nature neuroscience

Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex

Andrew E Jaffe[1–3], Yuan Gao[1], Amy Deep-Soboslay[1], Ran Tao[1], Thomas M Hyde[1,4,5], Daniel R Weinberger[1,4–7] & Joel E Kleinman[1,7]

DNA methylation (DNAm) is important in brain development and is potentially important in schizophrenia. We characterized DNAm in prefrontal cortex from 335 non-psychiatric controls across the lifespan and 191 patients with schizophrenia and identified widespread changes in the transition from prenatal to postnatal life. These DNAm changes manifest in the transcriptome, correlate strongly with a shifting cellular landscape and overlap regions of genetic risk for schizophrenia. A quarter of published genome-wide association studies (GWAS)-suggestive loci (4,208 of 15,930, $P < 10^{-100}$) manifest as significant methylation quantitative trait loci (meQTLs), including 59.6% of GWAS-positive schizophrenia loci. We identified 2,104 CpGs that differ between schizophrenia patients and controls that were enriched for genes related to development and neurodifferentiation. The schizophrenia-associated CpGs strongly correlate with changes related to the prenatal-postnatal transition and show slight enrichment for GWAS risk loci while not corresponding to CpGs differentiating adolescence from later adult life. These data implicate an epigenetic component to the developmental origins of this disorder.

DNA methylation (DNAm) is important for epigenetic regulation of gene expression, orchestrating tissue differentiation and development during fetal life, childhood and adolescence, and guiding functional activity in adulthood. Epigenetic control is especially important in the human brain, where gene expression is extremely dynamic during fetal and infant life, and becomes progressively more stable at later periods of development[1,2]. Dysregulation of these precise and coordinated gene expression changes through epigenetic mechanisms may have a vital role in the pathogenesis of neurodevelopmental disorders, such as schizophrenia (SZ) and related conditions[3–5]. Pathologically, these epigenetic changes, acting through gene expression, could disturb the formation of essential brain circuits, fitting into one prevailing set of hypotheses for the causes of schizophrenia, namely the neurodevelopmental hypothesis[6].

Exogenous factors have been associated with altering DNAm levels, both at specific loci and globally (averaged across all repeat elements), including changes in diet[7] and exposure to cigarette smoking[8] and arsenic[9]. Extensive research implicates environmental variables in the development of schizophrenia, especially during fetal and perinatal life, including maternal stress and infections, obstetric complications, and maternal nutrition during pregnancy[6]. For example, the Dutch famine of 1944–1945 led to a spike in the number of cases of schizophrenia two decades later[10]. Many of these factors have previously been associated with altering DNA methylation levels[11,12]. Lastly, several recent papers have explored the role of sequence variation on site- and region-specific DNA methylation[13,14]. The DNA sequence itself has a large role in the maintenance of DNAm[15], providing one potential mechanism, namely changes in DNAm, for the clinical associations of single nucleotide polymorphism (SNPs) from large GWAS.

We generated DNAm data from postmortem dorsolateral prefrontal cortex (DLPFC) brain tissue from 526 individuals using the Illumina HumanMethylation450 (450k) microarray, extending earlier DNAm maps of human frontal cortex in normal subjects by 20-fold increased genomic coverage (>485,000 versus 27,000 probes) in a much larger sample size (335 versus 108 normal subjects)[2] and creating a more comprehensive landscape of epigenetic development in the human brain than was previously available. Although previous efforts to comprehensively measure DNA methylation across the epigenome using whole genome bisulfite sequencing have identified many important features of brain development[16], our study complements this work, as we used a much larger sample at more continuous ages, albeit at lower genome-wide coverage, to obtain population-level spatial dynamics of DNA methylation across brain development.

RESULTS

Widespread DNAm changes contrast fetal from postnatal life

We obtained high-quality DNAm data on 526 individuals from across the lifespan, including adults diagnosed with schizophrenia (**Supplementary Table 1**) using the Illumina 450k microarray (Online Methods) after removing probes on the sex chromosomes and those containing known SNPs at the single-base extension and

[1]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, Maryland, USA. [2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [3]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [4]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. [5]Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. [6]Department of Neuroscience and the Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. [7]These authors contributed equally to this work. Correspondence should be addressed to A.E.J. (andrew.jaffe@libd.org).

Received 26 August; accepted 26 October; published online 30 November 2015; doi:10.1038/nn.4181

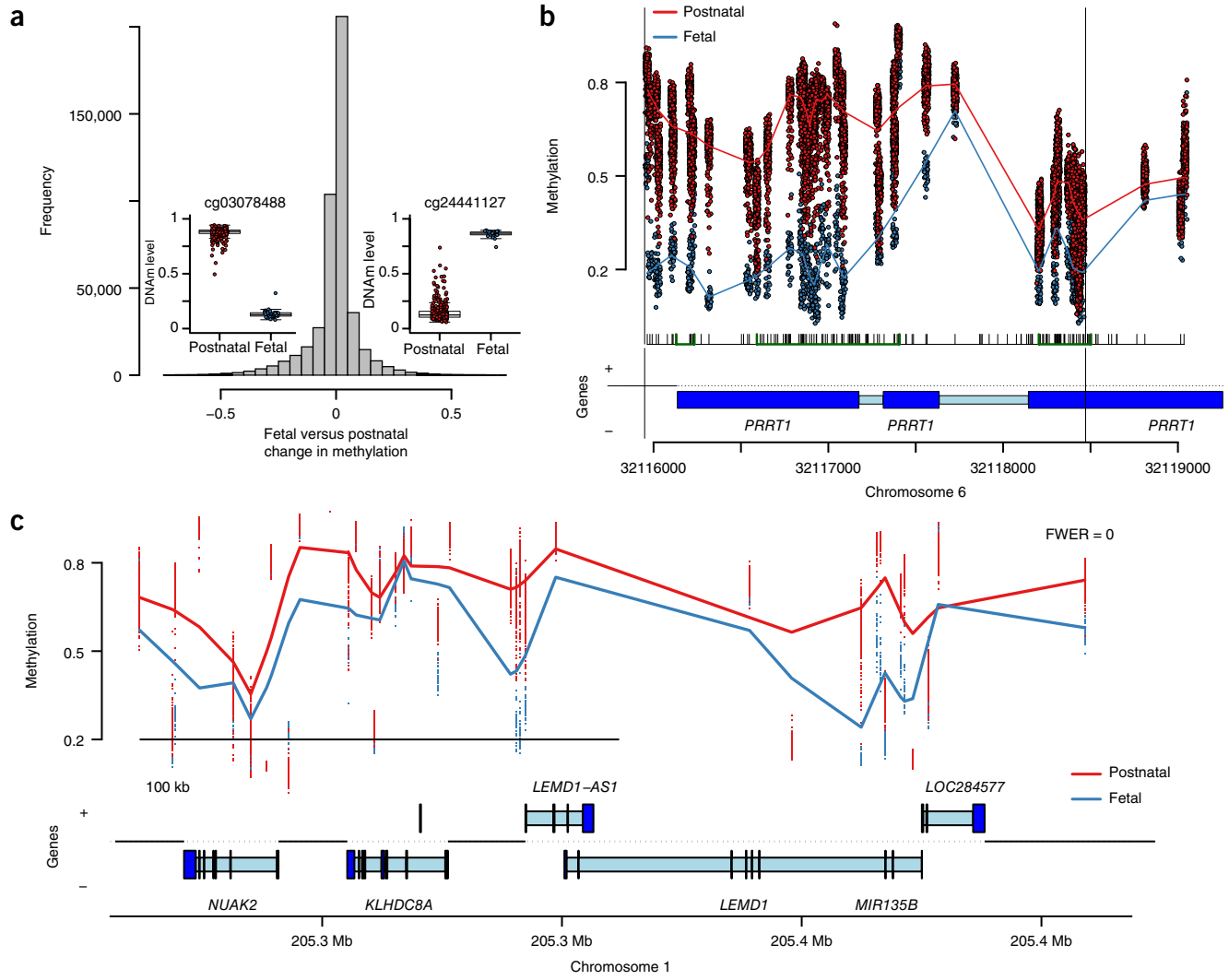

Figure 1 Differentially methylated loci comparing pre- and postnatal control subjects show large differences in DNA methylation. (**a**) Distribution of differences in DNAm across all individual CpGs and probes revealed many sites with large changes in DNAm. Insets, examples of differentially methylated loci. For box plots, center line is the median, limits are the interquartile range (IQR), and whiskers are 1.5× the IQR. (**b**) An example DMR representing regional differences in DNAm levels. (**c**) An example methylation block representing long-range changes. Proportion methylation is shown on the *y* axis of the insets in **a**–**c**. Gene annotation panels in **b** and **c** are based on Ensembl annotation; dark blue represents exons and light blue represents introns.

target CpG sites, leaving 456,513 autosomal probes for analysis. We compared fetal (n = 35) and postnatal (n = 300) non-psychiatric control samples (including newborns and children) to identify changes in DNAm associated with the transition from the second fetal trimester to postnatal life, at varying spatial scales, adjusting for negative control probe factors to control for potential 'batch' effects (Online Methods). At the single probe level, the majority of assayed CpGs (N = 231,415, 50.7%) were significantly differentially methylated (at Bonferroni-adjusted P value, $P_{bonf} < 0.05$, corresponding to $P < 1.10 \times 10^{-7}$; https://github.com/andrewejaffe/devMeth450k/blob/master/tables/jaffe_onlineTable1_DMPsWithExprs.csv.gz), suggesting a vastly different epigenetic landscape of the prefrontal cortex during fetal compared with postnatal life (**Fig. 1a**). Although the Illumina 450k largely targets CpGs in and around gene promoter regions (islands, shores and shelves)[17], almost every annotated gene (17,300 of 19,771 via UCSC knownGene hg19 table, 87.5%) contained at least one differentially methylated CpG within 5 kb. These effects were further evident across the entire array; using principal component analysis, the first principal component, explaining 55.6% of variability in the data, strongly correlated with pre- versus post-natal life (**Supplementary Fig. 1**).

These differentially methylated CpGs were classified into differentially methylated regions (DMRs) based on a 'bump hunting' approach[18] that required at least 10% changes in DNAm at adjacent CpGs, resulting in 6,480 statistically significant DMRs (at family-wide error rate, FWER < 5%, median width = 74 bp; **Fig. 1b**, **Supplementary Table 2** and **Supplementary Fig. 2**) which overlap (within 5 kb) 4,557 unique genes, which were strongly enriched for gene sets related to brain development and morphogenesis (**Supplementary Table 3** and **Supplementary Analysis**). We also identified 896 regions of long-range differential methylation (**Fig. 1c**), termed blocks[19], using an approach adapted to the Illumina 450k[20] from whole genome bisulfite sequencing (WGBS) data (median width = 91.8 kb; **Supplementary Table 4**). These blocks overlapped a combined 731 genes that were significantly enriched for 97 GO gene sets (at $P_{bonf} < 0.05$; **Supplementary Table 5**), largely related again to brain development, including neuron differentiation (109 of 1,201 genes in set, $P = 6.99 \times 10^{-18}$), generation of neurons (114 of 1,308, $P = 2.23 \times 10^{-17}$) and axonogenesis (65 of 569

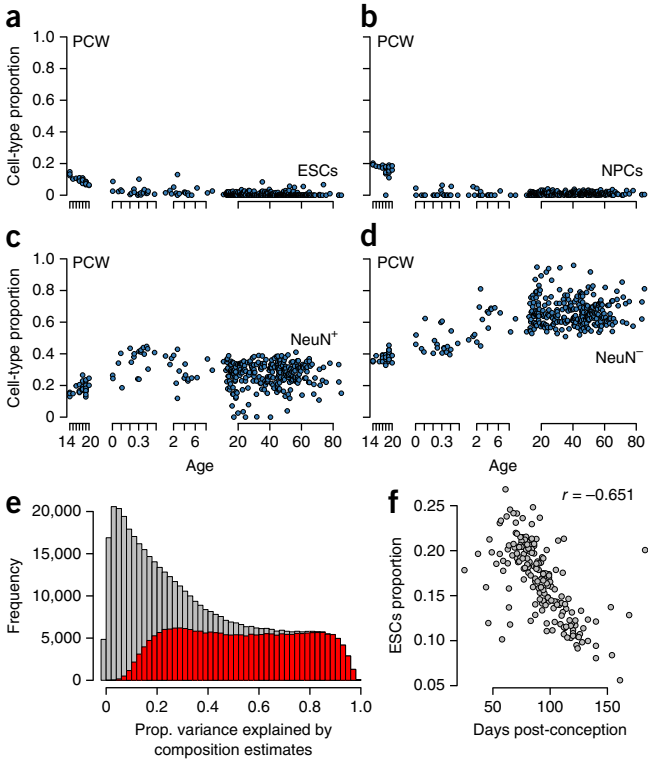

Figure 2 A changing neuronal phenotype across brain development. (**a**–**d**) Composition proportions per sample plotted versus age; the first subpanel in each represents age in post-conception weeks, and the remaining three subpanels show age in years. (**e**) Proportion of variance, R^2, explained by cell composition at each CpG (gray) in which the proportion of CpGs showing significant age stage-related (fetal versus postnatal) changes are shown in red. (**f**) The estimated proportion of embryonic stem cells (ESCs) versus post-conception days from a previous study[23] showed strong association. NPCs, neural progenitor cells; PCW, post-conception weeks.

genes, $P = 1.73 \times 10^{-15}$). We also found significant overlap of differentially methylated blocks with previously identified blocks associated with cancer (791 of the 896 blocks; 88.3%, odds ratio of enrichment: 1.56, χ^2 P value = 3.63×10^{-86}), including skewed directionality; fetal samples in these DLPFC-associated blocks were almost exclusively hypomethylated compared to adult samples, which mirrored the hypomethylated blocks associated with cancer (82.3% of overlapping blocks). These results support the idea that these blocks may represent more general developmental and/or proliferative phenomena[21], but also link these long-range changes in DNAm to mechanisms underlying human brain development.

Many of these DNAm changes were further confirmed using WGBS data comparing one fetal sample to two adult NeuN$^+$ and NeuN$^-$ samples from a previous study[16] (**Supplementary Analysis**) at all three spatial scales. Using Epigenome Roadmap data[22] , we found that many of the DNAm changes at birth were further associated with enhancer and repressive chromatin states in the adult DLPFC and suggest vastly different epigenetic landscapes of the prenatal versus postnatal human brain, consistent with our DNAm data (**Supplementary Table 6** and **Supplementary Analysis**). By matching these DNAm samples to publicly available gene expression data from the same donor[1], we confirmed that many of these DNA methylation changes were associated with nearby gene expression levels, including at the CpG (71.0%, **Supplementary Fig. 3**), DMR (87.3%, **Supplementary Fig. 4**) and block (73.3%) spatial scales (**Supplementary Analysis**). These data therefore suggest a functional role for many of these DNA methylation changes in the developing human brain.

DNAm changes reflect a shifting neuronal composition

In a small sample of 36 non-psychiatric control subjects (including 32 in this study), we recently uncovered evidence of significant differences in the neuronal composition across the lifespan, including a loss of a progenitor-like epigenetic signature after birth, and a rising non-neuronal signature in postnatal life[23]. We sought to more fully characterize these composition changes in this much larger sample, and again found strong evidence of age-dependent changes in cell composition based on cell-type epigenetic signature analysis (**Fig. 2**). We identified significant linear changes in measures of composition in the second trimester of fetal life, reflecting the large gene expression changes present in this developmentally important time period[1], including a decrease in progenitor-like cells based on DNAm signatures found in embryonic ($P = 1.21 \times 10^{-24}$; **Fig. 2a**) and neural progenitor (5.74×10^{-25}; **Fig. 2b**) cells, and a rise of mature adult neurons ($P = 3.57 \times 10^{-24}$; **Fig. 2c**) and non-neuronal cells ($P = 8.09 \times 10^{-86}$; **Fig. 2d**). The composition values of the fifth cell type in the data set (embryonic stem cell–derived dopamine neurons) did not differ between pre- and postnatal samples ($P = 0.99$). These composition profiles explained much of the variability in DNAm levels at individual CpGs, particularly at those differentially methylated from prenatal to postnatal life (**Fig. 2e**). In addition, by using the DNAm-derived composition profiles in the expression data, we found substantial, but lesser, association with gene expression levels (**Supplementary Fig. 5**). We hypothesize that these weakened associations with composition in gene expression data may result from additional layers of unexplored epigenetic regulation, as well as the fact that DNA and RNA from some of these samples were obtained from different tissue dissections.

We found these composition changes across brain development in publicly available data sets. For example, similar changes during fetal brain development are seen in a sample of 179 undissected fetal homogenate tissues[24], particularly in the proportion of pluripotent-like cells ($P = 6.03 \times 10^{-23}$; **Fig. 2f** and **Supplementary Analysis**), as well as the other cell types examined (**Supplementary Fig. 6**). Large effects of cellular composition were also observed in the developing and aging postnatal brain using DNAm data on 17 postnatal subjects and brains across 16 brain regions (including 11 neocortical) from the BrainSpan project[25] (**Supplementary Fig. 7** and **Supplementary Analysis**). These convergent findings suggest that DNAm changes in the developing and aging brain reflects a shifting cellular composition—namely loss of progenitor-like cells at birth followed by a rise of non-neuronal cells in postnatal life—that we have previously shown manifests in the transcriptome[23].

DNAm changes are enriched for schizophrenia risk loci

We next tested for significant enrichment between the CpGs that display epigenetic differences associated with the prenatal-postnatal transition and genomic loci associated with schizophrenia risk in the latest Psychiatric Genomics Consortium (PGC) genome-wide association study (GWAS)[26]. Of the 456,513 probes on the Illumina 450k used in differential methylation analysis, 5,476 were within the 108 genome-wide significant risk loci (specifically, the linkage disequilibrium (LD) blocks defined by the loci). These particular CpGs were more likely to be differentially methylated across the fetal-postnatal transition (2,903 of 5,476, 53.0%, in the loci compared with 228,512 of 451,037, 50.7% outside; odds ratio (OR) = 1.10, χ^2 P value = 5.75×10^{-4}; **Supplementary Table 7**). This association was driven by a majority

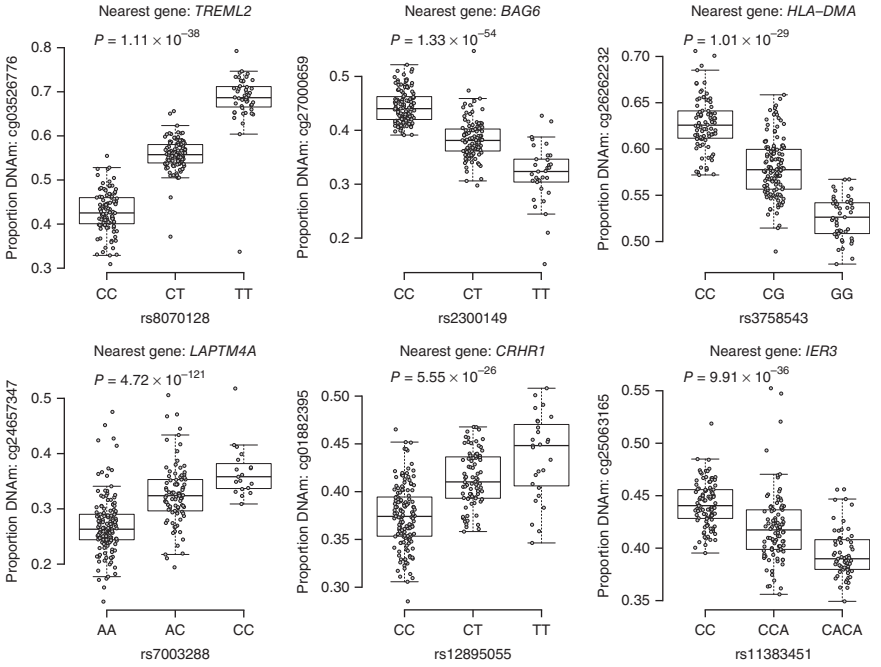

Figure 3 Examples of meQTLs for six GWAS-associated variants with nearby DNA methylation levels. *y* axis shows DNA methylation level at a particular probe and the *x* axis represents genotype at a particular SNP. *P* value corresponds to the effect of genotype on DNAm level, adjusting for ancestry and epigenetic principal components. For box plots, center line is the median, limits are the IQR, and whiskers are 1.5× the IQR.

subset of 3,607 CpGs in the PGC regions that were relatively more highly methylated in fetal than in postnatal life (1,848 of 3,607, 51.2%, versus 126,674 of 272,242, 46.5%; OR = 1.21, *P* value = 2.03×10^{-8}); the subset of CpGs more highly methylated in postnatal life were not relatively enriched in the SZ GWAS loci (OR = 0.98, *P* = 0.67).

We also considered GWAS-positive loci for Alzheimer's disease[27] (AD, *N* = 49), Parkinson's disease[28] (PD, *N* = 29) and type 2 diabetes[29] (T2D, *N* = 40) to determine the specificity of our results. Among the CNS disorders, there was no enrichment overall (*P* > 0.3), with only perhaps marginal enrichment among CpGs that were more highly methylated in fetal life among AD GWAS loci (OR = 1.23, *P* = 0.03), suggesting some specificity to schizophrenia. Notably, although we found significant enrichment among the T2D GWAS loci among all birth-associated CpGs (OR = 1.19, $P = 6.56 \times 10^{-3}$), this association was driven by CpGs that were more highly methylated in postnatal life (OR = 1.41, $P = 5.4 \times 10^{-4}$) with no association among the CpGs more highly methylated in prenatal life (OR = 1.01, *P* = 0.92). This finding may reflect adult lifestyle influences on epigenetic states that are associated with risk for T2D, including diet, body weight and exercise.

DNAm changes associated with the age of illness onset

As a sensitivity analysis to assess the specificity of the fetal-postnatal transition and to contrast possible neurodevelopmental mechanisms with epigenetic alterations around the time of schizophrenia illness diagnosis, we performed a differential methylation analysis comparing controls between 10–25-year-old individuals (*N* = 73) with those older than 25 (*N* = 190) to identify changes in DNAm associated with the typical age of onset of schizophrenia (Online Methods). We found 24,685 CpGs significant at $P_{bonf} < 0.05$ between these two age groups at lesser effect sizes than those associated with the prenatal-postnatal transition; only 58 CpGs showed changes in DNAm greater than 0.1, making it difficult to perform comparable DMR analyses (Online Methods). There were only 313 of the 24,685 CpGs linked to the adolescent period in the PGC2 loci (5.7%), which was not enriched compared with these age-of-onset–associated CpGs outside these loci (5.4%, OR = 1.06, *P* = 0.32). However, we note that stratifying by directionality resulted in significant enrichment among schizophrenia GWAS loci here among a minority of CpGs (*N* = 144) that were more highly methylated in adolescence than later adulthood (OR = 1.42, $P = 3.38 \times 10^{-5}$). Notably, these CpGs also tended to be more highly methylated in fetal than in postnatal life (80.6%), consistent with the general trend in GWAS-positive loci being associated with CpGs that were relatively hypermethylated in fetal compared with postnatal life. Overall, these data suggest that epigenetic changes around the age of onset of schizophrenia do not in general reflect genetic risk mechanisms, in contrast with the associations contrasting fetal to postnatal life, but they may contain a subset of risk loci relatively hypermethylated in late postnatal epigenetic regulation.

Extensive genetic regulation of local DNAm levels

Given the statistical enrichment between locations of early developmentally regulated CpGs and genetic risk for SZ, we attempted to directly determine the association between risk genotype among the genome-wide significant loci of SZ and DNAm levels. We genotyped and imputed these subjects to the 1000 Genomes reference panel, and retained 7,426,085 common SNPs (MAF > 5%, Online Methods) for methylation quantitative trait loci (meQTL) analysis to understand how common genetic variation influences DNAm levels. We reintroduced probes on the sex chromosomes (as genotypes across the genome should not correlate with sex) as well as probes with annotated SNPs at the target CpG site (as these could represent true biological signal) while still removing probes with annotated SNPs at the single base extension site (which represent technical signal if associated with genotype), leaving 477,636 CpGs for analysis. We conducted a meQTL analysis in the adult control samples (age > 13, *N* = 258) allowing for up to 20 kb between SNPs and CpGs, adjusting for factors related to ancestry and global epigenetic variation (Online Methods), and identified 4,107,214 significant SNP-CpG methylation associations (meQTLs) at FDR < 1% (corresponding to $P < 8.6 \times 10^{-4}$; http://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE74193&format=file&file=GSE74193%5Fjaffe%5FonlineTable2%5FmeQTLs%5FallPairs%2Ecsv%2Egz, https://github.com/andrewejaffe/devMeth450k/blob/master/tables/jaffe_onlineTable3_meQTLsMostSig.csv.gz and **Supplementary Analysis**). We note that many of these meQTLs identified in the adult samples appeared to be largely consistent in directionality (κ = 0.741) and magnitude (ρ = 0.707) with meQTLs in the smaller set of fetal samples (**Supplementary Fig. 8**). Although many meQTLs may not be cell type specific, as the composition of the fetal brains differed markedly from the adult samples (**Fig. 2**), a subset may impart fetal-specific effects not present in adult subjects[30].

Epigenetic marks such as DNAm have been hypothesized to underlie risk for common disease and to potentially mediate genetic risk

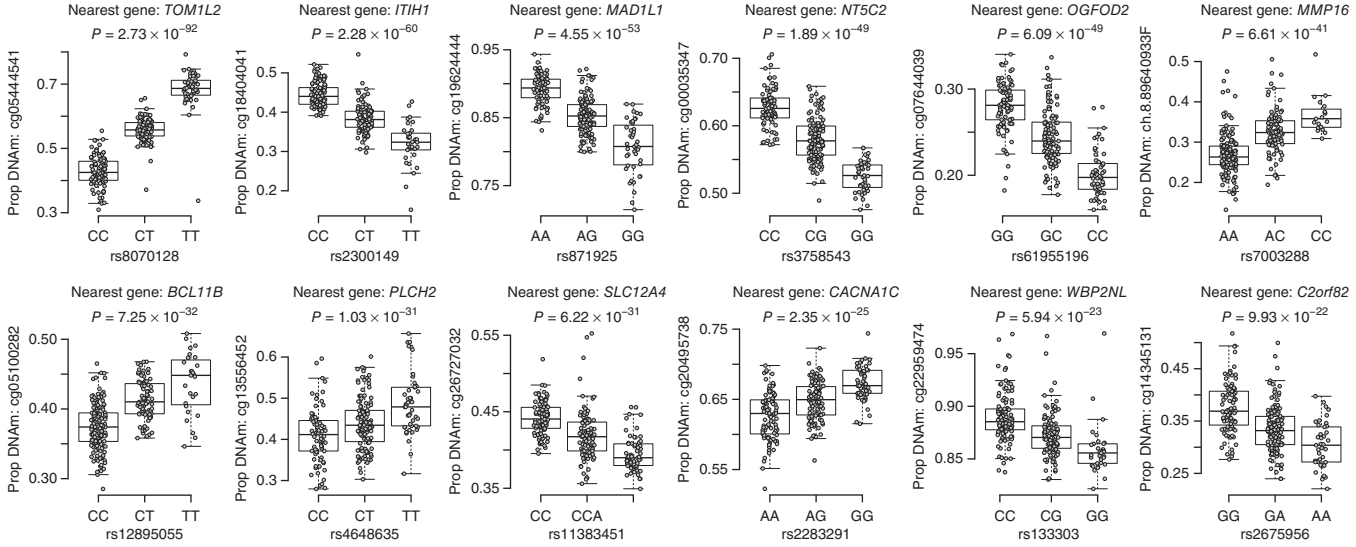

Figure 4 Examples of meQTLs for 12 GWAS-positive loci for schizophrenia. *y* axis shows DNA methylation level at a particular probe and *x* axis represents genotype at a particular SNP. *P* value corresponds to the effect of genotype on DNAm level, adjusting for ancestry and epigenetic principal components. For box plots, center line is the median, limits are the IQR, and whiskers are 1.5× the IQR.

identified from large GWAS[31]. We first examined all genetic variation previously identified in diverse GWAS using the NHGRI-EBI GWAS Catalog, which, at the time of access, contained 28,870 genome-wide significant associations across 1,290 disease traits (with 15,930 unique SNPs by rs number)[32]. We found that 4,208 GWAS-associated SNPs for any trait were significant DLPFC meQTL in our data (26.5% of all catalog SNPs, and 31.7% of those tested for meQTLs (**Supplementary Table 8**); this enrichment of DLPFC meQTLs in GWAS-associated SNPs was highly significant (odds ratio = 1.74, χ^2 $P < 10^{-100}$). Notably, the GWAS-associated SNPs were clinically associated with 877 different disorders and/or traits affecting many tissues in the body, further suggesting that many meQTLs are not cell type specific. We highlight examples of highly significant meQTLs where the clinical risk SNP disrupts a reference CpG dinucleotide (**Fig. 3**); these epigenetic effects could represent one possible mechanism by which these variants manifest risk[31] (**Supplementary Analysis**).

As many of the risk variants for schizophrenia in the NHGRI GWAS Catalog only reached genome-wide suggestive, but not significant, evidence for association, we sought to determine the proportion of PGC2 schizophrenia risk genotypes that associated with nearby DNAm levels, which would suggest a possible mechanism of risk. We examined all marginally (at $P < 1 \times 10^{-4}$, N = 1,302 of 2,107 in our data, Online Methods) and genome-wide (at $P < 5 \times 10^{-8}$, 104 of 111 in our data, Online Methods) significant 'index' SNPs and their highly correlated proxies ($R^2 > 0.6$) in the PGC2 49 study discovery data set. Among this subset of variants, we found that 579 of 1,302 (44.4%) of all marginally significant loci and 62 of 104 (59.6%) genome-wide significant loci interrogated had a risk or proxy SNP that was a meQTL in our cortex DNAm data (**Supplementary Table 9**). We hypothesize that these epigenetic signals highlight the particular risk gene in a locus with wide LD (12 loci that feature meQTL P values < 1.0×10^{-20} are shown in **Fig. 4**). These CpGs associated with risk variants mapped to *TOM1L2*, *ITIH1*, *MAD1L1*, *NT5C2*, *OGFOD2*, *MMP16*, *BCL11B*, *PLCH2*, *SLC12A4*, *CACNA1C*, *WBP2NL* and *C2orf82*. DNAm levels proximal to risk variants for schizophrenia may therefore influence or possibly mediate the effect of genotype on clinical risk for a large proportion of genome-wide significant loci.

DNAm changes and the diagnosis of schizophrenia

Finally, we performed a differential methylation analysis for schizophrenia diagnosis at the single CpG level by comparing 191 adult patients with schizophrenia to the 240 (of 335) non-psychiatric controls all with ages greater than 16 (**Supplementary Table 1**). The schizophrenia patients had relatively typical age of onset (mean = 22.5 years, s.d. = 7.0 years) and the majority were on anti-psychotic medications at the time of death (64.0%), assessed via chart review and/or toxicology on brain tissue (Online Methods). There were similar proportions of race, sex and causes of death (other than suicide) in the control and schizophrenia group, but the patients were more likely to be older and smoke, have a lower tissue pH, and longer postmortem interval (PMI). We further note that these patients had a relatively young average age at death compared with other post-mortem samples[33], potentially reducing the cumulative effect of chronicity, smoking and antipsychotics on DNAm levels. In these adult samples, we did not identify significant differences in composition of any estimated cell types between cases and controls ($P > 0.05$; **Supplementary Fig. 9a**), though cellular composition explained a large component of variability in the data (**Supplementary Fig. 9b**).

We modeled DNAm levels as a function of diagnosis and adjusted for age, race and the first four PCs of the negative control probes on the microarrays (Online Methods). These negative control PCs were strongly associated with processing plate and microarray slide (**Supplementary Fig. 10**); only the fourth negative control PC was marginally associated with diagnosis ($P = 0.003$). After initial differential methylation analyses using all 431 samples, we noticed that one of the three processing plates containing both patients with schizophrenia and controls drove much of the differential methylation signal (**Supplementary Fig. 11**); we removed the samples from this plate and reran the analysis on 244 subjects (108 patients with schizophrenia and 136 controls), identifying 2,104 probes/CpGs that were significant at a Bonferroni-adjusted $P < 0.05$ (**Supplementary Table 10**). Almost all of these changes involved hypomethylation of individual CpGs in patients compared with controls (N = 2,043, 97.1%) with much smaller differences in methylation levels than those that found in the prenatal and postnatal contrast analysis (mean absolute proportion change, 0.013; IQR, 0.010–0.01.8). This pattern of diagnosis (state) associated

hypomethylated CpGs contrasts with the association of risk-associated GWAS loci (trait) with principally relatively hypermethylated CpGs during the fetal and adolescent timeframe. Nevertheless, these differentially methylated CpGs were in or near (within 10 kb) genes significantly enriched for embryo development ($P = 1.9 \times 10^{-9}$), cell fate commitment ($P = 2.68 \times 10^{-7}$) and nervous system differentiation ($P = 1.9 \times 10^{-7}$) by Gene Ontology analysis[34] (**Supplementary Table 11**). One of the top differentially methylated CpGs lay in the promoter of *HAT1*, a histone acetyltransferase that primarily targets lysines on the H4 nucleosome[35], perhaps manifesting as differences in histone acetylation in conjunction with DNAm changes[36]. These hypomethylated CpGs were enriched for being in transcriptional start site chromatin states (70.6% compared with 21.7% for the entire array), and were depleted for enhancer, repressor and quiescent chromatin states (**Supplementary Table 7**) in independent adult DLPFC samplese[22], suggesting that they may influence gene expression. The lack of enrichment of these diagnosis-associated CpGs with adult-associated biology suggests that these CpGs reflect environmental influences on early developmental events that increase risk independent of genetic risk influences.

A recent study identified four significant CpGs that were differentially methylated in a smaller sample of brains of schizophrenia patients ($N = 20$) and controls ($N = 23$) across a discovery and replication cohort[37], but we found only one of these CpGs that was directionally consistent and marginally significant in our much larger data set (near *GSDMD* (cg26173173) at $P = 0.02$). We failed to replicate the other three CpGs: cg24803255 and cg08171022 were also higher and lower in cases than controls, respectively, but neither were significant ($P = 0.32$ and $P = 0.97$, respectively), and cg00903099 was marginally significant ($P = 0.02$), but higher in cases than in controls in our data rather than the reported hypomethylation. Conversely, only one of our CpGs was marginally significant (at $P < 0.05$) and directionally consistent in both their discovery and replication data sets when treating each separately, potentially highlighting uncertainties in case control analyses of DNAm, including heterogeneity in the clinical disorder, differing epiphenomena related to diagnosis and illness state, and differences in the ascertained tissue for postmortem human brain studies.

Despite concern about epiphenomena (**Supplementary Fig. 12** and **Supplementary Analysis**), we did find slight, but significant, enrichment of our 2,104 diagnosis-associated CpGs in the PGC loci—40 of 2,104 CpGs (1.9%) compared with only 1.3% of the rest of CpGs on the array (OR = 1.6, $P = 0.004$)—but none of these 40 were meQTLs to any SNPs, including the risk-associated SNPs identified in the PGC. Overall, among these 2,104 CpGs showing DNAm level differences between patients and controls, only 97 were genome-wide significant meQTLs, a sixfold decrease in enrichment than would be expected by chance (OR = 0.165, $P = 2.32 \times 10^{-86}$, Online Methods), even though there was strong global correlation among meQTLs identified in adult controls and in the patients with schizophrenia (**Supplementary Fig. 13**).

Finally, our data draw parallels to an earlier report of enrichment of DNAm changes during fetal life among schizophrenia diagnosis–related CpGs[37]. Even in our smaller, but regionally focused, fetal sample, 1,986 of 2,104 of the CpGs associated with diagnosis were significantly differentially methylated between fetal and postnatal life (at $P_{bonf} < 0.05$, OR = 16.5, enrichment P value $< 10^{-100}$). In contrast, these CpGs were strongly depleted (OR = 0.26, $P = 1.88 \times 10^{-15}$) for those CpGs showing significant differences comparing adolescent to adult controls, reflecting age-related changes occurring near the age of onset of schizophrenia; only 31 of 2,104 CpGs were associated with age-related changes around schizophrenia onset as well as diagnosis. These contrasts suggest that the diagnosis-associated CpGs are not related to epigenetic events germane to illness onset, but appear to reflect lifelong epigenetic states established early in development. This is further supported by the observation that these CpGs largely hypomethylated for diagnosis (compared with adult controls) were relatively highly methylated in fetal life ($\rho = -0.63$, $P < 10^{-20}$; **Supplementary Fig. 14**) and appeared to further diverge from fetal levels compared with the adult non-psychiatric controls. Thus, the schizophrenia associations at these CpGs strongly reflect DNAm changes related to early developmental events supporting a neurodevelopmental component not only to genetic risk, but also to environmental risk of this debilitating disorder.

DISCUSSION

We identified changes in DNA methylation levels associated with genetic sequence and developmental stage in one of the largest studies of postmortem human brain tissue to date. The most extensive changes in the methylome were found at local, regional and long-range spatial resolutions when comparing prenatal and postnatal specimens, and we suggest that this represents, in part, shifts in neuronal composition across the lifespan and corresponds to strong changes in gene expression profiles. Notably, these developmentally associated changes in DNAm were significantly enriched for genomic regions that confer clinical risk for schizophrenia. Many risk variants across the catalog of GWAS-associated loci in studies of common medical disorders themselves associated with nearby DNAm levels, termed meQTLs, suggesting potential mechanisms by which genetic risk propagates in the population. Lastly, we found that several thousand individual CpGs demonstrated small, but statistically significant, differences in DNAm levels between adult patients with schizophrenia and controls that did not appear confounded by cellular composition or smoking. The differences found between patients and controls appear to represent epigenetic marks that principally associate with early neurodevelopment and not with events that herald the onset of the disorder or that characterize adult brain biology. Overall, the data suggest that both the genetic and environmental risk components of schizophrenia involve early developmental influences.

The widespread methylome changes that occur across brain development, ranging from early fetal life[24], the transition into postnatal life and through adulthood[25], appear to track first the loss of immature neurons before birth followed by the rise of non-neuronal cell types in postnatal life through adulthood (**Fig. 2**). Although the quantitative estimates of cell composition employed here used a series of cell types that combined epigenetic data from adult human tissue and derived cellular systems, the proportion of pluripotent-like cells were quite consistent across two independent data sets and brain regions, with ~15% of the cells manifesting this signature by 14 post-conception weeks (**Fig. 2a**,**f**), and may relate to recent classifications of replicating versus quiescent fetal neurons using single-cell analysis[38].

Deviations from these essential developmental trajectories during critical windows of development from conception to young adulthood may interfere with the carefully coordinated temporal and spatial dynamics of gene expression through a combination of genetic and epigenetic factors[3,5,39,40] that may contribute to risk for schizophrenia and other neurodevelopmental disorders. Indeed, the CpGs that track this changing neuronal phenotype, such as those that differ in DNAm levels comparing pre- and postnatal samples, were enriched by genomic location of regions that confer genetic risk for schizophrenia[26], as are changes in transcriptome across brain development[23,41,42]. However the mechanisms by which DNAm changes that track shifting neuronal phenotypes alter risk for schizophrenia

appear to be largely unknown, and will likely require more cell type–specific assays to focus on individual cell populations across brain development to reduce the strong composition effects observed in homogenate brain tissue. This is likely to be a complex conundrum, as composition measures in homogenate tissue did not differ between cases and control samples and controlling for composition did not alter the CpG differences between patients and controls.

Conversely, homogenate brain tissue appears to represent a powerful tool for better understanding how genotypes identified in large population-based GWAS may manifest risk for neurodevelopmental and other brain disorders. Indeed, many meQTLs identified in the DLPFC during adult life appear to be consistent in fetal life and reach genome-wide significance in larger fetal samples[30], despite very different cellular compositions, suggesting that many of these variations serve conserved regulatory roles in multiple cell types. Furthermore, DNAm levels may be a more proximal read out of genetic variation than gene expression levels[43]; we identified that 62 of 104 (59.6%) genome-wide significant genetic loci for schizophrenia risk were associated with local DNAm levels, compared with a report suggesting that only 18 of 108 (16.7%) GWAS-positive loci are eQTLs across the human brain (**Supplementary Table 4**, Worksheet 2 in Schizophrenia Working Group of the Psychiatric Genomics[26]).

The small, but significant, differences in DNAm levels at individual CpGs between patients with schizophrenia compared with controls highlight the intrinsic tradeoff between statistical gains in increasing sample size versus the clinical and, likely by extension, molecular heterogeneity of the clinically ill state. The genetic heterogeneity of schizophrenia is reflected in small odds ratios (<1.1) for individual genomic loci in GWAS that reach genome-wide significance only because of very large sample sizes. These odds ratios look similar in magnitude to our results for differentially methylated CpGs, suggesting likely epigenetic heterogeneity as well. Unlike genetic sequence, which is largely determined at conception, these epigenetic signals are malleable across the lifespan, and the many epiphenomena that differentiate patients from controls may leave their marks on the epigenome, perhaps differently in different subpopulations of patients. These epiphenomena include the influence of medical therapy, chronic illness, nutrition, body weight, alcohol and cannabis use, etc. Untangling which epigenetic marks better relate to the causes versus the consequences of illness will be difficult. Indeed, only a fraction of the illness-associated CpGs, 4.6%, showed association to nearby genetic variants in the meQTL analysis, further suggesting that these findings may be more related to the epiphenomena of the illness state than to the genetic causes of the disorder. Lastly, although these diagnosis-associated CpGs were not confounded by cell composition, it is possible that they have larger effect sizes in individual cellular populations, and new consortia such as the psychENCODE project[44] can better identify the cellular specificity of, and potentially magnify, these effects.

Furthermore, although we observed significant enrichment of the PGC loci in CpGs differing between patients and controls, this was a marginal enrichment, small in comparison with the enrichment with loci showing epigenetic alterations from prenatal to postnatal life. These results suggest that the majority of DNAm differences associated with the illness state are likely unrelated to genetic mechanisms of causation and instead implicate environmental factors. In this context, and also germane to the issue of state-related epiphenomena, it is worth highlighting that the case control differences mapped to genes implicated in early developmental processes, even if they were not linked with genetic risk variation. Thus, the epigenetic associations with schizophrenia, both in terms of illness state and genetic risk, implicate factors, both genetic and environmental, that track with early development and not adult life. Consistent with this conclusion is the additional observation that epigenetic changes associated with adolescence and early adulthood, the typical time of onset of schizophrenia, did not show enrichment of either genetic risk loci or illness state-associated CpG alterations. This observation has potentially sobering implications for attributing a causative role of environmental influences that appear to coincide with the onset of the clinical disorder.

We also explored the relationship of genetic risk loci associated with Alzheimer's disease, Parkinson's disease and T2D. None of these disorders showed the enrichment of CpGs hypermethylated during fetal life associated with risk loci for schizophrenia, although the GWAS catalog of risk-associated variants is smaller in these cases. Notably, there was a small enrichment with CpGs showing significant hypermethylation in postnatal life and risk loci for T2D. Although the interpretation of this finding is highly speculative, it might reflect adult lifestyle influences on risk for T2D.

In conclusion, the epigenetic landscape represented by DNA methylation in the human brain varies markedly across development. Genetic loci implicated in risk for schizophrenia and other CNS disorders were enriched for loci expressing these shifting epigenetic states, particularly those that change from the transition from prenatal to postnatal life. Although these observations do not identify specific molecular mechanisms of the clinical associations, they suggest that there is an important epigenetic intermediate between sequence of risk and cell biology of risk.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw and processed data are available on the Gene Expression Omnibus (GSE74193). R code for data processing and analysis is available from the GitHub repository: https://github.com/andrewejaffe/devMeth450k.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful for the vision and generosity of the Lieber and Maltz Families who made this work possible. We thank the families who donated to this research and we thank A. Feinberg for helpful input on data analyses. This work was supported by the Lieber Institute for Brain Development. A.E.J. was partially supported by 1R21MH102791.

AUTHOR CONTRIBUTIONS

A.E.J. designed the study, performed the data analysis and oversaw the writing of the manuscript. Y.G. oversaw the data generation. A.D.-S. collected phenotype data on all subjects. R.T. performed DNA extractions and contributed to the data generation. T.M.H. collected brain samples and performed tissue dissections to obtain biological materials. D.R.W. designed the study, contributed to the data analysis and interpretation of the results, and oversaw the writing of the manuscript. J.E.K. collected brain samples and provided clinical interpretation of the results. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011).
2. Numata, S. *et al.* DNA methylation signatures in development and aging of the human prefrontal cortex. *Am. J. Hum. Genet.* **90**, 260–272 (2012).
3. Grayson, D.R. & Guidotti, A. The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. *Neuropsychopharmacology* **38**, 138–166 (2013).

4. Waterland, R.A. & Michels, K.B. Epigenetic epidemiology of the developmental origins hypothesis. *Annu. Rev. Nutr.* **27**, 363–388 (2007).
5. Jakovcevski, M. & Akbarian, S. Epigenetic mechanisms in neurological disease. *Nat. Med.* **18**, 1194–1204 (2012).
6. Weinberger, D.R. & Levitt, P. Neurodevelopmental origins of schizophrenia. in *Schizophrenia* (eds. Weinberger, D.R. & Harrison, P.J.) (Wiley-Blackwell, 2011).
7. Heijmans, B.T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. USA* **105**, 17046–17049 (2008).
8. Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* **88**, 450–457 (2011).
9. Reichard, J.F., Schnekenburger, M. & Puga, A. Long term low-dose arsenic exposure induces loss of DNA methylation. *Biochem. Biophys. Res. Commun.* **352**, 188–192 (2007).
10. Susser, E.S. & Lin, S.P. Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944–1945. *Arch. Gen. Psychiatry* **49**, 983–988 (1992).
11. Relton, C.L. & Davey Smith, G. Is epidemiology ready for epigenetics? *Int. J. Epidemiol.* **41**, 5–9 (2012).
12. Cortessis, V.K. *et al.* Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum. Genet.* **131**, 1565–1589 (2012).
13. Schilling, E., El Chartouni, C. & Rehli, M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res.* **19**, 2028–2035 (2009).
14. Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* **43**, 1091–1097 (2011).
15. Bird, A. Putting the DNA back into DNA methylation. *Nat. Genet.* **43**, 1050–1051 (2011).
16. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
17. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692–702 (2011).
18. Jaffe, A.E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* **41**, 200–209 (2012).
19. Hansen, K.D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43**, 768–775 (2011).
20. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
21. Timp, W. *et al.* Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* **6**, 61 (2014).
22. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
23. Jaffe, A.E. *et al.* Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat. Neurosci.* **18**, 154–161 (2015).
24. Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res.* **25**, 338–352 (2015).
25. BrainSpan. *Atlas of the Developing Human Brain*, (http://developinghumanbrain.org) (2011).
26. Schizophrenia Working Group of the Psychiatric Genomics. C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
27. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
28. Nalls, M.A. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
29. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
30. Hannon, E. *et al.* Methylation QTLs in the developing brain: enrichment in schizophrenia-associated genomic regions. *Nat. Neurosci.* advance online publication, doi:10.1038/nn.4182 (30 November 2015).
31. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**, 142–147 (2013).
32. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
33. Nichols, L. *et al.* The National Institutes of Health Neurobiobank: a federated national network of human brain and tissue repositories. *Biol. Psychiatry* **75**, e21–e22 (2014).
34. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
35. Verreault, A., Kaufman, P.D., Kobayashi, R. & Stillman, B. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr. Biol.* **8**, 96–108 (1998).
36. Gavin, D.P. & Sharma, R.P. Histone modifications, DNA methylation and schizophrenia. *Neurosci. Biobehav. Rev.* **34**, 882–888 (2010).
37. Pidsley, R. *et al.* Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biol.* **15**, 483 (2014).
38. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
39. Abdolmaleky, H.M. *et al.* Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **127B**, 51–59 (2004).
40. Mill, J. *et al.* Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* **82**, 696–711 (2008).
41. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
42. Parikshak, N.N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
43. Kleinman, J.E. *et al.* Genetic neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem human brains. *Biol. Psychiatry* **69**, 140–145 (2011).
44. The PsychENCODE Consortium. The PsychENCODE Project. *Nat. Neurosci.* **18**, 1708–1713 (2015).

ONLINE METHODS

Study samples. Brain specimens were donated through the Offices of the Chief Medical Examiners of the District of Columbia and of the Commonwealth of Virginia, Northern District to the NIMH Brain Tissue Collection at the National Institutes of Health in Bethesda, Maryland, according to NIH Institutional Review Board guidelines (Protocol #90-M-0142). Audiotaped informed consent was obtained from legal next-of-kin on every case. Details of the donation process are described elsewhere[45,46]. Additional specimens, including the 35 second-trimester fetal brain tissue samples, were obtained via a Material Transfer Agreement with the National Institute of Child Health and Human Development Brain and Tissue Bank. All postnatal non-psychiatric control donors ($N = 300$) were free from psychiatric and/or neurologic diagnoses and substance abuse according to DSM-IV. Every control donor had toxicology screening to exclude for acute drug and alcohol intoxication/use at time of death, and all fetal tissue was also screened for possible in utero drug exposure.

Tissue processing. All specimens were flash-frozen, and screened for macro- and microscopic neuropathological abnormalities, as previously described[46]. All specimens with significant evidence of neurological disorders, infarcts or other cerebrovascular abnormalities were excluded from study. Brain pH was measured, and postmortem interval (PMI, in hours) was calculated for every sample. Postmortem tissue homogenates of the prefrontal cortex (dorsolateral prefrontal cortex, DLPFC, BA46/9) were obtained from all subjects. Genomic DNA was extracted from 100 mg of pulverized dorsolateral prefrontal cortex (DLPFC) tissue with the phenol-chloroform method. Bisulfite conversion of 600 ng genomic DNA was performed with the EZ DNA methylation kit (Zymo Research).

DNA methylation microarray. DNA methylation was assessed using the Illumina HumanMethylation450 (450k) microarray, which measures CpG methylation across >485,000 probes covering 99% of RefSeq gene promoters[17]. Arrays were run following the manufacturer's protocols. A percentage of the samples were run in duplicate across multiple processing plates to assess technical variability related to DNA extraction and bisulfite conversion. A total of 675 microarrays were scanned on 534 unique subjects – however, note that all analysis used only a single microarray from each sample (see below).

Data processing and normalization. Red and green channel intensity files were obtained for each sample in the idat file format. These files were processed and normalized using the *minfi* Bioconductor package in R[20]. Red and green intensities were mapped to the M(eth) and U(nmeth) channels, and the average intensity for these channels were used to check for low quality samples (0 samples were dropped). Intensities from the sex chromosomes were used to predict sex, and we dropped 8 samples that had predicted sex different from its recorded value (indicating potential sample swaps). Then, the M and U channels were subsequently across-sample quantile normalized using an approach developed previously[20]. Briefly, this approach forces the distribution of type I and type II to be the same by first quantile normalizing the type II probes across samples and then interpolating a reference distribution to which the type I probes are normalized, stratified by region (for example, promoter, shore, island, shelf), which has previously been shown to best minimize the variability between replicates[20]. For all analyses, we retained a single array in the case of duplicates by choosing the sample that had the closest quality profile (via M and U signal intensity) to all other arrays.

Statistical analyses for differential methylation by development. We modeled differential methylation between pre- ($N = 35$) and post-natal ($N = 300$) non-psychiatric controls using linear modeling approaches. After normalization, probes on the sex chromosomes were dropped (which are more difficult to accurately normalize), as were probes annotated with single nucleotide polymorphisms (SNPs) at the target CpG or single base extension (SBE) site according to dbSNP142 with minor allele frequency > 1%, leaving 456,513 autosomal probes for age-associated DNAm analysis. All three approaches, single CpG, DMR and block, used the linear model

$$p_{ij} = \alpha_i + \beta_i Fetal_j + \sum_{k=1}^{4} \gamma_j^k negPC_j^k + \varepsilon_{ij}$$

where p_{ij} is the proportion methylation for probe i and subject j, $Fetal_j$ is a binary variable indicating if the jth sample is pre- or postnatal, and $negPC_j$ are the negative control principal components estimated from the microarray background probes. Therefore α_i represents the mean methylation proportion/level in the postnatal samples, and β_i is the difference in the fetal samples. For CpG-level analyses, we fit the above linear model with the limma R/Bioconductor package[47] to obtain mean differences, moderated t-statistics and corresponding P values, which we adjusted by the number of tests (that is, Bonferroni correction[48]) to conservatively control for multiple testing (as neighboring CpGs tend to be correlated[49], reducing the number of effective tests). While some previous manuscripts remove all probes containing SNPs in any position, we observe that the single CpGs that are significantly associated with birth are in fact depleted of probes containing SNPs (OR = 0.79, $P < 10^{-100}$), as 17.0% of significant CpGs contain an annotated probe SNP compared to 20.6% of non-significant probes.

The analyses at the longer spatial scales, DMRs and blocks, both used the above linear model; DMR analysis generally finds contiguous CpGs where $\beta_i \neq 0$ whereas block analysis first collapses the methylation proportions into one level across neighboring probes per sample (for example, $\overline{p_{ijk}} \rightarrow r_{ij}$) and fits the above model. Regional analysis to find differentially methylated regions (DMRs) and 'block finding' were performed using the *minfi* R package[20] using the *bumphunterEngine* and *blockFinder* functions, respectively, each with 1000 linear model bootstrap iterations[18] and a cutoff of 0.1 (corresponding to contiguous probes with a minimum 10% directionally consistent change in DNAm associated with birth, for example, $|\widehat{\beta_i}| > 0.1$), with other parameters in these *minfi* functions set to their default values.

CGDMRs were obtained from Lister *et al.*[16] which were relative to genome build hg18. We used the corresponding hg18-based probe coordinates from the Illumina 450k manifest file to perform the overlaps between the microarray and WGBS data at the CpG and DMR level, and then used the liftOver tool to map from hg19 to hg18 at the block level (since 450k probes are collapsed to probe groups and given an average coordinate)[50].

Secondary analyses for age-of-onset differentially methylated CpGs were fit using the same model above within all control samples greater than 10 years of age, comparing those samples younger than and older than 25 years (replacing the *Fetal* term in the model).

Analyses of chromatin state data. The 18-chromatin state data, derived using hidden Markov models (HMMs), was obtained for sample E073 in the Epigenome Roadmap project[22] (http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html). These data were derived from a mixture of DLPFC tissue from two subjects (75 and 81 years old). The chromatin states overlapping DMRs, blocks, and meQTL SNPs (by genomic coordinates, chr:start-end) were obtained, and compared to a background of all considered probe groups, collapsed probe groups, and all considered SNPs, respectively. Overlap was assessed based on the total coverage (in base pairs) of the chromatin states. Fold changes for enrichment and depletion >2 were reported. While the project generated data on two fetal brain samples, processed chromatin state data was currently unavailable. Additional details on the chromatin states are available at the above website.

Statistical analyses for gene expression correlation. Raw gene expression two-color microarray intensity data (available at GSE30272) were loess-normalized as previously described[1]. Probes were re-annotated to the hg19 genome using the Gemma tool[51] leaving 31,699 gene expression probes on 249 samples that had both Illumina 450k DNAm and expression data. Differential expression analysis comparing pre- and post-natal expression levels was performed using limma[47]. We annotated each 450k probe to its nearest gene in the expression data by distance, and computed the Pearson correlation between proportion DNAm and gene expression level, and converted these correlations to Z-scores and corresponding P values $\left(\text{for example, } Z \sim \frac{\rho}{\sqrt{(1-\rho)^2/(N-1)}}\right)$. For the DMR-expression analysis, we matched each DMR to all probes corresponding to the nearest gene, and retained the most correlated DMR-probe correlation. For block-expression analysis, we identified which probes, and their evidence for differential expression, were present in each block using genomic coordinates.

Composition estimation. We implemented *in silico* estimation of the relative proportions of five cell types (ESCs, ES-derived NPCs, and derived dopamine neurons from culture[52], and adult cortex neuronal and non-neuronal cells from

adult tissue[53]) using epigenome-wide DNAm data using a recently published algorithm[54]. All data was obtained using the Illumina HumanMethylation450 (450k) microarray platform from GEO[55]. After normalizing the publicly available data together using the preprocessQuantile function in the minfi Bioconductor package[20], we picked the cell type–discriminating probes as outlined previously[56] resulting in 405 unique probes that distinguished the five cell types. Projecting the brain samples onto these profiles results in a composition proportion for each cell type and sample (**Fig. 2**). We computed the percentage of variance explained (R^2), comparing a linear model with all five composition profiles to an intercept-only model at each Illumina 450k probe and gene expression probe.

We downloaded already-processed data from BrainSpan (http://download.alleninstitute.org/brainspan/Methylation/) and a previous study[24] from GEO (under accession GSE58885) and obtained composition estimates per sample using the same 405 probes and five cell types as above. Comparisons between composition estimates and age and/or brain region were performed using linear regression. Re-analysis of previous data[24] for a main effect of age, adjusting for the composition estimates from the five cell types, was performed using limma[46].

Enrichment for schizophrenia genetic risk. We analyzed the published 108 regions of schizophrenia risk in the latest Psychiatric Genomics Consortium (PGC) genome-wide association study for schizophrenia[25]. We calculated a Chi-squared statistic to determine whether CpGs within the PGC regions differed in their developmental DNAm effects, specifically comparing differential methylation effect sizes from the pre- versus post-natal analysis by whether each probe (of the 456,513) fell within the PGC regions (by genomic coordinates, eg chr:start-end) or not. We performed analogous analyses for the age-of-onset associated CpGs.

Genotype data processing. DNA for genotyping was obtained from the cerebellum of donors in the collection and performed with either the Illumina Human Hap 650v3,1M Duo V3, or Omni 5M BeadArrays as previously described[1]. We had genotype data on 520/526 samples measured on the Illumina 450k. Genotypes were called separately by genotyping platform using the crlmm software[57], and then cleaned separately for imputation (retaining SNPs with MAF > 0.5% and genotyping missing rate < 10%, then checking sex and heterozygosity)[58]. Genotypes were phased into haplotypes using SHAPEIT2 (ref. 59) and imputed in 6MB chunks using Impute2 (ref. 60) to the 1000 Genomes Phase 3 variant set for the autosomes and then Phase 1 variant set for chrX (as Phase 3 data was not available at time of analysis). Imputed genotypes were merged across the three platforms following imputation, and SNPs with MAF > 5%, HWE P value > 10^{-6}, and missing rate < 10% were retained across the 520 samples. LD-pruning generated an independent set of SNPs to perform genome-wide clustering to obtain multidimensional scaling (MDS) components for quantitative measures of ancestry.

meQTL analysis. We reintroduced the probes on the sex chromosomes and those CpGs that had a variant at the CpG site (as meQTLs to these CpGs would represent biological, and not technical, signal) resulting in 477,636 Illumina 450k probes and 7,426,085 common variants on 520 subjects. In the adult control subjects, we modeled the additive effect of genotype (number of minor alleles) on DNAm levels, adjusting for the first five MDS components from the genetic data and the first 11 PCs (based on the 100,000 most inter-individual variable probes for computational efficiency) using the MatrixEQTL package[61]. We allowed for a maximum distance of 20 kb between each SNP and CpG analyzed, resulting in 47,675,913 tests, and we adjusted for multiple testing using a false discovery rate (FDR) threshold of 0.01 to call meQTLs significant. *Post hoc* analyses of the most significant SNP-CpG pair per probe were calculated separately by Caucasians and African Americans, and then within fetal samples.

We downloaded the entire set of GWAS-suggestive and/or significant variants (by rs number) for thousands of traits from the NHGRI GWAS catalog[31]. We computed the odds ratio and corresponding P value for enrichment for GWAS-associated variants by determining which variants were meQTLs and GWAS-associated, meQTLs only, GWAS-positive only, and neither, among those 7,426,085 SNPs considered in our data set. For the PGC2 analysis, we obtained LD-clumped regions from the discovery data set from the May 2013 freeze in the Ricopili tool (PGC_SCZ52_may13) which contains marginally significant regions (down to P value < 10^{-4}) and LD-proxies to each index SNP. We matched these SNPs to our data by chromosome and position, and determined which had an meQTL in our data set.

Schizophrenia differential methylation analysis. We modeled differences in diagnosis, controlling for age, sex, race and the first four PCs from the negative control probes ultimately among 108 patients with schizophrenia and 136 controls. Sex was not associated with diagnosis, and therefore not included as an adjustment variable ($P = 0.16$; **Supplementary Table 1**). Smoking status was obtained by toxicology, where those positive for smoking tested positive for either nicotine and/or cotinine, and included in the above model to assess confounding by smoking. Similarly, cell composition estimates were included in the above model to assess potential confounding by composition differences between cases and controls. We performed a Chi-squared test (df = 1) to determine whether CpGs showing diagnostic effects were over- or under-enriched for meQTLs.

A **Supplementary Methods Checklist** is available.

45. Lipska, B.K. *et al.* Critical factors in gene expression in postmortem human brain: focus on studies in schizophrenia. *Biol. Psychiatry* **60**, 650–658 (2006).
46. Deep-Soboslay, A. *et al.* Reliability of psychiatric diagnosis in postmortem research. *Biol. Psychiatry* **57**, 96–101 (2005).
47. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 3 (2004).
48. Bland, J.M. & Altman, D.G. Multiple significance tests: the Bonferroni method. *Br. Med. J.* **310**, 170 (1995).
49. Irizarry, R.A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* **18**, 780–790 (2008).
50. Farrell, C.M.. *et al.* Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res.* **42**, D865–D872 (2014).
51. Zoubarev, A. *et al.* Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics* **28**, 2272–2273 (2012).
52. Kim, M. *et al.* Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum. Mol. Genet.* **23**, 657–667 (2014).
53. Guintivano, J., Aryee, M.J. & Kaminsky, Z.A. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* **8**, 290–302 (2013).
54. Houseman, E.A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
55. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
56. Jaffe, A.E. & Irizarry, R.A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15**, R31 (2014).
57. Scharpf, R.B., Irizarry, R.A., Ritchie, M.E., Carvalho, B. & Ruczinski, I. Using the R Package crlmm for genotyping and copy number estimation. *J. Stat. Softw.* **40**, 1–32 (2011).
58. Anderson, C.A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).
59. Williams, A.L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. *Am. J. Hum. Genet.* **91**, 238–251 (2012).
60. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
61. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).

doi:10.1038/nn.4181