1-700) or TET1-Mid (residues 701-1366) (**Fig. 2b**). On the other hand, to map the EGR1 domain mediating EGR1-TET1 interaction, we co-transfected HEK293T cells with HA-tagged TET1-CD along with Flag-tagged EGR1-N, EGR1-C or EGR1-FL. The immunoblotting analyses of anti-Flag

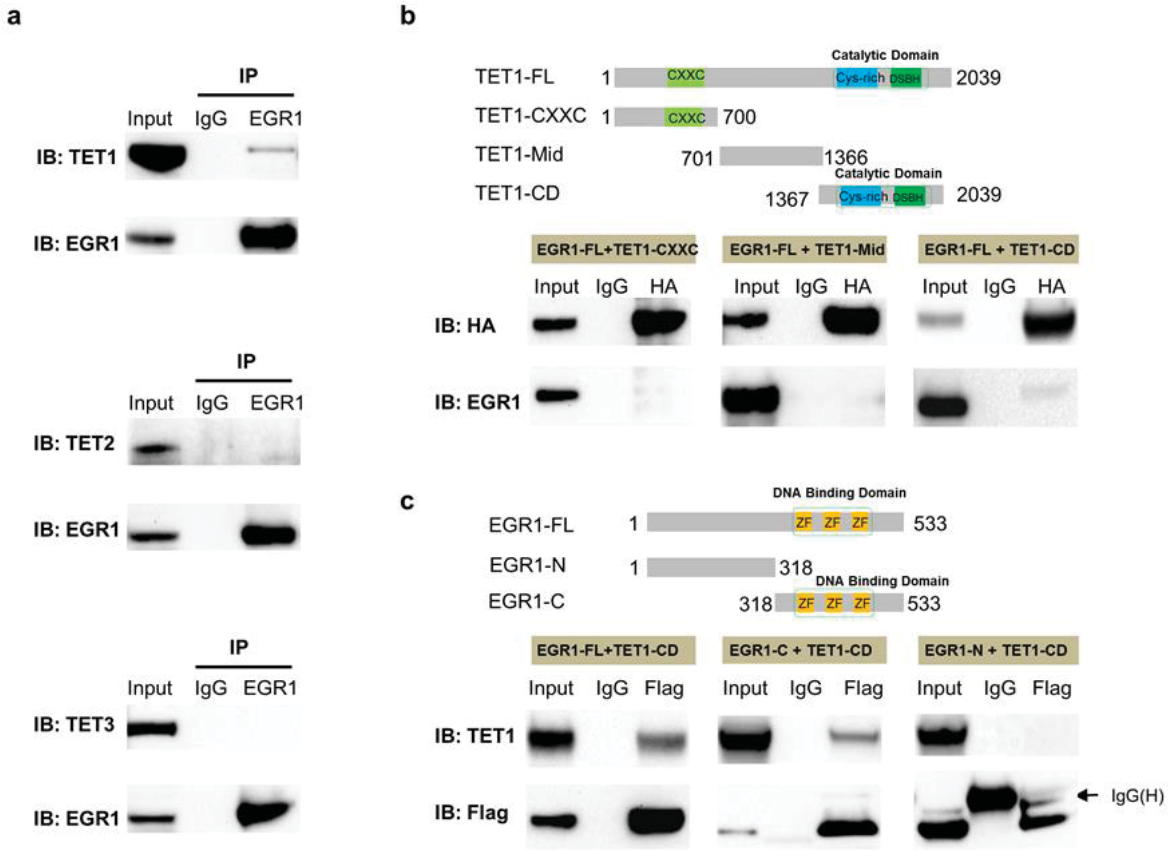


Figure 2. Identification of the protein-protein interaction between EGR1 and TET enzymes by co-immunoprecipitation.

(**a**) Endogenous association of EGR1 and TET proteins. EGR1 was immuno-precipitated from mouse frontal cortex, followed by western blot to detect TET1, TET2 and TET3. Normal rabbit IgG served as a negative control for immunoprecipitation. IP, immunoprecipitation. (**b**) Interaction between full-length EGR1 (EGR1-FL) and different deletion mutants of TET1. Flag-tagged EGR1-FL and HA-tagged TET1 deletion mutants as shown in the schematic illustration were co-expressed in HEK293T cells. (**c**) Interaction between TET1-CD and different deletion mutants of EGR1. HA-tagged TET1-CD and Flag-tagged Egr1 deletion mutants as shown in the schematic illustration were co-expressed in HEK293T cells. Protein-protein interaction was examined by IP-western blot using the antibodies indicated.

immunoprecipitates demonstrated that TET1-CD binds to the zinc-finger domain (residues 318-533) but not the N-terminal region (residues 1-318) of EGR1 (**Fig. 2c**). Altogether, these data demonstrate that EGR1 and TET1 may form a complex, and their interaction is mediated by the zinc-finger domain of EGR1 and the CD domain of TET1.

We next sought to investigate whether EGR1 recruits TET1 to its binding sites. For three EGR1 binding sites, we performed ChIP-qPCR analysis using frontal cortices from wild-type mice and observed the co-occupancy of EGR1 and TET1 (**Fig. 3a and Supplementary Fig. 4a**). To rule out the possibility that the concurrent enrichments may result from cell population heterogeneity, we conducted sequential ChIP-qPCR and confirmed EGR1 and TET1 indeed present at these loci simultaneously (**Fig. 3b**). In addition, we examined whether the depletion of EGR1 may affect the occupancy of TET1 at these loci. ChIP-qPCR assays showed TET1 enrichment on these loci was significantly decreased in the frontal cortices derived from *Egr1* knockout mice (Egr1KO) compared with those from wild-type mice (**Fig. 3c**). Taken together, these results support that EGR1 recruits TET1 to its binding sites.

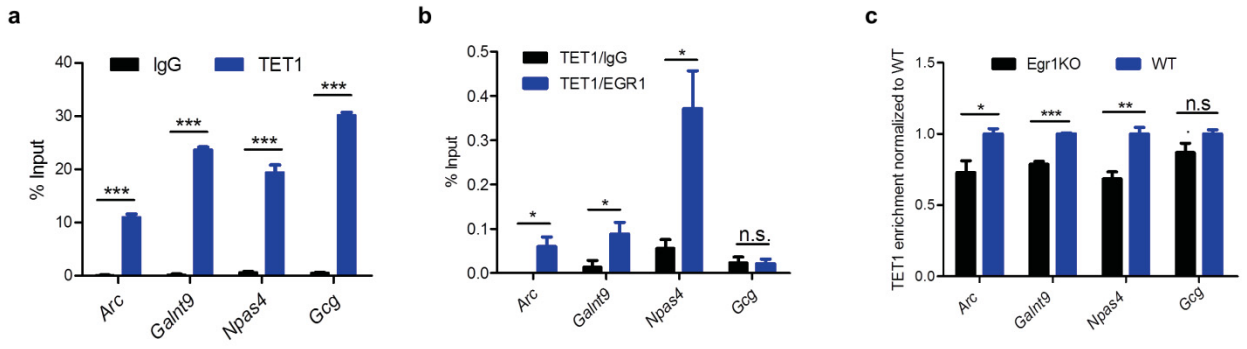


Figure 3. EGR1 recruits TET1 to its target sites.

(**a-b**) TET1 ChIP-qPCR and sequential ChIP-qPCR assay in wild-type mouse frontal cortices. *Gcg* locus serves as a negative control for EGR1 binding. (**c**) ChIP-qPCR assay in frontal cortices of Egr1KO and wild-type mice. P-values were calculated with t-test, *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$. n.s., not significant. Error bar represents s.d. from three technical replicates.

3.4.3 EGR1 coordinates with TET1 to epigenetically regulate gene expression.

We examined whether the interaction between EGR1 and TET would have an epigenetic regulatory effect. We started with three loci within *Arc*, *Galnt9* and *Npas4* genes, which were identified to be EGR1 binding sites by our ChIP-seq and validated by ChIP-qPCR to be enriched for H3K27ac enhancer mark (**Supplementary Fig. 4**). To test their enhancer activities, three

genomic fragments were individually cloned to the upstream of the EF1 promoter in the pCpG-free vector. Compared with the control vector, *Arc*, *Npas4* and *Galnt9* loci significantly promoted the gene expression from the basal EF1 promoter in the enhancer luciferase assays (**Fig. 4a**). To further examine whether their enhancer activities are under epigenetic control, prior to transfection, the constructs containing three loci were methylated *in vitro* with CpG methyltransferase, M.SssI. The methylation of these loci greatly reduced their enhancer activities (**Fig. 4 b-d**).

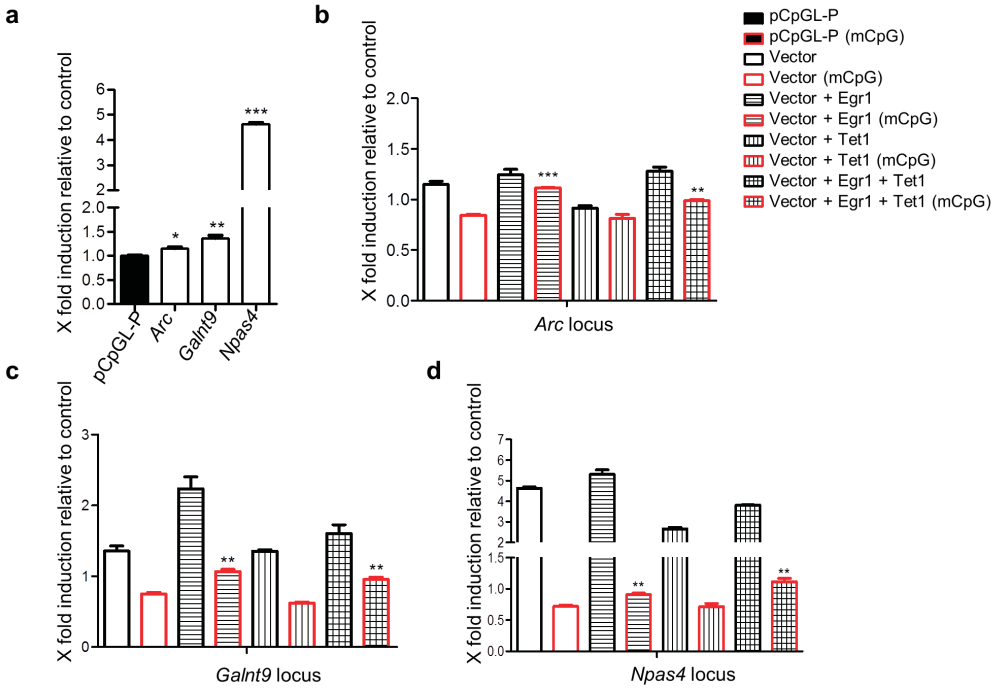


Figure 4. Cooperativity of EGR1 and TET1 modulate the enhancer activity of EGR1 binding sites.

(**a**) Luciferase reporter assays for the control vector pCpGL-P and constructs with *Arc*, *Galnt9*, *Npas4* locus. (**b-d**) Luciferase reporter assay for unmethylated and methylated *Arc* (**b**), *Galnt9* (**c**) and *Npas4* (**d**) Constructs under either EGR1 singularly or with TET1 co-expression in NSCs. Luciferase activity was measured at 48h after transfection and normalized against the activity of a co-transfected firefly construct. mCpG represents methylated constructs. P-values were determined by t-test, *, $p<0.05$, **, $p<0.01$. Values represent mean ± SD from three biological replicates. For figure 4b-d, statistical analysis was examined related to methylated vector without EGR1/TET1 overexpression.

We utilized the unmethylated and methylated reporter constructs to examine their enhancer activities in the cells with or without EGR1 or/and TET1 overexpression. For unmethylated reporter constructs, EGR1 overexpression alone could significantly increase luciferase signals of reporters. This is consistent with the fact that EGR1 as a transcriptional factor can induce the enhancer activities of its binding sites. By contrast, TET1 overexpression alone displayed no

increase in luciferase signal of these reporter vectors (**Fig. 4 b-d**). As a demethylation enzyme, TET1 may not recognize a specific DNA sequence and has no effect on enhancer activities of these loci. For methylated reporter constructs, we observed significant increases in enhancer activities when EGR1 co-overexpressed with TET1. This result suggested that EGR1 and TET1 may cooperate to activate the expression of EGR1 downstream gene.

We further examined whether the interaction between EGR1 and TET1 would lead to methylation changes on EGR1 binding sites. We focused on *Arc* and *Npas4* loci with enzyme recognition sites of HpaII, which is sensitive to DNA methylation. EGR1 and TET1 proteins increase in neuronal stem cells stimulated with KCl, which may serve as a model for the gain of function studies (**Supplementary Fig. 9a**). The loss of function studies was conducted using *Egr1* small hairpin RNA (shEgr1) and *Tet1* small hairpin RNA (shTet1-a+b), along with corresponding scrambled shRNA as negative controls. With RT-qPCR analysis, we first confirmed the downregulation of *Egr1* mRNA by shEgr1 and *Tet1* mRNA by shTet1-a+b, respectively (**Supplementary Fig. 9b**). We found that the methylation on both *Arc* and *Npas4* loci decreased in neural stem cells stimulated with KCl for 4 h, and such demethylation effect was partially abolished by the knockdown of *Egr1* or *Tet1* (**Supplementary Fig. 9c**). These data indicated both EGR1 and TET1 were required to induce the demethylation of *Arc* and *Npas4* loci.

3.4.4 Frontal cortices of Egr1KO and Tet1KO mice share aberrant DNA methylation and gene expression profiles.

It has been documented that Egr1KO mice show impaired long-term memory (Jones, Errington et al. 2001). Recent studies show *Tet1* knockout mice (Tet1KO) exhibited a significant deficiency in memory retention (Zhang, Cui et al. 2013), abnormal long-term depression and impaired memory extinction (Rudenko, Dawlaty et al. 2013). To examine the epigenetic effects of *Egr1* or *Tet1* loss, we performed methylome and transcriptome analyses for the frontal cortices derived from Egr1KO and Tet1KO mice. The genotypes of Egr1KO and Tet1KO mice were confirmed by the read coverage along *Egr1* and *Tet1* loci (**Supplementary Fig. 10**). Since EGR1 binding sites are enriched in promoters and CG rich regions, we performed reduced representation bisulfite sequencing using restriction enzymes MseI and MluCI to remove AT-rich regions. For four methylomes, we generated 211 to 287 million read pairs with an average of 140 million read pairs uniquely mapped to mouse reference genome (**Supplementary Table 3a**).

On average, we obtained methylation information for 48.9% of all CpG dinucleotides in the mouse genome and 18.1% of all CpG sites covered by at least 10 reads. Based on spiked-in unmethylated λ DNA control, the bisulfite conversion rates for four libraries were determined as 99.0% on average.

We observed strong correlations between biological replicates for two Egr1KO mice (**Supplementary Fig. 11a**) and two Tet1KO mice, respectively (**Supplementary Fig. 11b**). For the corresponding four transcriptomes, we generated 39 million read pairs on average, 86.0% of which were unambiguously mapped (**Supplementary Table 3b**). We determined pairwise Pearson's correlation at gene expression level and validated the consistency of RNA-seq results for biological replicates (**Supplementary Fig. 12**). Compared with the methylome of the frontal cortex from wild-type mice, we identified 49,991 DMSs in Egr1KO with 34,747 (69.5%) hypermethylated and 15,244 (30.5%) hypomethylated, and 113,488 DMSs in Tet1KO with 94,862 (83.6%) hypermethylated and 18,626 (16.4%) hypomethylated. To examine the association between EGR1 binding and methylation changes in KO mice, we determined the distribution of DMSs at the flanking of EGR1 binding sites. The density of hypermethylated DMSs in Egr1KO increases when approaching to the centers of EGR1 peaks, while hypomethylated DMSs in both Egr1KO and Tet1KO are depleted from EGR1 peaks (**Supplementary Fig. 13a**). When DMSs were clustered into differentially methylated regions (DMRs), the increased enrichment in EGR1 peaks was observed for hypermethylated DMRs from both Egr1KO and Tet1KO mice (**Supplementary Fig. 13b**). We next focused on the aforementioned 1,925 EGR1 binding sites, which display methylation loss from d0 to 6 weeks. Approximately 83.0% and 84.5% of these loci show increased methylation in KO mice; particularly, 19.4% and 24.7% loci are with hypermethylated DMSs in Egr1KO and Tet1KO mice respectively. These results indicate that EGR1 and TET1 are indispensable for the demethylation of some EGR1 binding sites during brain development.

Compared to wild-type mice, 322 and 2,373 DMRs were identified in the frontal cortices of Egr1KO and Tet1KO mice respectively, and these DMRs are significantly overlapped (Hypergeometric test, p-value=8.36e-14). In addition, the methylation correlation of the overlapping 184 DMRs between Egr1KO and Tet1KO is 0.88 (Pearsons' r). The knockout of *Tet1* has a broader and severer impact on the methylomes compared to the loss of *Egr1*. Intriguingly, for DMRs identified in Tet1KO mice only, moderate changes in methylation were

often observed in Egr1KO mice as well, and vice versa (**Fig. 5a**). We further divided DMRs into hypermethylated or hypomethylated in either Egr1KO or Tet1KO and obtained their methylation profiles across developmental stages and in neuronal cell types (**Supplementary Fig. 14**). Methylation loss during development and in neurons was observed for around 78.0%, and 56.2% of DMRs identified in Egr1KO and Tet1KO mice, respectively. Interestingly, 9.1% Egr1KO DMRs and 13.9% Tet1KO DMRs were found to be constantly hypomethylated across developmental stages but with increased methylation in KO mice. This result suggests EGR1 and TET1 are required for the maintenance of demethylation statues for some genomic loci, which are not or lowly methylated since early brain development (at E11.5 or earlier).

Compared to wild-type controls, 896 and 1,359 differentially expressed genes (DEGs) were determined in Egr1KO and Tet1KO mice respectively with 529 of them shared (**Fig. 5b**). Similar to what observed for methylomes, for these 529 genes, the correlation in the expression changes *vs* wild type is 0.94 (p-value=1.14e-256) between Tet1KO mice and Egr1KO mice. For all DEGs identified either in Egr1KO or Tet1KO mice, strong correlations were observed between the two kinds of KO mice. GO annotation analyses showed that both Egr1KO and Tet1KO DEGs are involved in several biological processes related to central nervous system development or neural tube development, including potassium ion transport and Notch signaling pathway which plays critical roles in brain development (Lasky and Wu 2005) (**Supplementary Fig. 15 and Supplementary Table 4***).* To explore the relationship between DMR methylation and gene expression, we calculated the Spearman's correlation coefficients of hypermethylated DMRs identified in Egr1KO or Tet1KO mice (**Supplementary Fig. 16**). Negative correlations between methylation level and gene expression were observed for DMRs in 5'-UTR, Promoter, and Distal Promoters. In addition, significant increases in methylation were observed in KO mice for the three EGR1 binding sites within *Arc*, *Galnt9* and *Npas4* genes (**Fig. 5c**). The methylation levels of these loci, *Arc* and *Galnt9* in particular, are negatively correlated with gene expression (**Fig. 5d**).

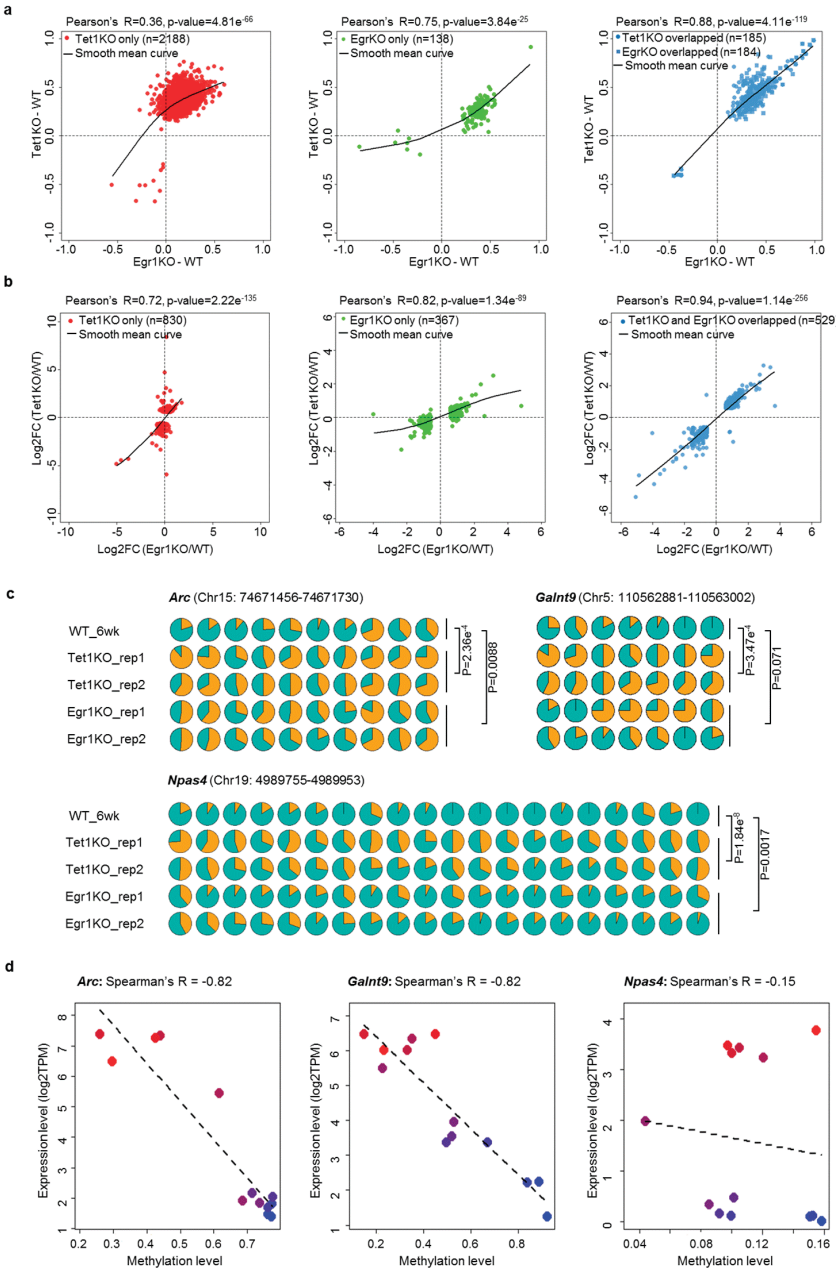


Figure 5. Correlations of DNA methylation and gene expression profiles between Egr1KO and Tet1KO frontal cortices.

Methylation correlations (a) and Gene expression correlations (b) between Egr1KO and Tet1KO mice. (c) Methylome sequencing results for Arc, Galnt9, Npas4 loci. Each CpG is represented by a circle; yellow in circles indicate the percentage of methylation in each CpG site. The statistical significance of methylation differences between Egr1/Tet1KO and WT mice was evaluated with the Wilcoxon rank-sum test. (d) The correlations between DNA methylation levels of Arc, Galnt9, Npas4 loci and corresponding gene expression during brain development from embryonic day 11.5 (E11.5) (denoted in blue color) to 22 months (22mo) (denoted in red color).

Lastly, hypermethylated DMRs identified in either Egr1KO or Tet1KO show low methylation in excitatory neurons compared with PV and VIP neurons (**Supplementary Fig. 14**). This prompts us to make use of single cell brain methylome data (Luo, Keown et al. 2017) for additional bioinformatics analyses on the cell-subtype-specific function of EGR1 bindings. We confirmed that the hypermethylated DMSs identified in Egr1KO mice are significantly enriched on excitatory-neuron-specific hypomethylated regions determined in single cell analyses (**Supplementary Fig. 17**), especially for excitatory-neuron-subtype mL5-1 (odds ratio=1.4, Binomial test, p-value=4.8e-42). EGR1 binding sites are significantly enriched on excitatory-neuron-specific hypomethylated regions but excluded from inhibitory-neuron-specific ones (**Supplementary Fig. 18**). For instance, the enrichment of EGR1 binding sites on hypomethylated regions in excitatory-neuronal subtype mL4 is highly significant (odds ratio: 1.7, Binomial test, p-value=6.7e-83). In addition, the enrichment of EGR1 binding is correlated with the bindings of other early response genes including transcription factors induced by neuronal activity, such as JUNB, FOSB, CFOS, and NPAS4 (**Supplementary Fig. 19**).

3.5 DISCUSSION

The link between epigenetic changes and neuronal activity has been well established, together with the gradual recognition of critical roles of TET DNA demethylases in learning and memory (Kaas, Zhong et al. 2013, Rudenko, Dawlaty et al. 2013, Yu, Su et al. 2015, Zhu, Girardo et al. 2016). Apparently, epigenetic changes upon neuronal activation are not random but TET enzymes do not display DNA binding specificity. Our study shows how TET1 gains its specificity *via* the interaction with EGR1, a sequence-specific DNA binding protein. On the other hand, as a key member of immediate early genes, *Egr1* has been known for decades to play an essential role in transcriptional response to environmental stimuli. *Egr1* is an important mediator of the effects of early-life experience and directly regulate genes controlling synaptic plasticity in both physiological and pathological conditions (Jones, Errington et al. 2001, Mataga, Fujishima et al. 2001, Weaver, Cervoni et al. 2004, Li, Carter et al. 2005). *Egr1* expression has been widely used as a marker for neuronal activation but how it leaves memory trace remains elusive. In this study, we provided a key piece of evidence that may help in solving this puzzle at the epigenetic level. Although neurogenesis and neuronal migration are largely completed at birth, the postnatal brains continue forming synapses and neural circuits and undergo activity-

dependent refinements. *Egr1* gene has been shown to control newborn neuron selection and maturation during the critical period of a few weeks after birth (Veyrac, Gros et al. 2013). The decoding of epigenetic machinery during this developmental period is critical for a complete understanding of the mechanisms that underlie late-stage refinement of maturing neuronal circuits. Of note, *Egr1* gene continues to have functions in the adult brain and may have pathological significance in the Alzheimer's disease (Renbaum, Beeri et al. 2003).

Our study provides several key pieces of evidence for the interaction of EGR1 and TET1. First, our result reveals that extensive DNA demethylation occurs in thousands of EGR1 binding sites during the postnatal frontal cortex development. Second, the zinc-finger domain of EGR1 and the CD domain of TET1 are required for the interaction of the two proteins. The co-occupancy of EGR1 and TET1 at target loci were confirmed with sequential ChIP analyses. In the presence of EGR1, TET1 is capable to achieve locus-specific demethylation and activate the expression of EGR1 downstream genes. Third, both EGR1 and TET1 are indispensable for the demethylation of a common set of EGR1 binding sites that show aberrant DNA methylation in Egr1KO and Tet1KO mice. Altogether, our data support a model that links environmental stimuli to brain methylome programming (**Supplementary Fig. 20***).* At birth, a subset of *Egr1*-mediated and neuronal activity-induced genes are silenced with methylated EGR1 binding sites. During early postnatal development, the over-expressions of *Tet1* and *Egr1* upon neuronal activation demethylate EGR1 binding sites and shift the genes to either “Poised” or “ON” states.

Our study raises a few interesting questions. Could EGR and TET interaction become a general mechanism for various kinds of cells to keep epigenetic memory in response to stimuli? EGR family members are involved in a variety of biological processes and the epigenetic memory may not be limited to the nervous system. For instance, *Egr1, Egr2* and *Egr3* have been shown to be critical for the response to external signals and to direct lineage differentiation in the immune system (Safford, Collins et al. 2005, Seiler, Mathew et al. 2012). In our study, we found Egr1KO and Tet1KO have significant effects but limited to less than 30% of the 1,925 EGR1 binding sites which show demethylation during postnatal brain development. Compared with the loss of *Egr1*, the *Tet1* knockout has a much severe outcome and broader impact in the epigenome. These results suggest a compensation mechanism to rescue the loss of *Egr1* or *Tet1* in the frontal cortex and TET1 may have other partners to recognize genomic loci where EGR1 can bind. Lastly, the integrated bioinformatics analyses with single neuron methylomes suggest the methylation

changes of EGR1 binding sites are largely restricted to a subset of excitatory neurons. It would be interesting to explore whether different kinds of neurons would adopt distinct epigenetic programming mechanisms in future.

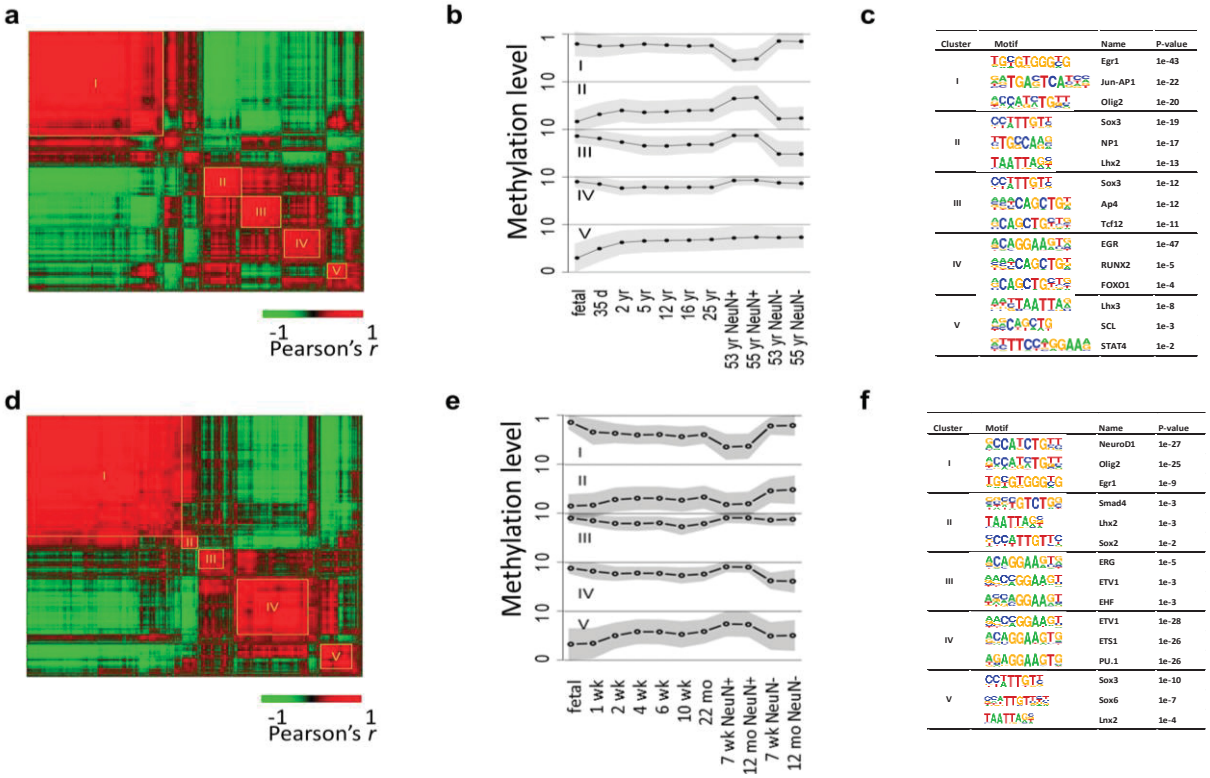
3.6 ACKNOWLEDGEMENTS

The authors thank Drs. Joseph, R. Ecker, Ryan Lister and Eran A. Mukamel for sharing brain methylome data, Drs. Bing Ren and Bradley E Bernstein for sharing ChIP-seq data, and the labs contributing to ENCODE project. This work was supported by NIH grant NS094574 and the Biocomplexity Institute faculty development fund for H.X.

3.7 AUTHOR CONTRIBUTIONS

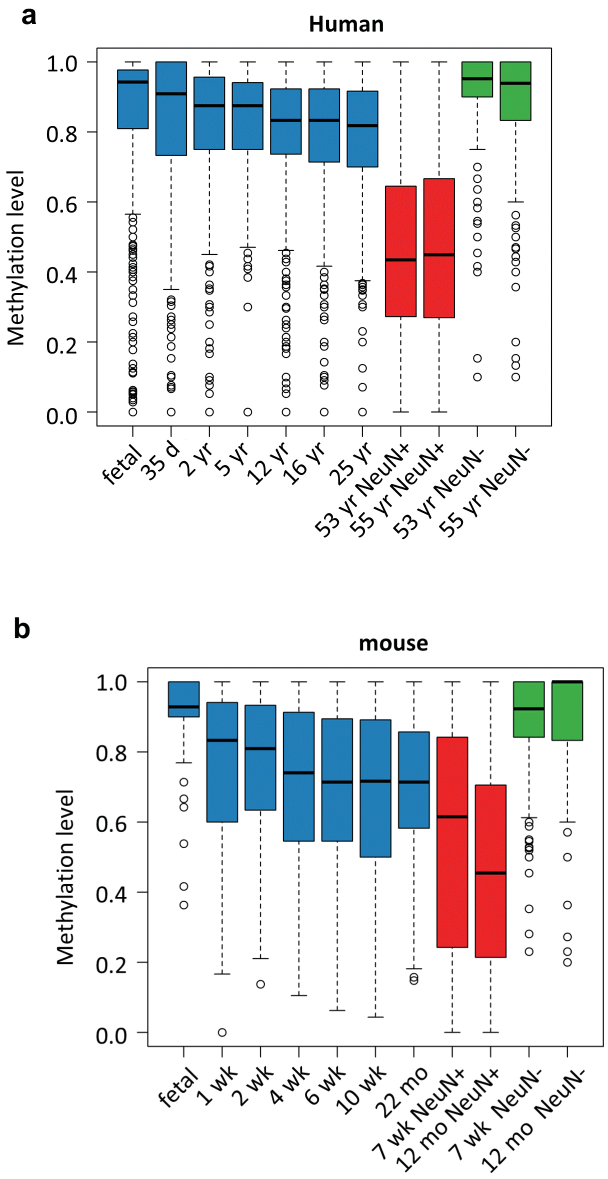
H.X. conceived and designed the study; Z.S. and A. M. conducted ChIP-seq, Co-IP experiments and luciferase assays; X.X. constructed the RRBS, RNA-seq libraries; M.S., E.X., and J. H. analyzed ChIP-seq, RNA-seq, RRBS data; X.W., E.M., J.X., L.L., J.Z., J.C., A.M., and M.T. collected and processed neural stem cells and tissue samples; Z.S., X.X, J.H., M.S. and H.X. wrote the manuscript. All authors discussed the results and commented on the manuscript.

3.8 SUPPLEMENTARY DATA



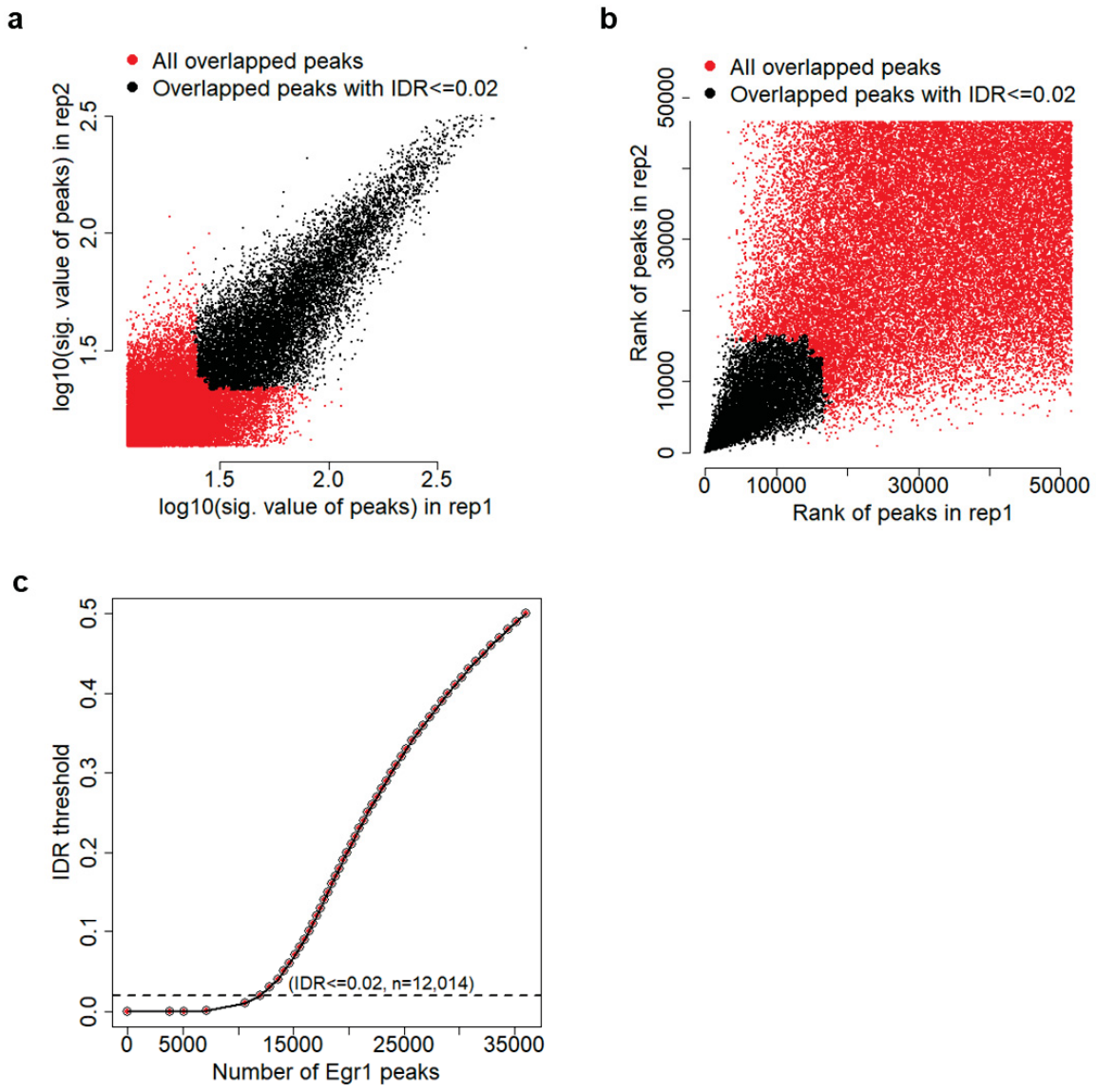
Supplementary Figure 1. Co-methylation analysis of bipolar methylated loci during mouse and human brain development.

(a, d) Heatmaps displaying the hierarchical clustering of human and mouse bipolar methylated loci, respectively. The top five largest co-methylated modules are marked with yellow squares. (b, e) The methylation level distributions of five co-methylated modules for human and mouse, respectively. The black solid line shows the average, and the grey color shows the standard deviation. (c, f) Top TF motifs enriched in each module identified in human and mouse brain methylomes, respectively (for the full lists, see **Supplementary Table 1**).



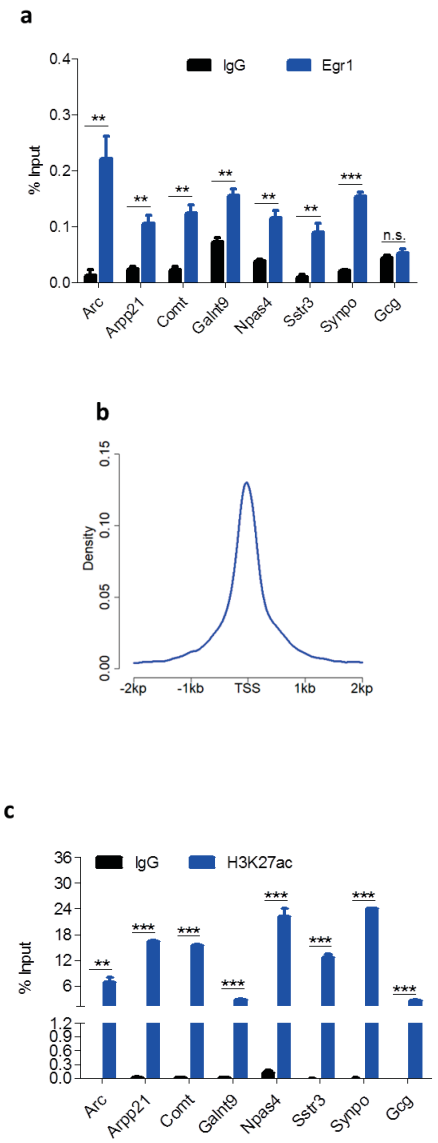
Supplementary Figure 2. Methylation profiles of EGR1 binding sites predicted within co-methylation module I.

The methylation profiles of the predicted EGR1 binding sites during human (a) and mouse (b) brain development and cell specifications. Only binding sites with at least ten methylation calls in every methylome were included.



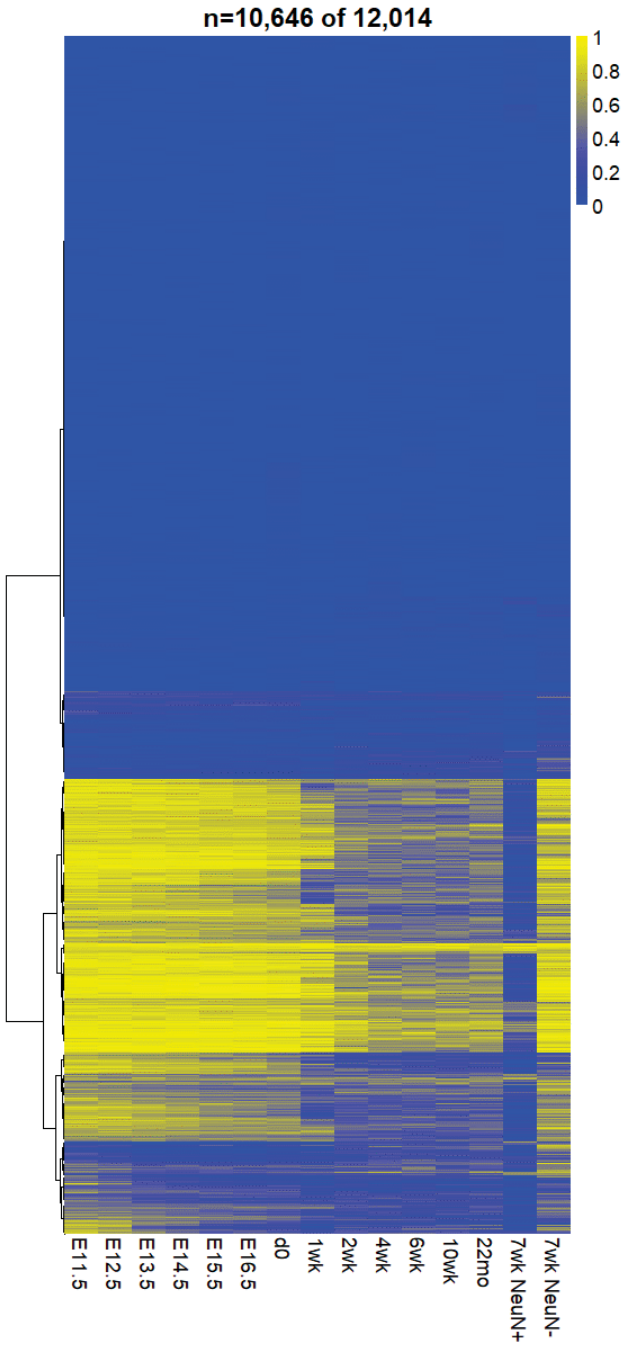
Supplementary Figure 3. Consistency of two biological replicates for EGR1 ChIP-Seq assays.

(a) Scatter plots of signal scores for peaks that overlap in two biological replicates. IDR: irreproducible discovery rate. (b) Scatter plots of ranks for peaks that overlap in two biological replicates. Note that low ranks correspond to high signals, and vice versa. (c) The estimated IDR as a function of different rank thresholds. (a, b) Black data points represent peak pairs that pass an IDR threshold of 2%, whereas the red data points represent peak pairs that do not pass the IDR threshold of 2%. Two EGR1 ChIP-seq replicates show high reproducibility with 12,014 peaks passing 2% IDR threshold.



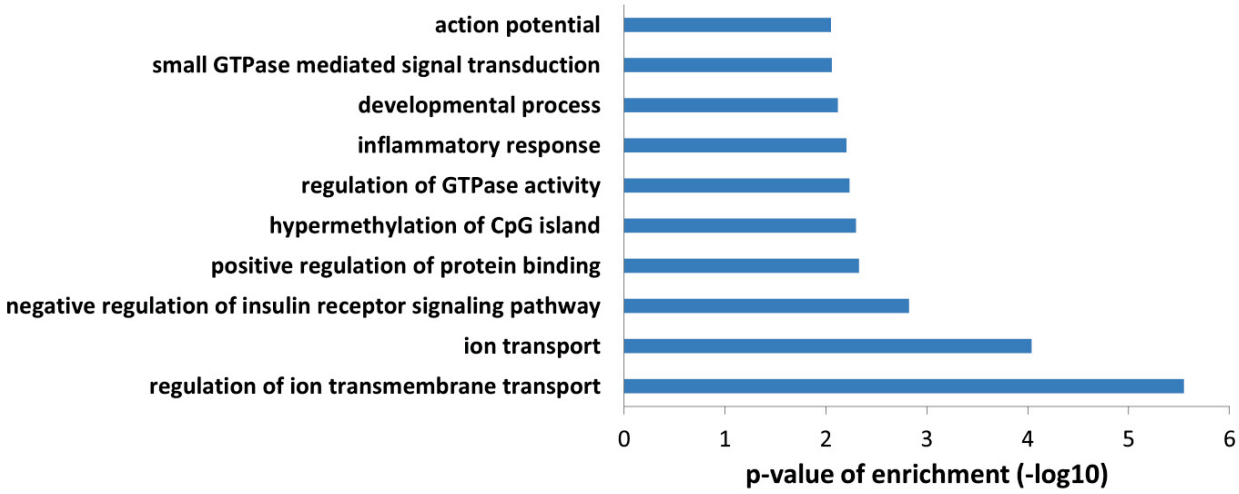
Supplementary Figure 4. Enrichment of EGR1 peaks around TSS and ChIP-qPCR validation for EGR1 binding and histone mark H3K27ac.

(a, c) ChIP-qPCR analysis for EGR1 (a) and H3K27ac (c) at eight genomic regions. *Arc* and *Gcg* are positive and negative control for EGR1 binding, respectively. (b) Enrichment of EGR1 peaks around TSS. Analysis is by t-test, *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$. n.s., not significant. Error bar represents standard deviation (s.d.) from three technical replicates.



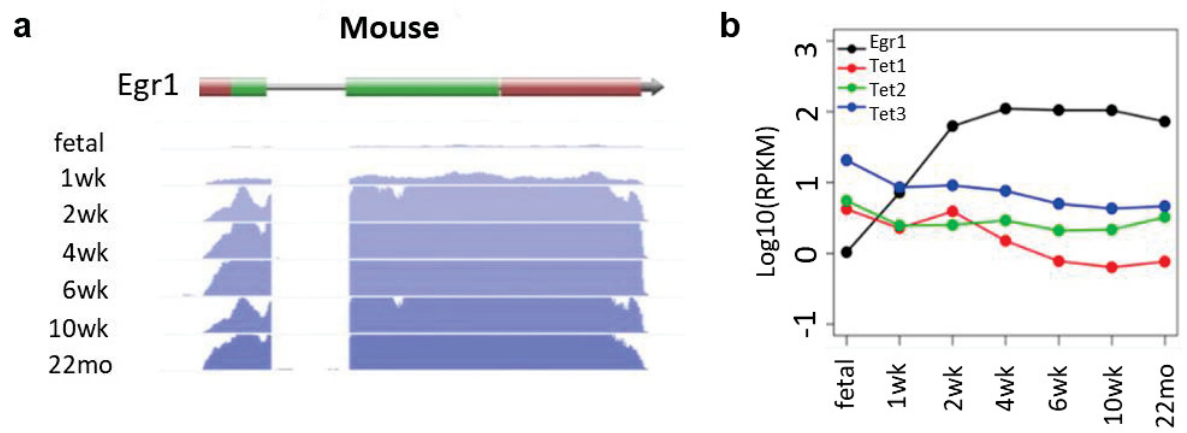
Supplementary Figure 5. Methylation profiles of EGR1 peaks during mouse brain development.

Methylation profiles of EGR1 peaks in mouse frontal cortices from embryonic day 11.5 to 22 months, neuron, and glia at 7 weeks. The average of methylation levels within EGR1 binding sites were calculated based on CpG sites with at least 5 reads covered. EGR1 binding sites with missing data in any of the 15 samples were discarded.



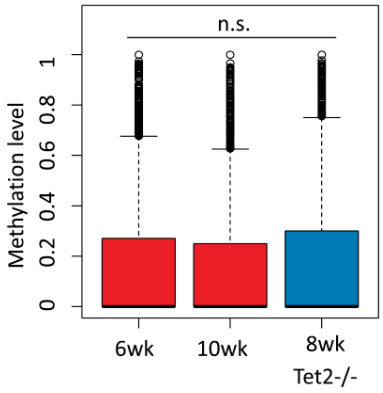
Supplementary Figure 6. GO analysis for genes (n = 598, which have been annotated by DAVID) with TSS flanking 10 kb of EGR1 peaks.

Genes were identified with 1,925 EGR1 peaks which show decreased methylation during frontal cortex development and hypomethylated in adult neurons compared to that in adult glial cells.

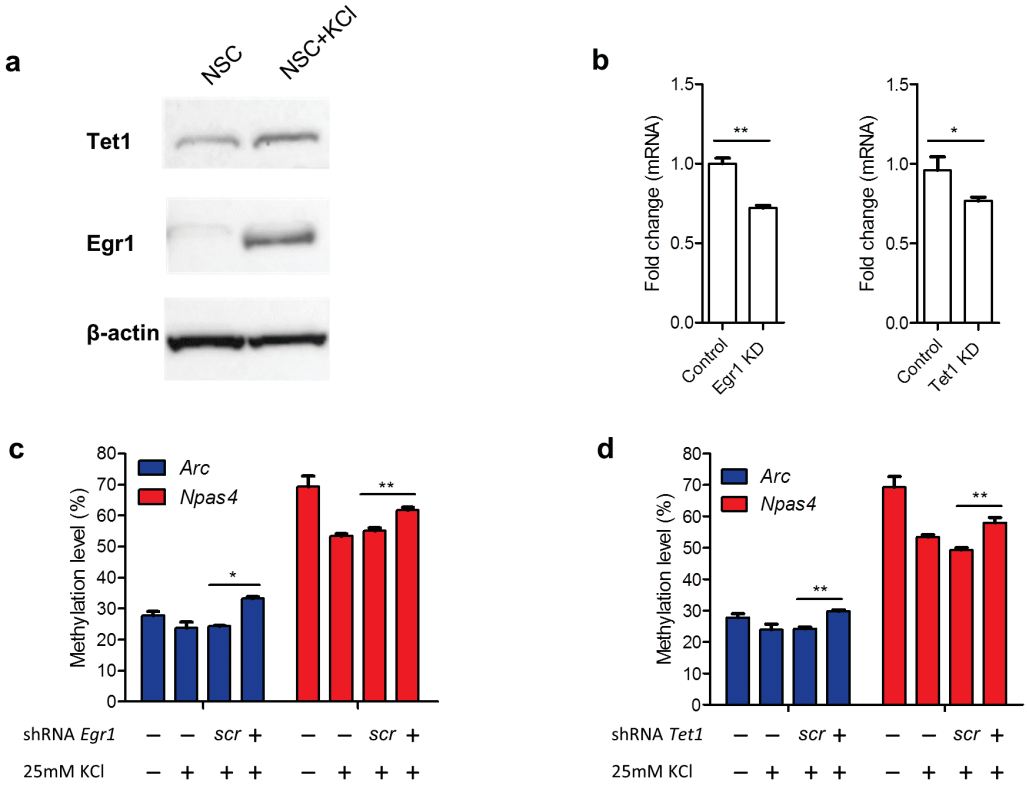


Supplementary Figure 7. Expression profiles of *Egr1*, *Tet* genes during mouse brain development.

(a) Transcript abundance for *Egr1* and (b) three *Tet* genes based on RNA-Seq data for frontal cortices during mouse brain development. RPKM, reads per kilobase of exon per million fragments mapped.

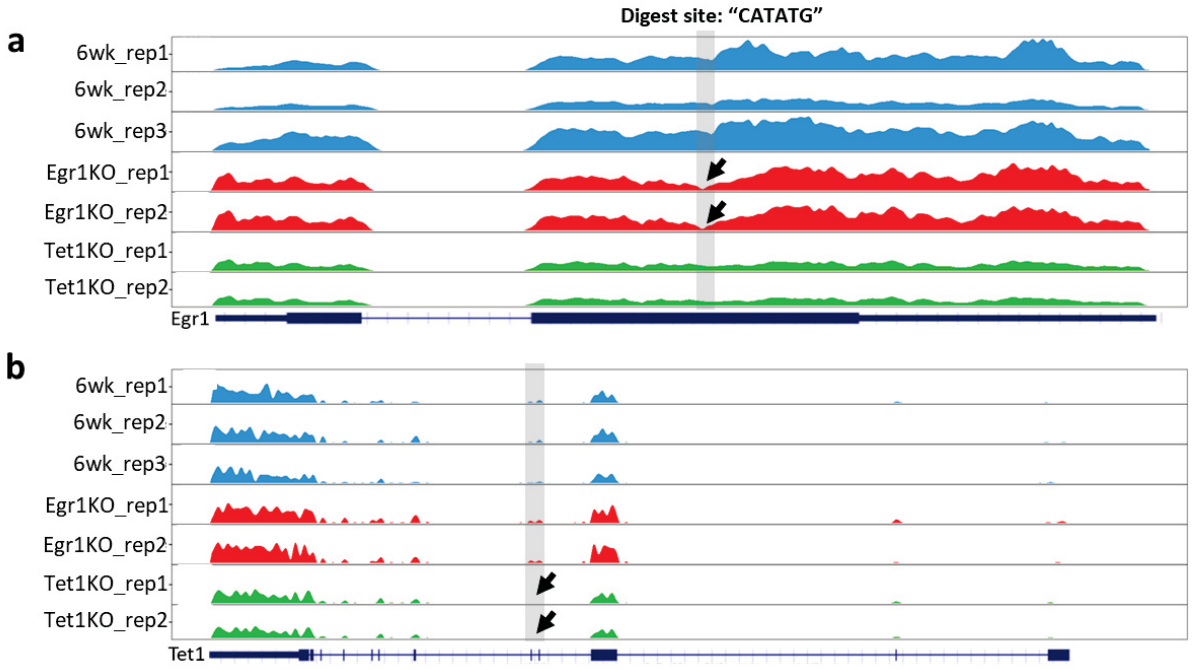


Supplementary Figure 8. Methylation profiles of EGR1 binding sites in frontal cortices of wild-type mice at 6-week, 10-week, and Tet2 knockout (Tet2-/-) mice at 8-week. ns, not significant.



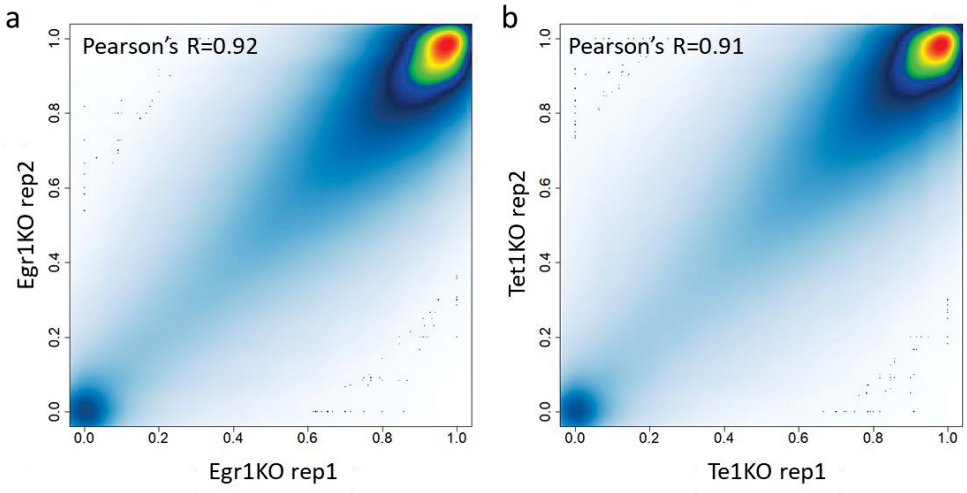
Supplementary Figure 9. Methylation analysis of *Arc*, *Npas4* loci by methylation-sensitive qPCR (MS-qPCR) in NSCs.

(a) Western blotting analysis of EGR1 and TET1 in NSCs with and without KCl stimulation. (b-c) RT-qPCR analysis of *Egr1* and *Tet1* expression in NSCs transfected with shEgr1, shTet1, or scrambled shRNA. (d) Methylation analysis of *Arc* and *Npas4* loci by MS-qPCR. Analysis is by t-test, *, $p<0.05$, **, $p<0.01$. Error bar represents standard deviation (s.d.) from three technical replicates.



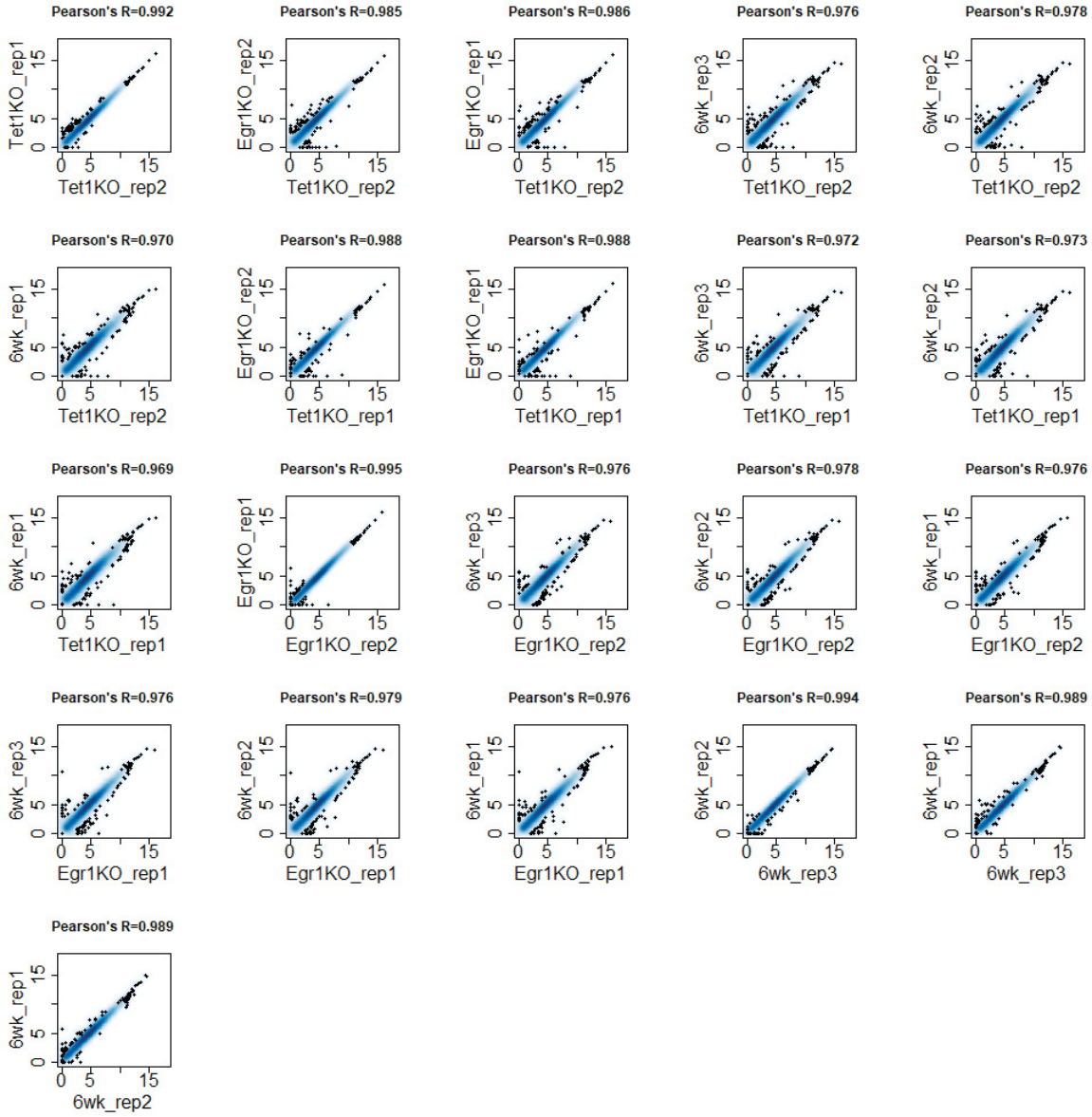
Supplementary Figure 10. Visualization of read coverage along (a) *Egr1* and (b) *Tet1* genes.

Each transcriptome has been normalized to 1 million reads. Gray boxes highlight the genomic regions and black arrows indicate specific genetic aberrations in the genomes of Egr1KO and Tet1KO mice.



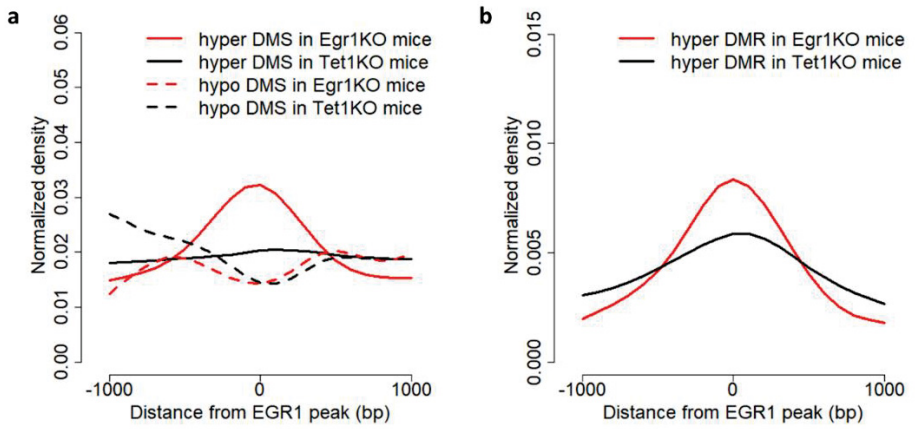
Supplementary Figure 11. Consistency of methylation on pairwise-replicate.

Consistency of RRBS-seq for two biological replicates of (a) Egr1KO mice and (b) Tet1KO mice. Methylation levels were calculated with CpG sites with at least 10 reads covered. Colors represent the density of CpG sites, in which red color denotes high density and blue color denotes low density. Black points represent the areas of lowest regional density.



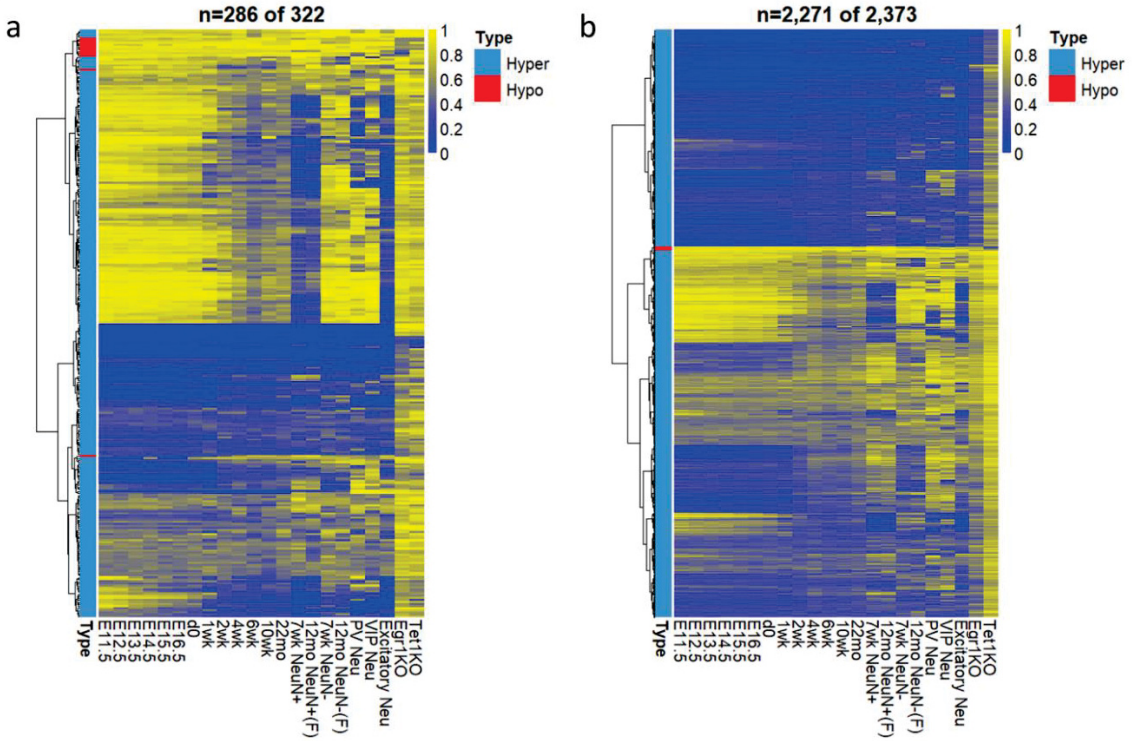
Supplementary Figure 12. Correlation analysis of the transcriptomes derived from the frontal cortices of Egr1KO, Tet1KO and wild-type mice.

Scatter plots represent the correlations of any sample pair. Each black point denotes a gene. Dark blue denotes high density of points. The expression level for each gene was presented by log2TPM. Two biological replicates were included in the analyses for Egr1KO or Tet1KO mice, and three biological replicates were included for age/gender-matched wild-type mice.



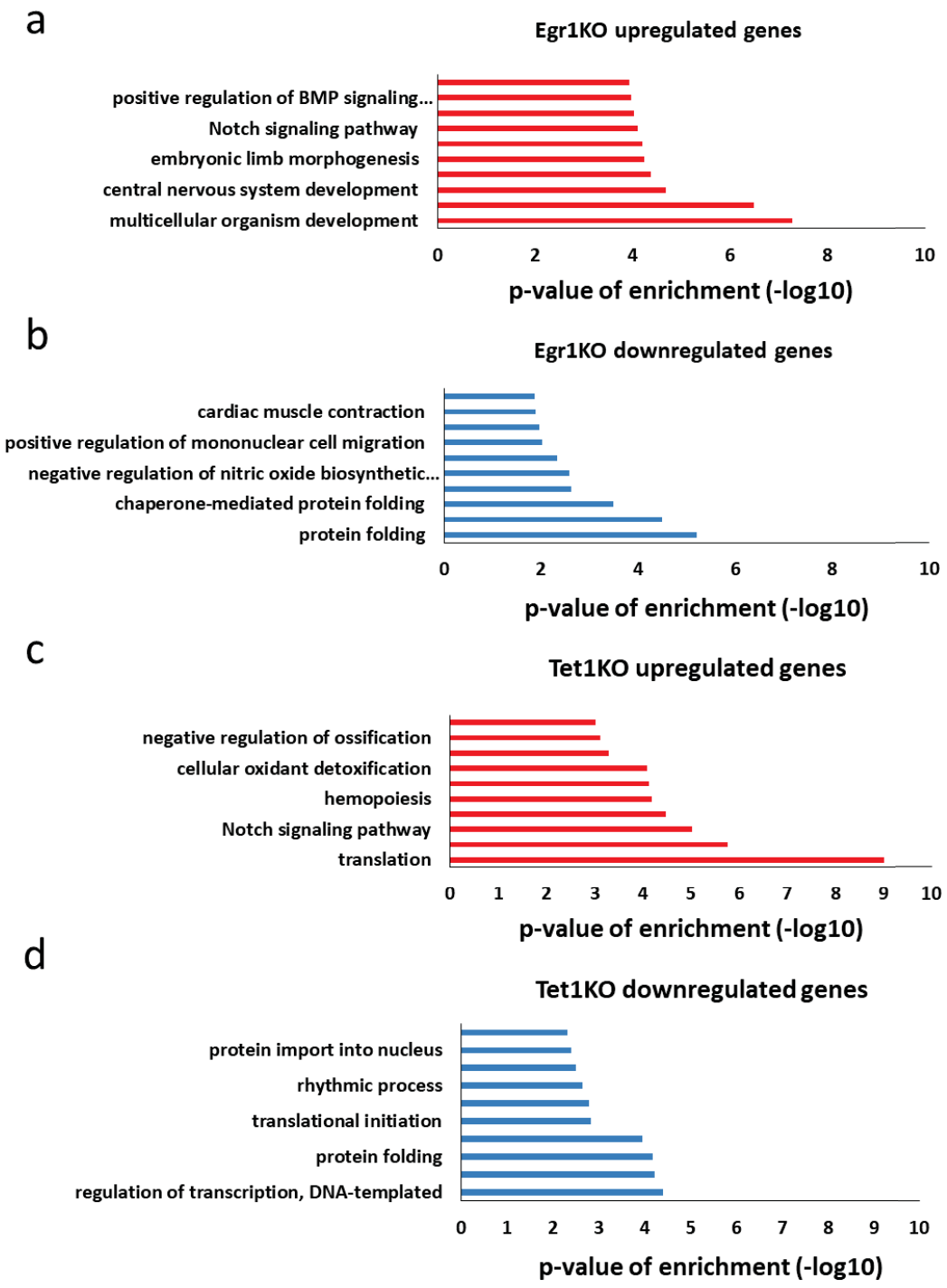
Supplementary Figure 13. Distribution of (a) differentially methylated CpG sites (DMSs) and (b) differentially methylated regions (DMRs) surrounding EGR1 binding sites.

Red color and black color represent differentially methylated loci from Egr1KO mice and Tet1KO mice, respectively. Solid line and dash line denote hypermethylated loci and hypomethylated loci, respectively.



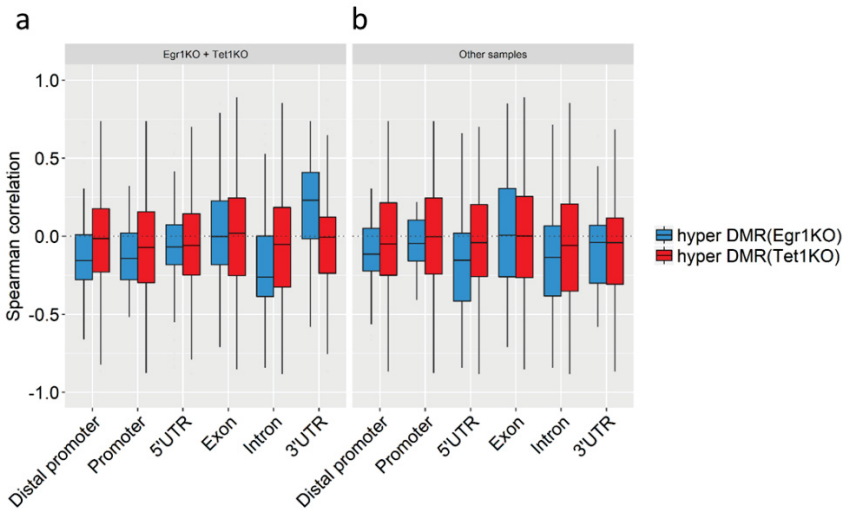
Supplementary Figure 14. Methylation profiles of DMRs identified from (a) Egr1KO mice and in (b) Tet1KO mice during mouse brain development and cell types.

The average of methylation levels within DMRs were calculated with CpG sites with at least 5 reads covered. Blue and red color bars represent hypermethylated and hypomethylated DMRs, respectively.



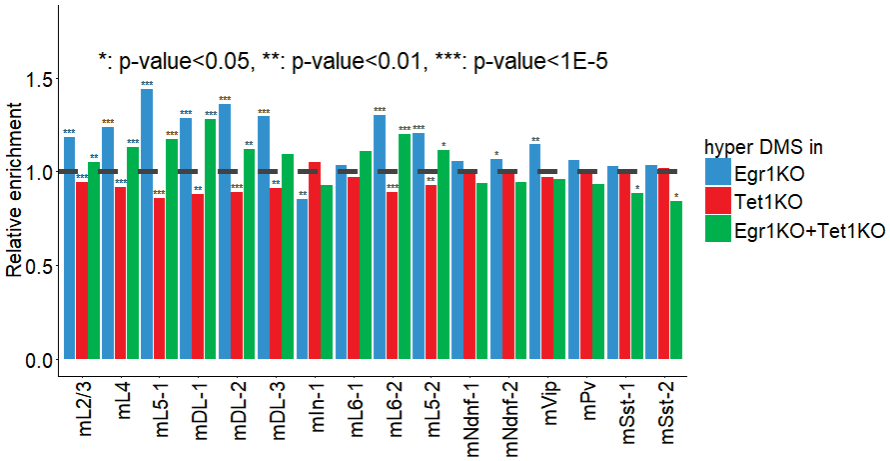
Supplementary Figure 15. GO analysis of differentially expressed genes identified in Egr1KO or Tet1KO frontal cortices.

GO analysis of (a) upregulated genes in Egr1KO, (b) downregulated genes in Egr1KO, (c) upregulated genes in Tet1KO and (d) downregulated genes in Tet1KO.



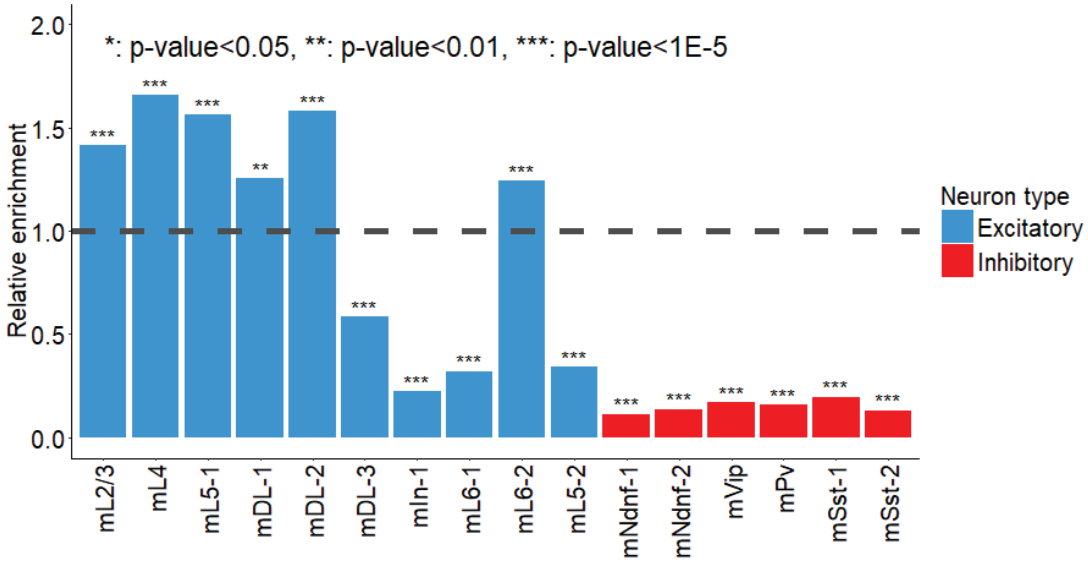
Supplementary Figure 16. Spearman's correlation between DMR methylation and adjacent gene expression.

(a) Correlations were calculated across genes overlapped with DMRs in Egr1KO mice and Tet1KO mice. (b) Correlations were calculated for genes overlapped with DMRs across 33 mouse brain samples with methylome and transcriptome data available (Supplementary Table 4). Blue and red color bars represent hypermethylated DMRs from Egr1KO mice and Tet1KO mice relative to 6-week old wild-type mice, respectively.



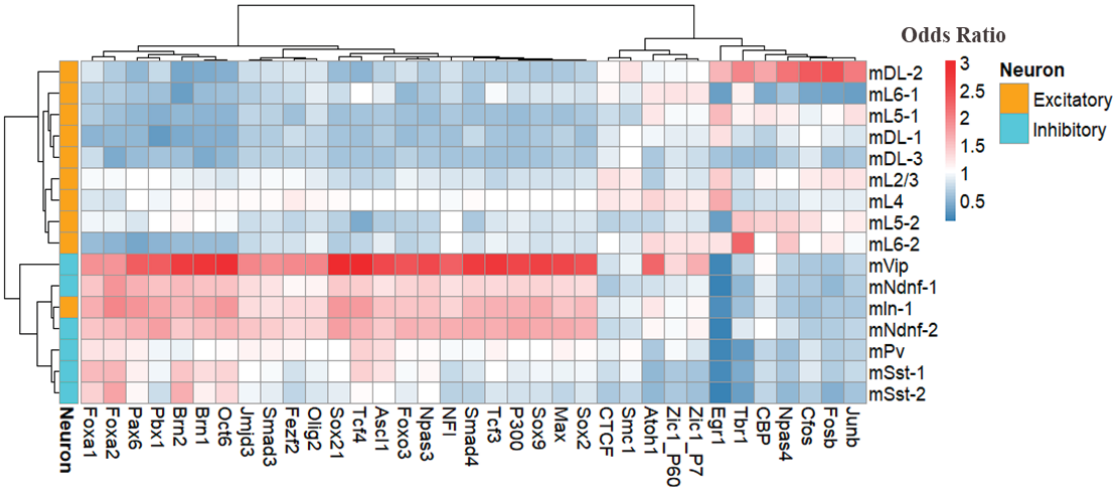
Supplementary Figure 17. Relative enrichment of hypermethylated DMSs on neuron subtype-specific DMRs.

Blue, red and green color represent the hypermethylated DMS from Egr1KO mice only, Tet1KO mice only, in both Egr1KO mice and Tet1KO mice, respectively. The neuron subtype-specific DMRs were reported by Luo *et al.* (Science, 2017). The enrichment significance was evaluated using Fisher Exact test across 16 neuron subtypes. Enrichment score denotes odds ratio and all hypermethylated DMSs were used as controls for three individual lists.



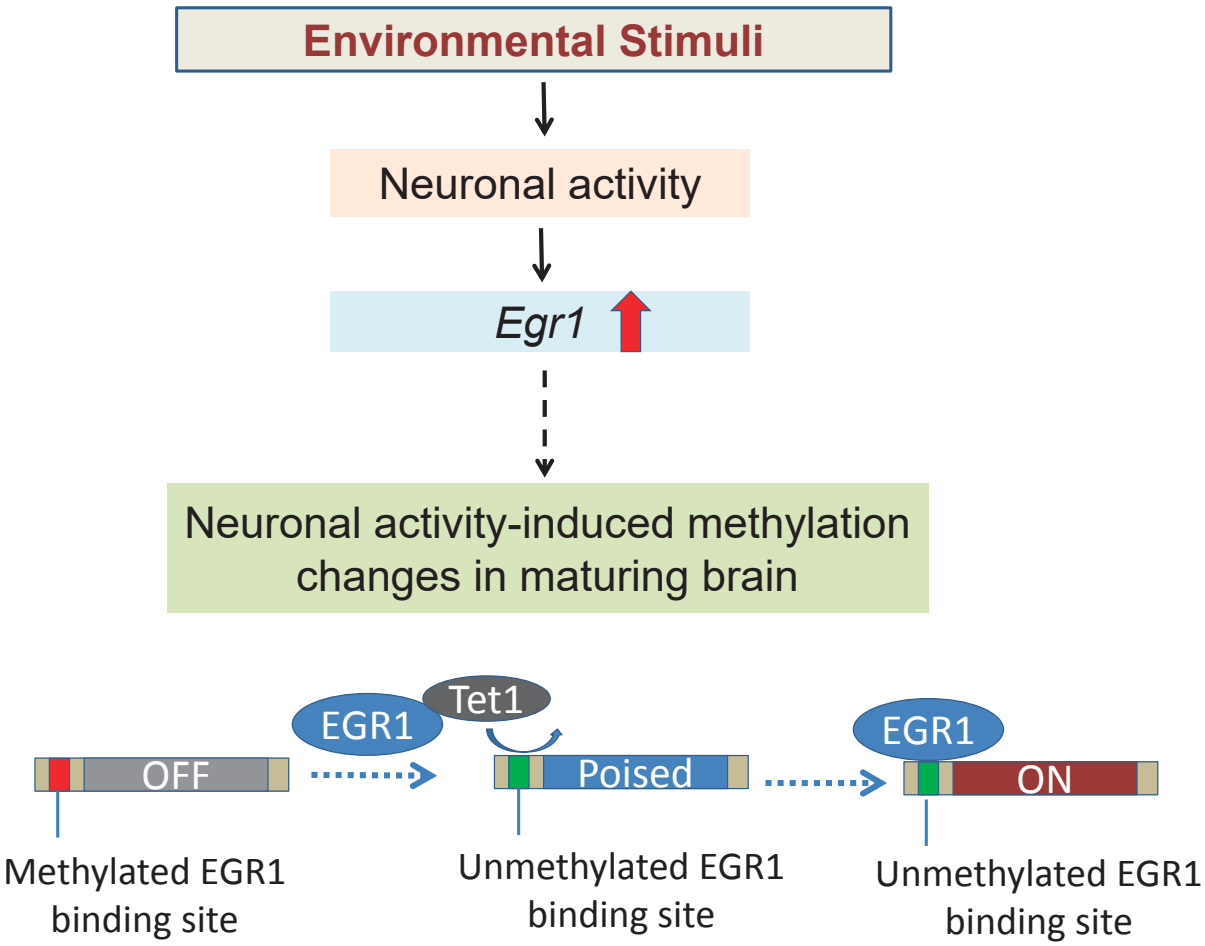
Supplementary Figure 18. Relative enrichment of EGR1 binding sites on neuron subtype-specific DMRs.

Blue color represents excitatory neuron subtypes and red color denotes inhibitory neuron subtypes. The significance of enrichment was evaluated by Fisher Exact test across 16 neuron subtypes. Enrichment score denotes odds ratio and all 16 neuron subtypes together were used as controls for each individual neuron subtype.



Supplementary Figure 19. Clustering of TFs relatively enriched on neuron subtype-specific DMRs.

Orange color in heatmap represents excitatory neuron subtypes and cyan color denotes inhibitory neuron subtypes. The significance of enrichment was evaluated by Fisher Exact test across 16 neuron subtypes. Enrichment score in each entity of matrix denote odds ratio and all 16 neuron subtypes together were used as controls for each individual neuron subtype.



Supplementary Figure 20. A simplified model for EGR1 and TET1 interaction linking environmental stimuli to brain methylome programming.

At birth, *Egr1*-mediated and neuronal activity-induced genes are silenced with methylated EGR1 binding sites. During postnatal development and upon neuronal activity, the increase in expression of *Tet1* and *Egr1* leads to the demethylation of EGR1 binding sites and shifts the genes to either "Poised" or "ON" states.

3.9 REFERENCES

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble (2009). "MEME SUITE: tools for motif discovery and searching." Nucleic Acids Res **37**(Web Server issue): W202-208.

Bancroft, T., C. Du and D. Nettleton (2013). "Estimation of false discovery rate using sequential permutation p-values." Biometrics **69**(1): 1-7.

Blecher-Gonen, R., Z. Barnett-Itzhaki, D. Jaitin, D. Amann-Zalcenstein, D. Lara-Astiaso and I. Amit (2013). "High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states." Nat Protoc **8**(3): 539-554.

Christy, B. and D. Nathans (1989). "DNA binding site of the growth factor-inducible protein Zif268." Proc Natl Acad Sci U S A **86**(22): 8737-8741.

Dawlaty, M. M., K. Ganz, B. E. Powell, Y. C. Hu, S. Markoulaki, A. W. Cheng, Q. Gao, J. Kim, S. W. Choi, D. C. Page and R. Jaenisch (2011). "Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development." Cell Stem Cell **9**(2): 166-175.

de la Rica, L., J. Rodriguez-Ubreva, M. Garcia, A. B. Islam, J. M. Urquiza, H. Hernando, J. Christensen, K. Helin, C. Gomez-Vaquero and E. Ballestar (2013). "PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation." Genome Biol **14**(9): R99.

Feng, J., Y. Zhou, S. L. Campbell, T. Le, E. Li, J. D. Sweatt, A. J. Silva and G. Fan (2010). "Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons." Nat Neurosci **13**(4): 423-430.

Gu, T. P., F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li and G. L. Xu (2011). "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes." Nature **477**(7366): 606-610.

Guo, J. U., D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao and H. Song (2011). "Neuronal activity modifies the DNA methylation landscape in the adult brain." Nat Neurosci **14**(10): 1345-1351.

Guo, J. U., Y. Su, C. Zhong, G. L. Ming and H. Song (2011). "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain." Cell **145**(3): 423-434.

Hashimoto, H., Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang and X. Cheng (2014). "Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence." Genes Dev **28**(20): 2304-2313.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.

Huang, D. W., B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane and R. A. Lempicki (2007). "The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists." Genome Biol **8**(9): R183.

Huang, H., X. Jiang, Z. Li, Y. Li, C.-X. Song, C. He, M. Sun, P. Chen, S. Gurbuxani, J. Wang, G.-M. Hong, A. G. Elkahloun, S. Arnovitz, J. Wang, K. Szulwach, L. Lin, C. Street, M. Wunderlich, M. Dawlaty, M. B. Neilly, R. Jaenisch, F.-C. Yang, J. C. Mulloy, P. Jin, P. P. Liu, J. D. Rowley, M. Xu, C. He and J. Chen (2013). "TET1 plays an essential oncogenic role in MLL-rearranged leukemia." Proceedings of the National Academy of Sciences of the United States of America **110**(29): 11994-11999.

Jones, M. W., M. L. Errington, P. J. French, A. Fine, T. V. Bliss, S. Garel, P. Charnay, B. Bozon, S. Laroche and S. Davis (2001). "A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories." Nat Neurosci **4**(3): 289-296.

Kaas, G. A., C. Zhong, D. E. Eason, D. L. Ross, R. V. Vachhani, G. L. Ming, J. R. King, H. Song and J.

D. Sweatt (2013). "TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation." Neuron **79**(6): 1086-1093.

Kennedy, A., E. M. Schmidt, A. P. Cribbs, H. Penn, P. Amjadi, K. Syed, J. E. Read, P. Green, B. Gregory and F. M. Brennan (2014). "A novel upstream enhancer of FOXP3, sensitive to methylation-induced silencing, exhibits dysregulated methylation in rheumatoid arthritis Treg cells." Eur J Immunol **44**(10): 2968-2978.

Kharchenko, P. V., M. Y. Tolstorukov and P. J. Park (2008). "Design and analysis of ChIP-seq experiments for DNA-binding proteins." Nat Biotechnol **26**(12): 1351-1359.

Koh, K. P., A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley and A. Rao (2011). "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells." Cell Stem Cell **8**(2): 200-213.

Landt, S. G., G. K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B. E. Bernstein, P. Bickel, J. B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K. I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A. J. Hartemink, M. M. Hoffman, V. R. Iyer, Y. L. Jung, S. Karmakar, M. Kellis, P. V. Kharchenko, Q. Li, T. Liu, X. S. Liu, L. Ma, A. Milosavljevic, R. M. Myers, P. J. Park, M. J. Pazin, M. D. Perry, D. Raha, T. E. Reddy, J. Rozowsky, N. Shoresh, A. Sidow, M. Slattery, J. A. Stamatoyannopoulos, M. Y. Tolstorukov, K. P. White, S. Xi, P. J. Farnham, J. D. Lieb, B. J. Wold and M. Snyder (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia." Genome Res **22**(9): 1813-1831.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

LaPlant, Q., V. Vialou, H. E. Covington, 3rd, D. Dumitriu, J. Feng, B. L. Warren, I. Maze, D. M. Dietz, E. L. Watts, S. D. Iniguez, J. W. Koo, E. Mouzon, W. Renthal, F. Hollis, H. Wang, M. A. Noonan, Y. Ren, A. J. Eisch, C. A. Bolanos, M. Kabbaj, G. Xiao, R. L. Neve, Y. L. Hurd, R. S. Oosting, G. Fan, J. H. Morrison and E. J. Nestler (2010). "Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens." Nat Neurosci **13**(9): 1137-1143.

Lasky, J. L. and H. Wu (2005). "Notch signaling, brain development, and human disease." Pediatr Res **57**(5 Pt 2): 104R-109R.

Li, B. and C. N. Dewey (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." BMC Bioinformatics **12**: 323.

Li, L., J. Carter, X. Gao, J. Whitehead and W. G. Tourtellotte (2005). "The neuroplasticity-associated arc gene is a direct transcriptional target of early growth response (Egr) transcription factors." Mol Cell Biol **25**(23): 10286-10300.

Li, T., D. Yang, J. Li, Y. Tang, J. Yang and W. Le (2014). "Critical Role of Tet3 in Neural Progenitor Cell Maintenance and Terminal Differentiation." Mol Neurobiol.

Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens and J. R. Ecker (2013). "Global epigenomic reconfiguration during mammalian brain development." Science **341**(6146): 1237905.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Luo, C., C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, B. Bui, T. J. Sejnowski, T. T. Harkins, E. A. Mukamel, M. M. Behrens and J. R. Ecker (2017). "Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex." Science **357**(6351): 600-604.

Mataga, N., S. Fujishima, B. G. Condie and T. K. Hensch (2001). "Experience-dependent plasticity of mouse visual cortex in the absence of the neuronal activity-dependent marker egr1/zif268." J Neurosci **21**(24): 9724-9732.

Mathelier, A., X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C. Y. Chen,

A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin and W. W. Wasserman (2014). "JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles." Nucleic Acids Res **42**(Database issue): D142-147.

Mora-Lopez, F., N. Pedreno-Horrillo, L. Delgado-Perez, J. A. Brieva and A. Campos-Caro (2008). "Transcription of PRDM1, the master regulator for plasma cell differentiation, depends on an SP1/SP3/EGR-1 GC-box." Eur J Immunol **38**(8): 2316-2324.

Murgatroyd, C., A. V. Patchev, Y. Wu, V. Micale, Y. Bockmuhl, D. Fischer, F. Holsboer, C. T. Wotjak, O. F. Almeida and D. Spengler (2009). "Dynamic DNA methylation programs persistent adverse effects of early-life stress." Nat Neurosci **12**(12): 1559-1566.

Nakagama, H., G. Heinrich, J. Pelletier and D. E. Housman (1995). "Sequence and structural requirements for high-affinity DNA binding by the WT1 gene product." Mol Cell Biol **15**(3): 1489-1498.

Piccolo, F. M., H. Bagci, K. E. Brown, D. Landeira, J. Soza-Ried, A. Feytout, D. Mooijman, P. Hajkova, H. G. Leitch, T. Tada, S. Kriaucionis, M. M. Dawlaty, R. Jaenisch, M. Merkenschlager and A. G. Fisher (2013). "Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion." Mol Cell **49**(6): 1023-1033.

Quinlan, A. R. (2014). "BEDTools: The Swiss-Army Tool for Genome Feature Analysis." Curr Protoc Bioinformatics **47**: 11 12 11-34.

Renaudineau, S., B. Poucet, S. Laroche, S. Davis and E. Save (2009). "Impaired long-term stability of CA1 place cell representation in mice lacking the transcription factor zif268/egr1." Proc Natl Acad Sci U S A **106**(28): 11771-11775.

Renbaum, P., R. Beeri, E. Gabai, M. Amiel, M. Gal, M. U. Ehrengruber and E. Levy-Lahad (2003). "Egr-1 upregulates the Alzheimer's disease presenilin-2 gene in neuronal cells." Gene **318**: 113-124.

Ritchie, M. F., C. Yue, Y. Zhou, P. J. Houghton and J. Soboloff (2010). "Wilms tumor suppressor 1 (WT1) and early growth response 1 (EGR1) are regulators of STIM1 expression." J Biol Chem **285**(14): 10591-10596.

Rudenko, A., M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, T. Le, K. F. Faull, R. Jaenisch and L. H. Tsai (2013). "Tet1 is critical for neuronal activity-regulated gene expression and memory extinction." Neuron **79**(6): 1109-1122.

Safford, M., S. Collins, M. A. Lutz, A. Allen, C. T. Huang, J. Kowalski, A. Blackford, M. R. Horton, C. Drake, R. H. Schwartz and J. D. Powell (2005). "Egr-2 and Egr-3 are negative regulators of T cell activation." Nat Immunol **6**(5): 472-480.

Seiler, M. P., R. Mathew, M. K. Liszewski, C. J. Spooner, K. Barr, F. Meng, H. Singh and A. Bendelac (2012). "Elevated and sustained expression of the transcription factors Egr1 and Egr2 controls NKT lineage differentiation in response to TCR signaling." Nat Immunol **13**(3): 264-271.

Shen, Y., F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov and B. Ren (2012). "A map of the cis-regulatory sequences in the mouse genome." Nature **488**(7409): 116-120.

Stroud, H., S. C. Su, S. Hrvatin, A. W. Greben, W. Renthal, L. D. Boxer, M. A. Nagy, D. R. Hochbaum, B. Kinde, H. W. Gabel and M. E. Greenberg (2017). "Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States." Cell **171**(5): 1151-1164 e1116.

Sun, M. A., Z. Sun, X. Wu, V. Rajaram, D. Keimig, J. Lim, H. Zhu and H. Xie (2016). "Mammalian Brain Development is Accompanied by a Dramatic Increase in Bipolar DNA Methylation." Sci Rep **6**: 32298.

Supek, F., M. Bosnjak, N. Skunca and T. Smuc (2011). "REVIGO summarizes and visualizes long lists of gene ontology terms." PLoS One **6**(7): e21800.

Theus, M. H., J. Ricard and D. J. Liebl (2012). "Reproducible expansion and characterization of mouse neural stem/progenitor cells in adherent cultures derived from the adult subventricular zone." Curr Protoc Stem Cell Biol **Chapter 2**: Unit 2D 8.

Veyrac, A., A. Besnard, J. Caboche, S. Davis and S. Laroche (2014). "The transcription factor Zif268/Egr1, brain plasticity, and memory." Prog Mol Biol Transl Sci **122**: 89-129.

Veyrac, A., A. Gros, E. Bruel-Jungerman, C. Rochefort, F. B. Kleine Borgmann, S. Jessberger and S. Laroche (2013). "Zif268/egr1 gene controls the selection, maturation and functional integration of adult hippocampal newborn neurons by learning." Proc Natl Acad Sci U S A **110**(17): 7062-7067.

Wang, Y., M. Xiao, X. Chen, L. Chen, Y. Xu, L. Lv, P. Wang, H. Yang, S. Ma, H. Lin, B. Jiao, R. Ren, D. Ye, K. L. Guan and Y. Xiong (2015). "WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation." Mol Cell **57**(4): 662-673.

Weaver, I. C., N. Cervoni, F. A. Champagne, A. C. D'Alessio, S. Sharma, J. R. Seckl, S. Dymov, M. Szyf and M. J. Meaney (2004). "Epigenetic programming by maternal behavior." Nat Neurosci **7**(8): 847-854.

Wei, F., Z. C. Xu, Z. Qu, J. Milbrandt and M. Zhuo (2000). "Role of EGR1 in hippocampal synaptic enhancement induced by tetanic stimulation and amputation." J Cell Biol **149**(7): 1325-1334.

Wu, X., M. A. Sun, H. Zhu and H. Xie (2015). "Nonparametric Bayesian clustering to detect bipolar methylated genomic loci." BMC Bioinformatics **16**(1): 11.

Yu, H., Y. Su, J. Shin, C. Zhong, J. U. Guo, Y. L. Weng, F. Gao, D. H. Geschwind, G. Coppola, G. L. Ming and H. Song (2015). "Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair." Nat Neurosci **18**(6): 836-843.

Zandarashvili, L., M. A. White, A. Esadze and J. Iwahara (2015). "Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1." FEBS Lett **589**(15): 1748-1753.

Zhang, R. R., Q. Y. Cui, K. Murai, Y. C. Lim, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H. P. Wu, W. Liu, Z. M. Xu, L. Yang, Y. Q. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi and G. L. Xu (2013). "Tet1 regulates adult hippocampal neurogenesis and cognition." Cell Stem Cell **13**(2): 237-245.

Zhu, X., D. Girardo, E. E. Govek, K. John, M. Mellen, P. Tamayo, J. P. Mesirov and M. E. Hatten (2016). "Role of Tet1/3 Genes and Chromatin Remodeling Genes in Cerebellar Circuit Formation." Neuron **89**(1): 100-112.

Chapter 4 – Conclusion

DNA methylation is a well-studied epigenetic mechanism that occurs by addition of methyl (-CH3) group to the fifth position of cytosine, playing pivotal roles in diverse biological processes, including regulation of transcription, imprinting, X-chromosome inactivation, and silencing of transposons. The deposition and removal of DNA methylation are finely tuned by DNMTs and TET proteins, and its aberration can cause devastating consequences and contributes to many diseases, including cancers, neurological disorders (e.g., autism spectrum disorder (ASD), Rett Syndrome).

In mammals, DNA methylation acts as an essential epigenetic mechanism with critical roles in brain development, neural differentiation and identity maintenance, synaptic plasticity, learning and memory. To gain a better understanding of epigenetic mechanisms underlying brain function at the molecular and cellular level, accumulating studies start to focus on the interrogation of DNA methylation in mammalian brains on a genome-wide scale, or at single-base resolution and even at single-cell level.

Mammalian brains consist of millions to billions of neurons and glia, which can be further categorized into many distinct types of cells with assumedly distinct DNA methylation profiles. Therefore, some genomic loci from cell population may demonstrate bipolar DNA methylation pattern, i.e., hypermethylated in one cell subset but hypomethylated in others. Although the technological advances, including the most recent technique single-nucleus methylome sequencing (snmC-seq), greatly expedite the exploration of epigenomic diversity, how extensive methylation patterns vary among brain cells is still largely limited. Moreover, our understanding of epigenetic regulatory mechanisms underlying the establishment and maintenance of cell type-specific methylation is quite limited.

To investigate these two questions, in Chapter 2, we first inferred CSM loci among human and mouse brain cells at different development stages of frontal cortices. By utilizing the genome-scale hairpin bisulfite sequencing technique, we observed high methylation fidelity ($\geq$94.9%) in human fetal and adolescent brain methylomes. This high methylation fidelity in the brain demonstrated that the majority of putative CSM loci likely resulted from the methylation differences among brain cells rather than from asymmetric DNA methylation between DNA

double strands. Strikingly, the frequency of putative CSM loci increased dramatically during early postnatal stages of brain development, suggesting that dramatic changes in functions and complexities of the brain may be accompanied by a rapid change in epigenetic heterogeneity. Additionally, we examined the occurrence of putative CSM loci in regulatory elements including promoters and enhancers, and found these regions enriched in non-CGI promoters and active/poised enhancers. Furthermore, these CSM regions show significant enrichment in brain disease/trait-associated SNPs. These properties suggest that putative CSM loci play important roles in brain development, and are probably involved in the pathogenesis of brain diseases, including Parkinson's disease and Schizophrenia.

To further explore epigenetic regulatory mechanisms during brain development, in Chapter 3, we utilized unbiased data-driven approaches to re-analyze methylome for human and mouse frontal cortices at different developmental stages. We predicted that *Egr1*, a transcriptional factor with important roles in neuron maturation, synaptic plasticity, long-term memory formation and learning, plays an essential role in brain epigenetic programming. After performing EGR1 ChIP-seq, we found that thousands of EGR1 binding sites are with cell-type specific methylation patterns established during postnatal frontal cortex development. More specifically, the CpG dinucleotides within these EGR1 binding sites become hypomethylated in mature neurons but remain heavily methylated in glia. We further demonstrated that EGR1 recruits a DNA demethylase TET1 to remove the methylation marks at EGR1 binding sites and activate downstream genes. Also, we found that the frontal cortices from the knockout mice lacking *Egr1* or *Tet1* share strikingly similar profiles in both gene expression and DNA methylation. Collectively, our study reveals that EGR1 programs brain methylome together with TET1 during postnatal development and provides new insights into how life experience and neuronal activity may shape the brain methylome.

Brain development is a dynamic process that involves the interaction between genes and environmental exposures. Although the process continues into adulthood, early stage of postnatal brain development represents a critical period for brain structure and function maturation (Hubel, Wiesel et al. 1977). During this period, environmental inputs can lead to divergent developmental trajectories, leaving long-term effects on later life behaviors (Tierney and Nelson 2009, Stiles and Jernigan 2010). However, we still have a limited understanding of the mechanisms by which early-life experience, such as stress and childhood neglect/abuse, can

translate into long-term gene expression aberration and neuronal dysfunction, predisposing the individual vulnerability to mental disorders or neurodegenerative diseases later in life (Kundakovic and Champagne 2015). Mounting evidence reveals that epigenetic factors (i.e., DNA methylation, histone modification and small non-coding RNA) can function as a molecular bridge between genes and environments (Gapp, Jawaid et al. 2014, Champagne 2018). The aberrant DNA methylation patterns caused by early-life adversity can lead to dysregulation of several essential genes, posing a risk of cognitive impairments or psychiatric disorders to individuals (McGowan, Sasaki et al. 2009).

In this dissertation, we reveal a mechanism of brain methylome programming mediated by an activity-dependent transcription factor, EGR1, together with TET1. As a member of the immediate early gene family that provides a molecular framework for rapid and dynamic response to neuronal activity and environmental stimuli, *Egr1* has been shown to regulate synaptic plasticity, control newborn neuron selection, maturation and experience-dependent functional integration (Veyrac, Gros et al. 2013). Mice with global *Egr1* deletion exhibit impaired learning and long-term memory, while conditional transgenic mice with *Egr1* overexpressed in forebrain neurons display strengthened memory with enhanced resistance to extinction (Baumgartel, Genoux et al. 2008). Studies in humans indicate that EGR1 expression is positively correlated with expression levels of marker genes of neuropsychiatric disorders, and down-regulated in the prefrontal cortex of patients with psychiatric disorders (i.e., depression, schizophrenia), highlighting its role in human cognitive function (Duclot and Kabbaj 2017). Although animal and human studies demonstrate the role of EGR1 in brain development and its implication in psychiatric disorders, the underlying mechanisms at the molecular and cellular levels remain largely unknown.

Our work provides an initial glimpse into EGR1-mediated epigenetic programming in mouse brain during early postnatal development. To gain a better understanding of how EGR1-mediated epigenetic programming translates early life experience into long-lasting effects on behaviors in mammals, further research should be conducted to explore the roles of EGR1 binding cis-regulatory elements and associated genes in cognitive processes including long-term memory and learning. It is worth noting that both *Egr1* and *Tet1* express universally among brain cells, whereas the global deletion of either gene exhibit cell-type specific methylation changes. We found that the identified DMRs in Egr1KO or Tet1KO show low methylation in excitatory

neurons in relative to PV and VIP neurons, in addition, EGR1 binding sites are significantly enriched on excitatory-neuron-specific hypomethylated regions but excluded from inhibitory-neuron-specific ones, this suggests different types of neural cells may adopt distinct epigenetic networks to program their methylome during brain development. Thus, future work on methylome and transcription analyses of isolated neural cell populations can shed insights into this question.

Early-life adversity can elicit aberrant methylation in different brain cells, in this study, we identified hundreds to thousands of DMRs and thousands of putative CSM loci which enriched in brain disease/trait-associated SNPs. These data could be used to screen biomarkers for the early detection of brain diseases, or employed to select genomic loci for epigenetic editing to correct aberrant methylation at specific sites in a cell type-specific manner.

REFERENCES

Baumgartel, K., D. Genoux, H. Welzl, R. Y. Tweedie-Cullen, K. Koshibu, M. Livingstone-Zatchej, C. Mamie and I. M. Mansuy (2008). "Control of the establishment of aversive memory by calcineurin and Zif268." Nat Neurosci **11**(5): 572-578.

Champagne, F. A. (2018). "Beyond the maternal epigenetic legacy." Nat Neurosci **21**(6): 773-774.

Duclot, F. and M. Kabbaj (2017). "The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders." Front Behav Neurosci **11**: 35.

Gapp, K., A. Jawaid, P. Sarkies, J. Bohacek, P. Pelczar, J. Prados, L. Farinelli, E. Miska and I. M. Mansuy (2014). "Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice." Nat Neurosci **17**(5): 667-669.

Hubel, D. H., T. N. Wiesel and S. LeVay (1977). "Plasticity of ocular dominance columns in monkey striate cortex." Philos Trans R Soc Lond B Biol Sci **278**(961): 377-409.

Kundakovic, M. and F. A. Champagne (2015). "Early-life experience, epigenetics, and the developing brain." Neuropsychopharmacology **40**(1): 141-153.

McGowan, P. O., A. Sasaki, A. C. D'Alessio, S. Dymov, B. Labonte, M. Szyf, G. Turecki and M. J. Meaney (2009). "Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse." Nat Neurosci **12**(3): 342-348.

Stiles, J. and T. L. Jernigan (2010). "The basics of brain development." Neuropsychol Rev **20**(4): 327-348.

Tierney, A. L. and C. A. Nelson, 3rd (2009). "Brain Development and the Role of Experience in the Early Years." Zero Three **30**(2): 9-13.

Veyrac, A., A. Gros, E. Bruel-Jungerman, C. Rochefort, F. B. Kleine Borgmann, S. Jessberger and S. Laroche (2013). "Zif268/egr1 gene controls the selection, maturation and functional integration of adult hippocampal newborn neurons by learning." Proc Natl Acad Sci U S A **110**(17): 7062-7067.