Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex

Iris Cheung^{a,1}, Hennady P. Shulha^{b,1}, Yan Jiang^a, Anouch Matevossian^a, Jie Wang^c, Zhiping Weng^{b,c,d,2}, and Schahram Akbarian^{a,2}

^aBrudnick Neuropsychiatric Research Institute, Department of Psychiatry, ^cProgram in Bioinformatics and Integrative Biology, and ^dDepartment of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01604; and ^bBiomedical Engineering, Boston University, Boston MA 02215

Edited* by Edward G. Jones, University of California, Davis, CA, and approved April 8, 2010 (received for review February 10, 2010)

Little is known about the regulation of neuronal and other cell-type specific epigenomes from the brain. Here, we map the genome-wide distribution of trimethylated histone H3K4 (H3K4me3), a mark associated with transcriptional regulation, in neuronal and non-neuronal nuclei collected from prefrontal cortex (PFC) of 11 individuals ranging in age from 0.5 to 69 years. Massively parallel sequencing identified 12,732–19,704 H3K4me3 enriched regions (peaks), the majority located proximal to (within 2 kb of) the transcription start site (TSS) of annotated genes. These included peaks shared by neurons in comparison with three control (lymphocyte) cell types, as well as peaks specific to individual subjects. We identified 6,213 genes that show highly enriched H3K4me3 in neurons versus control. At least 1,370 loci, including annotated genes and novel transcripts, were selectively tagged with H3K4me3 in neuronal but not in nonneuronal PFC chromatin. Our results reveal age-correlated neuronal epigenome reorganization, including decreased H3K4me3 at approximately 600 genes (many function in developmental processes) during the first year after birth. In comparison, the epigenome of aging (>60 years) PFC neurons showed less extensive changes, including increased H3K4me3 at 100 genes. These findings demonstrate that H3K4me3 in human PFC is highly regulated in a cell type- and subject-specific manner and highlight the importance of early childhood for developmentally regulated chromatin remodeling in prefrontal neurons.

cerebral cortex | neuron | histone methylation | human | brain | gene expression

Developmentally regulated changes in histone modifications and DNA methylation, shaping gene expression patterns and genome organization, are critical intermediates for numerous genetic and environmental factors affecting neuronal functions in healthy and diseased brains (1). For example, there is increasing evidence that epigenetic alterations in the cerebral cortex and hippocampus play an important role in the etiology of schizophrenia and other neurodevelopmental disease (2, 3). Cortical neurons permanently exit from the cell cycle during the fetal period, before the dramatic changes in functional connectivity, both on a micro- (e.g., synapse) and macroscale (e.g., network activity, cortical gray matter volumes), that extend into early childhood years and continue throughout adolescence and even beyond (4, 5). To date, however, comprehensive and genome-wide maps of neuronal epigenomes, and their developmental trajectories, do not exist. This critical deficiency in epigenetic information, as it pertains to the human—and more generally animal—brain, finally can be addressed because recently it became possible to efficiently separate neuronal chromatin from other chromatin in tissue, thereby avoiding potential confounds such as the highly dynamic changes in glia cell densities during cortical ontogenesis and maturation (6). Here, we employ ChIP-Seq (7) to study the genome-wide distribution of histone H3K4 trimethylation (H3K4me3)—an epigenetic mark highly enriched at start sites

of actual or potential transcription (8)—in neuronal nuclei from the human prefrontal cortex. We present neuronal epigenomes from human brains, highlight commonalities and differences between individual subjects, and identify neuron-specific chromatin signatures and novel transcripts. Our results also indicate that large scale neuronal chromatin remodeling occurs in early childhood. Exploration of cell type-specific epigenomes will provide insights into mechanisms of normal and diseased neurodevelopment, thereby potentially transforming human brain research.

Results

H3K4me3 Landscapes of PFC Neurons. We isolated nuclei from the poles of the frontal lobes of 11 subjects (age range 0.5–69 years, median = 4.71 years; *Materials and Methods*) and sorted for the NeuN positive (NeuN+) population by FACS (Fig. 1A). NeuN is a marker specific for mature neurons and expressed by a large majority of glutamatergic (excitatory) and GABAergic (inhibitory) neurons (9, 10). We generated genome-wide maps of the H3K4me3 mark from chromatin prepared by micrococcal nuclease digest, using ChIP-Seq in conjunction with an Illumina Solexa sequencing platform (*Materials and Methods*). For each neuronal (NeuN+) sample, 2.1–8.3 million 36-nt sequence tags were obtained, 67–87% of which mapped to unique locations in the human genome (Table S1). For 2 of the 11 subjects, the NeuN– samples, which are primarily comprised of nonneuronal cells (including glia, microglia, and endothelium), were also processed by ChIP-Seq.

To obtain a genome-wide assessment of the 13 datasets (11 NeuN+ and 2 NeuN–), we computed Pearson correlation coefficients (r) between all pairs of datasets based on tag counts in 20,324 annotated promoters [the RefSeq track of the University of California Santa Cruz (UCSC) genome browser]. We also included three publicly available H3K4me3 ChIP-Seq datasets as control: CD4⁺ T lymphocytes (11), and the K562 [chronic myelogenous leukemia (CML)] and GM12878 (lymphoblastoid) cell lines (12). The 11 prefrontal NeuN+ samples are highly correlated (r in the range of 0.92–0.98), substantially higher than their correlations with the 2 NeuN– datasets ($r = 0.68–0.78$), and further higher than the 3 lymphocyte-derived datasets ($r = 0.46–0.60$) (Fig. 2 Lower). Furthermore, the highest

Author contributions: I.C., Z.W., and S.A. designed research; I.C., Y.J., and A.M. performed research; H.P.S., J.W., and Z.W. analyzed data; and I.C., H.P.S., Z.W., and S.A. wrote the paper. The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the National Center for Biotechnology Information, (accession no. GSE21172).

¹I.C. and H.P.S. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: schahram.akbarian@umassmed.edu or zhiping.weng@umassmed.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001702107/-DCSupplemental.

Fig. 1. Epigenome profiling in PFC neurons. (A) Representative examples of fluorescence-activated sorting of NeuN+ tagged neuronal nuclei. (Scale bar, 10 μ m.) (B and C) H3K4me3 profiles for the 11 PFC NeuN+ neuronal samples (red), and for comparison, NeuN- (orange) and CD4+ lymphocytes (blue) in a (B) 20-kb region of Chr 17 and (C) 1.3-MB region of Chr 8.

correlations were found between NeuN+ samples of similar ages (Fig. 2 Upper). No gender-specific pattern is apparent from the correlation analysis. Five male-specific H3K4me3 peaks are near annotated genes on the Y-chromosome (Fig. S1A and B), one of which is associated with *NLGN4Y*, a cell adhesion gene highly expressed in the PFC of young males (13) that could play a role in some cases of autism (14). Some of the remaining genes, including *ZFY*, were also mentioned in an earlier work on gender-related transcriptome differences in the human PFC (15).

From the above genome-wide correlation analysis on known promoters we draw two conclusions. First, the H3K4me3 epigenome is cell type-specific: it is differentially regulated in neurons in comparison with nonneuronal cells in the prefrontal cortex of an individual. Second, the H3K4me3 epigenome undergoes age-correlated reorganization in prefrontal neurons.

There were 12,732–19,704 statistically enriched regions for the 11 neuronal samples compared with the 3 lymphocyte samples (Materials and Methods) (Table S2). We will refer to these regions as neuron-enriched H3K4me3 peaks (or simply peaks when there is no ambiguity) throughout this paper. The peaks average approximately 1.3 kb, and 69–81% of them were proximal to (within 2 kb of) the transcription start site (TSS) of a RefSeq gene (Table S2). When normalized for sequencing depth, the percentages of proximal peaks became even higher (77–84%) (Table S2). Postmortem confounds—tissue pH and postmortem interval—do not show significant correlations with the percentage of gene proximal peaks ($r = 0.50$ and 0.58 ; P value = 0.06 and 0.12), in agreement with a previous study based on individual genes (16). Fig. 1B shows the H3K4me3 peak at the TSS of *PPP1R1B* (commonly known as *DARPP32*), a well studied target gene of dopaminergic pathways (17), and Fig. 1C shows peaks at multiple TSSs of neuregulin 1 (*NRG1*), a prominent schizophrenia susceptibility gene (18).

Neuron-Specific Signatures Within the H3K4me3 Epigenome. The neuron-enriched H3K4me3 peaks among the 11 neuronal samples overlapped greatly, with 90–99.9% of the peaks in one sample overlapping with the peaks in at least one other sample (Table S2). We identified 7,949 peaks that are enriched in all 11 neuronal samples compared with 3 lymphocyte samples (see Materials and Methods), and 6,394 (80%) of these peaks are proximal to 6,213 RefSeq genes. Strikingly, the genes near these 6,394 peaks are enriched in 179 tightly connected gene ontology (GO) categories (false discovery rate (FDR) cutoff 0.0001), all of which are related to neuronal signaling (Table S3). Under an even stricter cutoff (see Materials and Methods) there are still 1,721 neuron-enriched peaks, proximal to 1,682 genes. The neuron-related GO enrichment became even more significant for these 1,682 genes (the most significant GO biological processes are shown in Fig. 3 and the entire list is in Table S3). Among the most specific categories in biological process is axon guidance (26/68 genes, FDR = 7.6×10^{-11}), in cellular compartment is synaptic vesicle (24/60 genes, FDR = 6.0×10^{-10}), and in molecular function are voltage-gated potassium (34/108 genes) or calcium (14/30 genes) or extracellular glutamate-gated channel activities (15/20 genes) (FDR = 8.4×10^{-12} , 9.4×10^{-7} , and 4.6×10^{-11}). Fig. 3 shows that there are two tightly connected components in the enriched biological processes, one centered on neuron development and the other on synaptic transmission. Among the genes with neuron-specific H3K4me3 peaks are many key regulators of prefrontal function and cognition, including multiple NMDA and AMPA receptor subunits (*GRIN1/2A/2B*, *GRIA2*) and neuron-specific nitric oxide synthase (*NOS1*) (19). To investigate why these neuron-specific H3K4me3 peaks are not present in lymphocytes, we aggregated two repressive histone marks, H3K27me3 and H3K9me3, around these peaks in CD4+ T cells and found H3K27me3 to be substantially elevated compared with the average signals around the H4K4me3 peaks in T cells (Fig. S1C and D). We conclude that there are thousands of loci that are significantly enriched in H3K4me3 in the neurons of all 11 examined individuals compared with the 3 lymphocyte samples. These loci are defined by high levels of (open chromatin-associated) H3K4me3 in PFC neurons, and, conversely, high levels of (repressive chromatin-associated) H3K27me3 in lymphocytes.

We also identified H3K4me3 peaks in each of the two NeuN- samples compared with the three lymphocyte samples (Table S2). Among the aforementioned 7,949 peaks that are enriched in all 11 neuronal samples, 3,531 peaks overlap with peaks in both NeuN- samples (among them 2,834 peaks are proximal to 2,777 genes) and 1,370 peaks do not overlap with any peaks in either of the NeuN- samples (among them 909 peaks are proximal to 893 genes). The 2,777 genes that are marked by H3K4me3 in both neuronal (NeuN+) and nonneuronal (NeuN-) samples are enriched in 103 highly brain-specific GO categories, with the most significant biological process being nervous system development (Table S4). The 893 genes that are marked by H3K4me3 in all neuronal samples, but not in either of the NeuN- samples, are also enriched in highly neuron-specific GO categories, with the most significant GO biological process being synaptic transmission (Table S5). Note that nervous system development and synaptic transmission fall into the two tightly connected components of

Fig. 2. The H3K4me3 epigenome of PFC neurons is different from nonneuronal cells. Heatmaps showing Pearson correlation coefficients for pair-wise comparison of raw tag counts within 2 kb of annotated TSS. (Lower) 11 neuronal (NeuN+) samples from subjects 1–11 (female pink, male blue), two nonneuronal (NeuN–) samples from subjects 6 and 11, and CD4⁺ T cells and two lymphocyte-derived cell lines K562 and GM12878. (Upper) The same comparison for the 11 neuronal samples as in *Bottom*, but with a different color scale to highlight differences between neuronal samples. Notice (Lower) very high levels of correlation between NeuN+ samples, but not between NeuN+ and NeuN– samples or lymphocytes. Notice further (Upper) that correlations are higher for neuronal samples closer in age.

Fig. 3, suggesting that genes that are marked by H3K4me3 in NeuN– cells are also important for the development of neurons but less important for transmitting nerve impulses. Many of the 893 NeuN+ cell specific genes are key regulators for prefrontal cognition and function, including the aforementioned *NOS1* and *NRG1*, for which multiple transcription start sites are marked by sharp H3K4me3 peaks in all 11 NeuN+ samples but none of the 2 NeuN– or 3 lymphocyte samples (Fig. 1C).

Neuron-Specific H3K4me3 Localizes to Hundreds of Nonannotated Transcripts. Among the 1,370 peaks that are shared by all NeuN+ neuronal samples but not present in lymphocytes or the NeuN– samples, 461 are more than 2 kb from the 5′-ends of any RefSeq genes. Most of these peaks (337/461, or 73%) overlap with annotated CpG islands (Table S6). This level of overlap with CpG islands is only slightly lower than the neuron-specific peaks that are within 2 kb of RefSeq genes (91%), indicating that a high percentage of the 461 peaks correspond to bona fide TSSs. Indeed, 404 of the 461 (88%) peaks were positioned within 2 kb of expressed sequence tags (ESTs) (Table S6). We experimentally tested a subset of the peaks that were positioned more than 10 kb from annotated genes for RNA expression in PFC neurons by qRT-PCR and in situ hybridization. One transcript, located on Chr 9 (Fig. 4A), which appears to be primate-specific, shows particularly high level of expression in PFC pyramidal neurons positioned in cortical layer III (Fig. 4B). We conclude that H3K4me3 mapping can serve as a guide to uncover potentially hundreds of unannotated novel and cell-specific transcripts in the brain.

H3K4me3 Peaks Distinct in Individual Epigenomes. Each neuronal sample contains peaks that are specific to that individual. Among these subject-specific peaks, 64–91% overlap with ESTs. In addition, 5–66% overlap with CpG islands (Table S7). In the two individuals for which we obtained both NeuN+ and NeuN– data, 4/81 (5%) and 301/517 (58%) of subject-specific peaks in the NeuN+ fraction are also present in the NeuN– fraction of the same individuals. Taken together, it is unlikely that these subject-specific peaks are artifacts.

We analyzed the distribution of the subject-specific peaks and found that only a smaller fraction of them are located within 2 kb of an annotated TSS (Fig. 5B), in striking contrast to the large portion (80%) of gene-proximal peaks common to all 11 individuals (Fig. 5A). Conversely, the proportion of subject-specific peaks positioned 2–10 kb from annotated TSSs is 10–25%, and the proportion

positioned more than 10 kb from annotated TSSs is 26–69%, or 2- to 4-fold higher than the proportions of TSS distal peaks common to all samples (Fig. 5B). Thus, some of these subject-specific peaks may correspond to novel noncoding transcripts.

Remodeling of PFC Neuron Chromatin During Postnatal Development. Pearson correlation analyses for H3K4me3 marking within 2 kb of annotated TSS revealed that PFC NeuN+ samples show age-related correlation (Fig. 2 Upper). To further explore this age-related remodeling of the H3K4me3 epigenome, we compared the three youngest samples (all less than 1 year of age) with those from the three oldest samples (all more than 60-years-old). Indeed, 589 gene-proximal peaks (corresponding to 556 genes) are marked in infant neurons but not in any of the oldest samples. For example, genes specific to infant brains include *NEUROD1* and other transcription factors implicated in neurogenesis, multiple members of the semaphorin and cadherin families of adhesion molecules (*SEMA3A/3F/6A*, *CDH3/9*), and many others that are of fundamental importance for neuronal growth and differentiation and connectivity formation in cerebral cortex and other brain regions (Table S8). In contrast, fewer H3K4me3 peaks are specific to the three oldest (more than 60-years-old) samples compared with infants; 101 genes are selectively marked by H3K4me3 in the aged samples, but no significant GO enrichment was observed. Our results indicate significant remodeling of H3K4 methylation during postnatal development and aging of prefrontal neurons, with the number of de(H3K4 tri)methylated genes outnumbering genes with increased H3K4me3 methylation in a 6:1 ratio.

Taken together, our observations point to significant epigenome reorganization during the transition from the postnatal period and early infancy to later phases of childhood. Although the epigenomes of PFC neurons are highly similar between all individuals we examined (Fig. 2 Lower), greater similarity was observed for individuals closer in age (Fig. 2 Upper).

Discussion

In this study, we mapped genome-wide H3K4me3 in the prefrontal neurons of 11 humans. We observed sharp peaks, most of which are proximal to annotated TSSs. The majority of the remaining peaks overlap annotated CpG islands and are near ESTs, suggesting that they correspond to novel TSSs. Thousands of genes are associated with peaks common to all 11 samples, but not present in CD4⁺ lymphocytes and two lymphocyte-derived cell lines, and several

Fig. 4. H3K4me3 map as guide to find novel transcripts expressed in prefrontal neurons. (A) 4-kb region of Chr 9 showing, proximal to ESTs, a representative example (subject no. 11) of a peak present in all NeuN+ samples but not present in lymphocytes or NeuN– samples. y axis, number of reads normalized by sequencing depth. Purple tag marks sequence used for in situ hybridization of sections from adult PFC specimen shown in (B Left to Right) Nissl, antisense, and sense probe. Notice intense labeling of layer III pyramidal neurons. (Scale bar, 100 μm.)

reads per sample), it should be feasible to explore the epigenetic underpinnings of individual differences in cognitive abilities, personality traits, or maladaptive behaviors (2) by ChIP-Seq screening of chromatin marks in larger cohorts. Finally, the approach present here can be easily extended to map neuronal epigenomes across different anatomical areas or even for specific neuronal subpopulations. A better understanding of region- and cell-specific epigenomes is likely to shed new light on the developmental and evolutionary history of the human and primate brain.

Materials and Methods

Postmortem Brain Samples. Human prefrontal cortex samples used in this study were obtained from the Brain and Tissue Bank for Developmental Disorders, University of Maryland (Director: R. Zielke) and a brain bank at the University of California, Irvine (Director: W. E. Bunney, Jr.). Procedures were approved by the Institutional Review Board of the University of Massachusetts Medical School. See Table S1 for detailed and case-by-case information on demographics and postmortem parameters.

Fig. 5. Numbers and distribution of H3K4me3 peaks in PFC neurons. (A) Distribution of 7,949 peaks common to all 11 subjects in relation to annotated TSSs (<2 kb, 2–10 kb, > 10 kb as indicated) (B) Same as in A but for subject specific peaks. Numbers are expressed as mean ± SEM with $n = 11$. Notice different distribution patterns of “common,” in A, and “subject specific” peaks, in B.

Nuclei Extraction and FACS. Nuclei extraction was carried out essentially as described in Jiang et al. (6). 250 mg of tissue was homogenized in 5 mL of lysis buffer (0.32 M sucrose, 10 mM Tris pH 8.0, 5 mM CaCl_2 , 3 mM Mg acetate, 1 mM DTT, 0.1 mM EDTA, 0.1% Triton X-100) by douncing 50 times in a 7-mL dounce homogenizer. Lysate is transferred to a 14 mL ultracentrifugation tube (Beckman; 14 × 95 mm; 344061), and 9 mL of sucrose solution (1.8 M sucrose, 10 mM Tris pH 8.0, 3 mM Mg acetate, 1 mM DTT) is pipetted directly to the bottom of the tube. Four tubes were prepared for each sample so in total 1 g of tissue was used. Ultracentrifugation was carried out at 24,400 rpm for 2.5 h at 4 °C (Beckman; L8-70 M; SW80 rotor). After centrifugation, the two layers of solutions were removed by aspiration. Each nuclei pellet was resuspended in 500 μL PBS and all four were pooled together.

Nuclei were incubated in staining mix [0.72% normal goat serum, 0.036% BSA, 1:1200 anti-NeuN (Millipore), 1:1400 Alexa488 goat anti-mouse secondary antibody (Invitrogen)] for at least 45 min by rotating in dark at 4 °C. A small fraction of the nuclei was used in stained control where anti-NeuN was omitted. FACS was done at the University of Massachusetts Medical School Flow Cytometry Core Lab on a FACS Vantage SE flow cytometer. To pellet the sorted nuclei, 2 mL of sucrose solution, 50 μL of 1 M CaCl_2 and 30 μL of Mg acetate were added to 10 mL of nuclei in PBS, incubated on ice for 15 min, then centrifuged at 3,000 rpm for 15 min. The nuclei pellet was resuspended in a buffer containing 10 mM Tris (pH 7.5), 4 mM MgCl_2 , and 1 mM CaCl_2 . Fluorescence images of nuclei were acquired by Slidebook 4 (Intelligent Imaging Innovations) on a Axiovert 200 M (Carl Zeiss).

Chromatin-Immunoprecipitation. Prewarmed nuclei in 400 μL were digested with micrococcal nuclease (4 U/mL) at 37 °C for 5 min. Reaction was stopped and nuclei were lysed by adding 40 μL of 0.5 M EDTA and kept on ice; 900 μL of 0.2 mM EDTA with 0.1 mM Benzamidin, 0.1 mM PMSF, and 1 mM DTT was added to lysed nuclei and precleared by Protein G Agarose (Millipore) for 0.5 h at 4 °C. After spinning down beads, 1 mL of supernatant was transferred to a new tube, and chromatin immunoprecipitation was carried out by addition of 50 mM NaCl, 20 mM Tris pH 7.5, 5 mM EDTA, and anti-H3K4me3 (1:315; Upstate; 07-473) at 4 °C overnight. Immunoprecipitated chromatin was recovered by rotating with Protein G Agarose for 1 h, and beads were washed for 3 min each at RT in low salt buffer (150 mM NaCl, 10 mM Tris pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), followed by high salt buffer (500 mM NaCl, 20 mM Tris pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), followed by lithium chloride buffer (250 mM LiCl, 10 mM Tris pH 8.0, 1 mM EDTA, 1% Deoxycholic acid, 1% IGEPAL), followed by TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). Bound chromatin was eluted in 0.1 M NaHCO_3 and 1% SDS twice, first by rotating for 15 min, then by vortexing 15 min at room temperature (RT); 40 mM Tris pH 6.5, 10 mM EDTA, and 50 μg/mL of Proteinase K were added to the eluted chromatin and incubated at 52 °C for 3–4 h. DNA was purified by phenol-chloroform extraction, and precipitated by ethanol overnight. DNA pellet was washed twice with 70% ethanol and resuspended in 4 mM Tris pH 8.0.

ChIP-Seq Library Construction. ChIPed DNA was first blunt-ended by End-it DNA Repair Kit (Epicentre) at RT for 1 hr. Reaction was cleaned by column (Qiagen), and DNA was A-tailed by exo^- Klenow DNA Polymerase (Epicentre) at RT for 1 hr. Reaction was again cleaned by column, and DNA was ligated to Genomic Adaptor Oligo Mix (Illumina) by Fast-link DNA Ligation Kit (Epicentre) at RT for 2 h followed by 16 °C O/N. Reaction was cleaned by column, and ligated DNA was amplified using PfuUltra II Fusion HS DNA Polymerase (Stratagene) with 250 μM dNTPs and 1 μL of each of the Genomic PCR Primers (Illumina) in 100 μL volume with the following cycling conditions: 95 °C 1 min; 15 cycles of 95 °C 50 sec, 65 °C 1 min, 72 °C 30 sec; 72 °C 5 min. PCR product was cleaned and volume reduced by column before loading onto a 1.8% agarose gel. Typically a strong smear of 160–230 bp (including 92 bp from the Illumina primers) was observed. Occasionally another smear of less intensity at around 400 bp, likely representing di-nucleosomal DNA, was also observed. The smaller smear was cut out and gel purified. Libraries were sequenced by an Illumina Genome Analyzer (GA II). Images were analyzed using GAPIipeline (versions 1.0 and 1.4).

Data Analysis. Length of sequence reads was 36 bp. We used Bowtie (version 0.11.3) allowing up to one mismatch (29) to map all sequence reads to the gender appropriate human genome HG18, and 67–87% of the reads in the neuronal samples mapped to one unique location in the genome. Reads that were mapped to multiple locations in the genome were discarded for subsequent data analysis. Summary statistics on all 11 NeuN+ and the 2 NeuN– samples are presented in Table S1. Additional H3K4me3 ChIP-Seq data were obtained for CD4+ T cells (11), and for two cell lines (GM12878 and K562) from the UCSC ENCODE webpage (12).

We analyzed genome-matching reads with the MACS software ((30) (version 1.3.5) with $\text{bw} = 230$ bp, as defined experimentally by PCR, $t\text{Size} = 36$ bp, and

other parameters set at default. MACS reports genomic regions that are statistically enriched in H3K4me3; these regions are called H3K4me3 peaks or simply peaks when there is no ambiguity. We detected peaks that were enriched in each of the 11 NeuN+ and 2 NeuN- samples against the three lymphocyte samples pooled using MACS, by requiring MACS P value $< 1e^{-5}$ and normalized tag density in the neuronal samples > 0.005 ppm (normalized for sequencing depth).

To calculate the Pearson correlation heatmap (Fig. 2), each distinct TSS was accounted for separately and the region within 2 kb of a TSS was defined as the promoter of the TSS. The TSS set was defined with the RefSeq gene set from UCSC genome browser, which contained 21,432 genes, 32,914 transcripts and 24,791 unique TSSs. Promoters for TSSs that were less than 2 kb apart were merged to avoid double counting. Number of tags within each promoter was tallied and divided by size of the regions and the resulting tag densities for all annotated TSSs were used to compute Pearson correlation coefficients between each pair of samples.

To detect peaks that were enriched in H3K4me3 in all 11 neuronal samples compared with the 3 lymphocytes, we used the following criteria: (i) P value $< 1e^{-10}$ computed by the MACS algorithm run with all 11 neuronal samples pooled against 3 lymphocyte samples pooled; (ii) log-ratio (base 2) of average tag density between pooled neuronal and lymphocyte samples > 1 ; (iii) normalized tag density in the pooled neuronal sample > 0.005 ppm; (iv) P value for t test on tag density between 11 neuronal samples against 3 lymphocyte samples $< 2e^{-2}$; and (v) the peaks defined with the pooled neuronal sample against the pooled lymphocyte sample overlap with peaks in each of the 11 neuronal samples against the pooled lymphocyte sample (MACS P value $< 1e^{-5}$). These criteria resulted in 7,949 peaks and 6,394 of them are proximal to TSS. The list of these peaks is provided in Table S3.

We also used a stricter set of criteria for detecting peaks that are enriched in H3K4me3 in all 11 neuronal samples compared with the 3 lymphocytes: (i) MACS P value $< 1e^{-100}$ run with all 11 neuronal samples pooled against 3 lymphocyte samples pooled; (ii) log-ratio (base 2) of average tag density between pooled neuronal and lymphocyte samples > 2 ; (iii) same as above; (iv) P value for t test on tag density between 11 neuronal samples against 3 lymphocyte samples $< 1e^{-4}$; and (v) same as above. These criteria resulted in 2,148 peaks and 1,721 of them are proximal to TSS.

We used the following criteria to detect peaks that were enriched in H3K4me3 in the three infant neuronal samples compared with the three oldest neuronal samples: (i) MACS P value $< 1e^{-10}$ run with the three infant samples pooled against the three oldest samples pooled; (ii) log-ratio (base 2) of average tag density between pooled neuronal and lymphocyte samples > 0.7 ; (iii) normalized tag density in the pooled neuronal sample > 0.005 ppm; (iv) P value for t test on tag density between three infant samples against the three old samples < 0.15 . We obtained 1,292 peaks that are enriched in H3K4me3 in

the three infant samples and 589 are proximal to TSS. We used the reciprocal criteria and detected 317 peaks that were enriched in H3K4me3 in the three oldest samples and 101 were proximal to TSS.

We used the following criteria to detect subject-specific peaks, which were unique to each neuronal sample against the other 10 neuronal samples (1): MACS P value $< 1e^{-5}$ run with the sample of interest against 10 other neuronal samples pooled (2); log-ratio (base 2) of tag density between the given and 10 other samples > 1 (3); normalized tag density in the given neuronal sample > 0.005 ppm.

We used the DAVID 2008 web-server (31) for detecting enriched GO (Gene Ontology) categories (32).

Peaks from different individuals were compared with each other and with other publicly available ChIP-Seq datasets. Overlap was recorded if two peaks overlapped by one or more bps. All analyses were performed with custom-built PERL scripts.

To test whether a peak was in the vicinity of genomic annotations (e.g., TSSs, CpG islands, and EST), we used datasets on ESTs, CpG islands, and TSSs downloaded from the UCSC genome browser (<http://genome.ucsc.edu>).

As sequencing depth increases, weaker peaks become detectable. Therefore, datasets with a larger number of tags could result in a larger number of significant peaks at the same FDR cutoff. To ensure that the results we obtained were not biased by this tendency, we performed exactly the same calculations for normalized sets, by randomly removing tags from the samples until the total number of tags became equal for each sample. The statistics about unique/common samples, overlap with two ENCODE cell lines, and distance to TSSs remained similar with the normalized datasets.

Analysis of the correlation of percent proximal peaks with PMI and pH values of the samples was performed on the normalized sets. Sampling was repeated ten times for each brain and the average was reported.

In Situ Hybridization. In vitro transcribed RNA was labeled with digoxigenin according to manufacturer's instructions (Roche). RNA probes used were between 113 nt and 191 nt. In situ hybridization was done as described previously (22).

ACKNOWLEDGMENTS. We thank Dr. Ellen Kittler and Dr. Maria Zapp and the staff of the UMMS Deep Sequencing Core for expert advice and support, Dr. Oliver Rando for insightful discussions, and Dr. Yin Guo for excellent technical assistances. We are also grateful for the H3K4me3 datasets on GM12878 and K562, which were produced by the Bradley Bernstein group in the ENCODE consortium and distributed by the ENCODE data coordination center at UCSC. This work was supported by National Institute of Mental Health Grants 5RC1MH08047 and 5R01-MH071476 and by funds from the International Mental Health Research Organization (to S.A.) and National Science Foundation Grant DBI 085008 (to Z.W.).

- Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447:433–440.
- McGowan PO, et al. (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* 12:342–348.
- Tsankova N, Renthal W, Kumar A, Nestler EJ (2007) Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci* 8:355–367.
- Shaw P, et al. (2008) Neurodevelopmental trajectories of the human cerebral cortex. *J Neurosci* 28:3586–3594.
- Rakic P, Ayoub AE, Breunig JJ, Dominguez MH (2009) Decision by division: Making cortical maps. *Trends Neurosci* 32:291–301.
- Jiang Y, Matevosian A, Huang HS, Straubhaar J, Akbarian S (2008) Isolation of neuronal chromatin from brain tissue. *BMC Neurosci* 9:42.
- Park PJ (2009) ChIP-seq: Advantages and challenges of a maturing technology. *Nat Rev Genet* 10:669–680.
- Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447:407–412.
- Connor CM, Guo Y, Akbarian S (2009) Cingulate white matter neurons in schizophrenia and bipolar disorder. *Biol Psychiatry* 66:486–493.
- Tamamaki N, et al. (2003) Green fluorescent protein expression and colocalization with calretinin, parvalbumin, and somatostatin in the GAD67-GFP knock-in mouse. *J Comp Neurol* 467:60–79.
- Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.
- Birney E, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
- Weickert CS, et al. (2009) Transcriptome analysis of male-female differences in prefrontal cortical development. *Mol Psychiatry* 14:558–561.
- Yan J, et al. (2008) Analysis of the neuroligin 4Y gene in patients with autism. (Translated from eng). *Psychiatr Genet* 18:204–207.
- Vawter MP, et al. (2004) Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology* 29:373–384.
- Huang HS, Matevosian A, Jiang Y, Akbarian S (2006) Chromatin immunoprecipitation in postmortem brain. *J Neurosci Methods* 156:284–292.
- Svenningsson P, et al. (2004) DARPP-32: An integrator of neurotransmission. *Annu Rev Pharmacol Toxicol* 44:269–296.
- Harrison PJ, Weinberger DR (2005) Schizophrenia genes, gene expression, and neuropathology: On the matter of their convergence. *Mol Psychiatry* 10:40–68.
- Donohoe G, et al. (2009) Influence of NOS1 on verbal intelligence and working memory in both patients with schizophrenia and healthy control subjects. *Arch Gen Psychiatry* 66:1045–1054.
- Shilatifard A (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol* 20:341–348.
- Bandeira F, Lent R, Herculano-Houzel S (2009) Changing numbers of neuronal and non-neuronal cells underlie postnatal brain growth in the rat. *Proc Natl Acad Sci USA* 106:14108–14113.
- Iwase S, et al. (2007) The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* 128:1077–1088.
- Adegbola A, Gao H, Sommer S, Browning M (2008) A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *Am J Med Genet A* 146A:505–511.
- Akbarian S, Huang HS (2009) Epigenetic regulation in human brain-focus on histone lysine methylation. *Biol Psychiatry* 65:198–203.
- Huang HS, et al. (2007) Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. *J Neurosci* 27:11254–11262.
- Lim DA, et al. (2009) Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature* 458:529–533.
- Kim SY, Levenson JM, Korsmeyer S, Sweatt JD, Schumacher A (2007) Developmental regulation of Eed complex composition governs a switch in global histone modification in brain. *J Biol Chem* 282:9962–9972.
- Tan HY, Callicott JH, Weinberger DR (2009) Prefrontal cognitive systems in schizophrenia: Towards human genetic brain mechanisms. *Cogn Neuropsychiatry* 14:277–298.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
- Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.
- Dennis G, Jr, et al. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 4:P3.
- Ashburner M, et al. (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29.