HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2015 July 01.

Published in final edited form as:

Nat Neurosci. 2015 January ; 18(1): 154–161. doi:10.1038/nn.3898.

Developmental regulation of human cortex transcription and its clinical relevance at base resolution

Andrew E. Jaffe[1,2,3,*], **Jooheon Shin**[1], **Leonardo Collado-Torres**[1,2], **Jeffrey T. Leek**[2,4], **Ran Tao**[1], **Chao Li**[1], **Yuan Gao**[1], **Yankai Jia**[1], **Brady J. Maher**[1,5,6], **Thomas M. Hyde**[1,5,6,7,8], **Joel E. Kleinman**[#1], and **Daniel R. Weinberger**[#1,4,5,6,7,*]

[1]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore MD 21205

[2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD 21205

[3]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore MD 21205

[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore MD 21205

[5]Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore MD 21205

[6]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore MD 21205

[7]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21205

[8]Department of Biological Sciences, Johns Hopkins School of Medicine, Baltimore, MD

[#] These authors contributed equally to this work.

Abstract

[*]To whom correspondence should be addressed: Andrew E. Jaffe 855 N Wolfe St, Ste 300 Baltimore, MD 21205 1-443-287-6864 andrew.jaffe@libd.org Daniel R. Weinberger 855 N Wolfe St, Ste 300 Baltimore, MD 21205 1-410-955-1000 drweinberger@libd.org.

Data Availability: BioProject PRJNA245228; the Track Hub is currently available from: https://s3.amazonaws.com/DLPFC_n36/humanDLPFC/hub.txt. Gene RPKM values are available in Data S1. R code from analyses is available in Data S2.

Author Contributions: All authors contributed to the writing of the manuscript, plus the following individual contributions:
AEJ: designed the study, performed data analyses on summarized DERs - BrainSpan, mouse, cell/tissue types, histone tail- and disease-associated enrichments, and cell composition
JS: performed data analysis involving processing the RNAseq data
LCT: performed data analysis involving the initial global derfinder approach
JTL: performed data analysis involving the initial global derfinder approach
RT: performed RNA extractions and cytosolic separations
CL: performed RNA extractions and cytosolic separations
YG: created sequencing libraries and oversaw the data generation for the discovery data
YJ: created sequencing libraries and oversaw the data generation for the validation data
BJM: assisted in the biological interpretation of the computational findings
TMH: provided brain tissue and demographic data, assisted in biological interpretation of the computational findings
JEK: oversaw the project, provided brain tissue and demographic data, assisted in biological interpretation of the computational findings
DRW: designed the project, oversaw the project, assisted in biological interpretation of the computational findings.

Transcriptome analysis of human brain provides fundamental insight about development and disease, but largely relies on existing annotation. We sequenced transcriptomes of 72 prefrontal cortex samples across six life stages, and identified 50,650 differentially expression regions (DERs) associated with developmental and aging, agnostic of annotation. While many DERs annotated to non-exonic sequence (41.1%), most were similarly regulated in cytosolic mRNA extracted from independent samples. The DERs were developmentally conserved across 16 brain regions and within the developing mouse cortex, and were expressed in diverse cell and tissue types. The DERs were further enriched for active chromatin marks and clinical risk for neurodevelopmental disorders like schizophrenia. Lastly, we demonstrate quantitatively that these DERs associate with a changing neuronal phenotype related to differentiation and maturation. These data highlight conserved molecular signatures of transcriptional dynamics across brain development, some potential clinical relevance and the incomplete annotation of the human brain transcriptome.

Keywords

gene expression; brain development; postmortem human brain; RNA sequencing

Introduction

The transcriptome of the human brain changes dramatically across development and aging, with the largest gene expression changes occurring during fetal life, tapering into infancy[1,2]. Developmental brain disorders often involve genes that are differentially expressed in fetal compared with postnatal life[3,4]. While exploration of the brain transcriptome has been an important approach to understanding brain development and brain disease, previous transcriptome characterizations have used primarily microarray technologies based on probe sequences that capture only a limited proportion of transcriptome diversity. The technological advances of RNA sequencing (RNA-seq) now permit a flexible and potentially unbiased characterization of the transcriptome at high resolution and coverage[5]. Yet, existing published RNA-seq-based characterizations of brain development have utilized gene- and/or exon-level count-based summarizations[4,6,7], which require an accurate and complete gene annotation. Such feature-based read counts lack the ability to reliably identify novel transcriptional activity, but generally limit the inherent difficulty in transcript assembly and characterization based on short read sequencing technologies [8].

We have implemented a method for RNA-seq analysis at single base resolution to more fully characterize transcription dynamics, which leverages the benefits of both count- and transcript-based methods. We describe herein the results of deep coverage sequencing of the polyA+ transcriptomes of human dorsolateral prefrontal cortex (DLPFC) samples across 6 important life stages – fetal (2nd trimester), infant, child, teen, adult and elderly (Table S1) – and implemented an annotation-agnostic differential expression analysis to leverage the power of RNA-seq without the difficulties of transcript assembly [9]. This method, called *derfinder*, identifies differential expression at base-pair resolution, and forms differentially expressed regions (DERs) by joining adjacent differentially expressed bases. We tested for differences in average expression levels across the six age groups and used statistical

permutation to calculate a measure of genome-wide significance for each DER [10]. A DER represents a differentially expressed unspliced segment of RNA (here, across age) that can originate from a full-length, or potentially spliced, transcript. The *derfinder* approach therefore interrogates transcript-level changes in gene expression via differentially expressed segments using only coverage-level RNA-seq data. This approach allows an unconstrained and unbiased search of the transcriptome to identify fragments of interest for more detailed molecular characterization of corresponding full length transcripts.

After application of this approach to a discovery dataset of 36 brain samples, we carried forward DERs that had significant differential expression in a replication dataset of an additional 36 DLPFC samples. Significant and replicated DERs were mapped onto existing reference transcriptomes in databases such as Ensembl [11], UCSC [12], and Gencode [13] to characterize their locations in the genome. We further explored the expression levels within DERs to a wide range of publicly available resources, including RNA-seq data from 16 human brain regions [14], the developing mouse cortex [15], and a variety of other cell [16] and tissue [17] types to understand these patterns in a broader context (summarized in Figure 1). Lastly, we identify significant enrichment for functional epigenomic marks associated with gene expression and for disease-associated genetic loci from recent GWAS. The results highlight conserved signatures of gene expression across development and aging in the human brain, including many non-exonic sequences that appear to be mature mRNAs, and identify biological fingerprints of age-associated changes in neuronal phenotypes and CNS disorder associated genes.

Results

Extensive transcriptional changes across brain development

We identified 50,650 DERs associated with development and aging that were both genome-wide significant in our discovery dataset (at FWER ≤ 5%) and were also differentially expressed in a second independent sample of 36 human brains distributed across the same age ranges (at $p < 0.05$, see Methods, Table S1). These DERs represent 8.63 megabases (Mb) of expressed sequence (Table S2), annotated to 5,985 unique RefSeq (and overlapped 6,549 unique Ensembl) genes. There were, on average, 7.51 DERs annotated to each RefSeq gene (median = 4, IQR: 2–10) – only 1,454 genes contained a single DER (24.3%).

The RefSeq genes containing DERs were strongly enriched for many general developmental and metabolic processes including organelle organization (GO:0006996, 976/2368 genes, $p=7.13\times10^{-29}$), regulation of gene expression (GO:0010468, 1314/3442 genes, $p=8.62\times10^{-23}$) and regulation of transcription, DNA-dependent (GO:0006355, 1127/2916 genes, $p=3.78\times10^{-21}$) (Table S3A). A more focused gene ontology analysis using the 1000 most significant DERs revealed more specific enrichment for neuron projection morphogenesis (GO:0048812, 49/575 genes, $p=4.98\times10^{-11}$), neuron development (GO:0048666, 61/838 genes, $p=1.29\times10^{-10}$), axonogenesis (GO:0007409, 43/509 genes, $p=1.08\times10^{-9}$) and nervous system development (GO:0007399, 100/1784 genes, $p=3.84\times10^{-10}$, Table S3B).

The majority of the DERs have highest expression levels (adjusted for sequencing depth) in the fetal developmental period (N=41,405; 81.7%), followed by adolescent (N=3,104; 6.1%) and adult expression levels (N=2,621; 5.2%). The genes containing DERs most highly expressed from infancy through adulthood are consistently enriched for synaptic transmission (GO:0007268; p-value range: 5.0×10^{-12}-5.5×10^{-24}), cell-cell signaling (GO: 0007267; p-value range: 4.0×10^{-7}-1.7×10^{-17}), and other related signaling processes (Supplementary Tables 3D–G). Interestingly, genes containing DERs most expressed in later life (50+) were not enriched for these signaling processes, and instead were enriched for processes related to cellular respiration and energy-related processes (Table S3H).

Principal component analysis (PCA) of the normalized coverage estimates across the 50,650 DERs revealed that the first principal component (PC) represents a linear scaling (either positive or negative) of expression across the lifespan (72% of variance explained, Figure S1A). The second and third PCs explain lesser variance (combined 15.1%), and represent dynamic expression from infanthood to adolescence with relatively similar levels of expression in fetal life and adulthood (Figure S1B,C). However, almost all DERs had much higher correlation to the first PC (49,698; 98.1%) than the second or third PCs (605 and 346, representing 1.2% and 0.7%, respectively), suggesting that most DERs represent “scaling” of gene expression, i.e. one-directional change, across the lifespan.

Several of the genes containing the most significant DERs showed patterns consistent with the canonical biology of brain development (see Figure S2). These include the high expression of previously identified developmentally significant genes during fetal life, such as *SOX11* (also shown in Figure 1) which encodes a transcription factor involved in the regulation of embryonic development [18] and *DCX*, which is involved in the migration and organization of neuroblasts [19]. Expression of *SLC6A1* (*GAT1*), a sodium/chloride dependent GABA transporter that removes GABA from the synaptic cleft, follows the well-studied early developmental expression of the Gabaergic system [20]. DERs overlapping *NRGN* and *CAMK2A*, two calcium binding proteins important for learning and memory and neuropsychiatric disorders [21,22], become most highly expressed in infant and teenage life periods, respectively. Interestingly, several DERs that have the highest expression during postnatal life have been implicated in brain disorders thought to be developmental, including *RGS4*, a G-protein signaling regulator associated with schizophrenia [23] that has highest expression during adolescence, and *CNTNAP1*, a contactin-associated protein associated with autism [24] with highest expression during adulthood.

Many of the genes associated with DERs also showed developmental regulation across the lifespan using previously published microarray data on 269 non-psychiatric individuals [1] (obtained from GSE30272, see Methods), which highlights both confirmation of the developmentally-regulated genes identified with the DERs and the gains made from using sequencing-based approaches over microarrays. Notably, many of individuals in the present RNA-seq study discovery dataset (N=28/36) were interrogated in this array-based dataset. Most (4,955/5,985 (82.8%)) of the DER-associated genes were present in the processed microarray data and almost all of these genes were differentially expressed across the lifespan: 4,920 (99.3%), 4,684 (94.5%), and 4,304 (86.9%) were significant at $p < 0.05$, $p <$ 10–6, and $p <$ 10–11, respectively. Of the 1,030 genes showing significant differential

expression only in the RNA-seq data, 432 genes were removed during QC steps performed in Colantuoni et al [1] (suggesting they may be more difficult to measure using oligonucleotide probes), and the remaining 598 genes were not included on the microarray design. These genes did not differ in functionality from those included on the microarray (all GO enrichment p-values > 10–6).

Widespread differential expression of unannotated sequence

Surprisingly, many of the age-associated DERs, while contained within genes, contained expressed sequence annotated as intronic – i.e. 21,033 significant regions (41.5%) overlapped at least one Ensembl-annotated intron (minimum overlap = 20 base pairs, see Methods). Additionally, 4,214 regions (8.3%) do not map to any Ensembl annotated genes (i.e. exonic or intronic regions), which we term “intergenic”; 29,813 regions (58.9%) cross at least one annotated exon (Figure S3). Not surprisingly, the exonic DERs had, on average, much higher expression across all samples than DERs annotating to non-exonic sequence (140.8 normalized reads compared to 14.0 and 8.2 normalized reads for intergenic and intronic DERs, respectively, $p < 10^{-100}$) and were longer (190.3 bp versus 150.4 and 139.4 bps respectively, $p < 10^{-20}$). Nevertheless, of the 3,056 Ensembl genes containing intron-annotated DERs, 1,765 (57.7%) genes contained both intronic and exonic DERs. We note these intronic changes are not likely due to technical artifacts and we observe significant enrichment of lncRNAs in the intergenic DERs (see Supplemental Note). There were similar percentages of overlapping annotated features using the UCSC hg19 knownGene (based on RefSeq) database (19,575 / 6,676 / 26,886 for introns/intergenic/exons, respectively) and Gencode v19 (21,107 / 3,994 / 30,016), further suggesting that the transcriptome contained in commonly accessed databases is quite incomplete, at least across human brain development.

The widespread differential expression across development and age of previously-annotated intronic sequence may be due to an abundance of nuclear pre-mRNA present in the total RNA. We therefore sought to better distinguish pre-mRNA from spliced exonic mRNA by sequencing nuclear and cytosolic preparations from six additional independent brain samples (three fetal and three adult, Table S4). Quantifying the relative concentration of mRNA in the cytosolic and nuclear mRNA fractions provided initial evidence that our differentially expressed regions were present in the cytosol – the mean concentrations of cytosolic to nuclear RNA were 204.0:17.6 (ng/ul; 11.6x) in the fetal samples and 137.0:17.6 (7.7x) in the adult samples, showing that the majority of polyadenylated RNA in total polyadenylated RNA originates from the cytosol. We sequenced each mRNA fraction from each sample to characterize the significant and widespread differential expression observed in the total RNA. The relative $\log_2$ fold changes of expression, comparing fetal to adult levels were highly correlated across total and cytosolic polyA+ mRNA DERs (ρ=0.914), including expression of annotated intronic (ρ=0.664) and intergenic (ρ=0.820) regions (Figure S4). There was especially high concordance in the directionality of the non-exonic fetal versus adult fold changes − 96.4% were directionally consistent overall between cytosolic and total polyA+ mRNA. These results implicate the developmental regulation of a potentially large subset of intron-containing mRNA in the cytosolic fraction of the human frontal cortex.

Age-associated DERs lack regional specificity

We next explored the representation of our age-associated DERs in other brain regions, including other cortical and subcortical nuclei, and cerebellum using publicly available BrainSpan data [14], which included RNA-seq data across prenatal and postnatal developmental periods in 16 brain regions. Our DLPFC-identified DERs show consistent age-related changes across each brain region with little inter-regional variability. The first principal component (PC) of only the BrainSpan normalized mean coverage data across the 50,650 DERs (explaining 59% of the variability) strongly correlates with age, particularly fetal versus post-natal, and not brain region (Figure 2). The second principal component (explaining 8.7% of the variability) strongly correlates with RNA quality (Figure S5), subsequent lesser principal components differentiate the neocortical regions from the subcortical region and cerebellum (Figure S6). Within a secondary PCA on only non-exonic DERs, the first principal component remains age (here explaining 40.6% of the variance, Figure S7). There was also significant correlation between $\log_2$ fold changes comparing fetal samples to adults in our DLPFC dataset and the same fetal versus post-natal comparison within each brain region, including within previously annotated intronic and intergenic sequences (Table 1). We note the high correlations between fetal versus adult comparisons in our DLPFC samples and the BrainSpan DLPFC samples constitute an additional independent validation of our identified DERs, including the non-exonic sequences.

Age-associated DERs are conserved in the mouse cortex

We further examined our DERs, particularly the preponderance of non-exonic expression, by leveraging genetic synteny in mice to validate differential expression using a cross-species approach. We downloaded and renormalized publicly available data from mouse cerebral cortex, comparing E17 (N=4) to adult (N=3) C57BL/6 mice [15] which had previously been interrogated for differences in gene-level expression across development. We lifted over the DERs [12] to the mouse genome (mm10), of which 37,428 mapped (73.9%, average synteny = 88.7%) and 25,372 had an average coverage > 5 reads in at least one sample (22,195, 423, and 2,764 in human-annotated exonic, intergenic, and intronic sequence, respectively), suggesting a subset of these DERs are expressed in the developing mouse cortex. We identified significant correlation between the relative differences in fetal and adult human expression compared to E17 versus adult mouse expression within these syntenic regions (Figure 3, $\rho = 0.771$, $p<10^{-100}$). The magnitude and directionality of the expression fold changes were consistent for many human sequences (directionality concordance = 84.1% overall), including those annotated as intronic and intergenic, suggesting these age-associated DERs represent conserved expression signatures in the mammalian developing brain.

Age-associated DERs expressed in other cells and tissues

We also explored the cell-type specificity of these DERs, and respective intronic and intergenic expression, using publicly available RNA-seq data from human stem cells [16] and somatic adult tissues [17]. After re-aligning and processing these public datasets, we observed that the majority of the DERs had on average > 5 reads in at least one stem cell (86.4%) or tissue (84.0%) type, including non-exonic brain-expressed sequences (75.3% and 67.1% of

non-exonic DER expression in at least one stem cell or tissue group, respectively). Furthermore, 53.3% of all DERs, and 26.5% of non-exonic DERs were expressed in all five stem cell conditions in the dataset with coverage > 5 reads (ES, BMP4-treated ES, then differentiation to mesenchymal, mesendodermal and neural progenitor cells), while only 0.4% of the DERs were expressed in all 16 tissue types (see legend of Figure 4).

We identified global expression similarities of these age-associated DERs (via PCA) between the fetal brain samples, and the stem cell and somatic tissue data (PC1) – notably, it was the postnatal brain samples that appear qualitatively different than the diverse cell and tissue types with respect to these DERs (Figure 4A). While the DERs overlapping intronic and intergenic Ensembl-annotated sequence align with the stem cells in its first PC (Figure 4B), these non-exonic DERs appear most unique to the fetal human brain. We then contrasted these patterns to the clustering of the global transcriptome (based on read counts for all Ensembl-annotated genes, available in Data S1) – here PC1 distinguishes the brain (fetal and post-natal) from non-brain (stem cells and somatic tissues) samples, and PC2 distinguishes developmentally active tissues (fetal brains and stem cells) from somatic postnatal tissues (including postnatal brains, Figure 4C). Gene-level expression patterns across the entire transcriptome highlight tissue specific features, while the DERs target more general developmental transitions. Thus, while the overall transcriptomes of cells at different stages of early differentiation are clearly distinct, the DERs reflect common features of these differentiating cells.

Age-associated DERs overlap open chromatin

We next sought to better characterize the DERs with regard to functionality, using publicly-available histone data on human fetal brain [25]. We downloaded and performed peak calling on ChIP-seq data on six histone tail marks (H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac and H3K9me3) and DNase-seq data in fetal brain [25,26] (see Methods), and calculated the overlap with the DERs (see Methods). There was highly significant overlap (at empirical $p < 10^{-100}$, see Methods for permutation procedure) between the DERs and histone marks associated with active chromatin, including H3K36me3 (OR=13.32), H3K4me1 (OR= 3.00), H3K4me3 (OR=5.66), and H3K9ac (OR=4.82). Notably, approximately half of the exonic (48.9%; 14,582/29,813) and intronic (49.4%; 8,204/16,616) DERs were within 1kb of a significant H3K36me3 peak; a smaller proportion of the intergenic DERs were also within 1kb (22.7%, 960/4,221). There was also significant overlap between open chromatin and the DERs (OR=3.13, via the DNase-seq data). Conversely there was little enrichment for histone marks associated with repression, including H3K27me3 (OR=1.04) and H3K9me3 (OR=1.43). These effects are largely consistent between the DERs annotated to exonic and intronic sequence, and weakened within the DERs annotated to intergenic sequence (Table S5), demonstrating that the DERs largely reside in actively transcribed regions in the human fetal brain.

Age-associated DERs overlap disease-associated loci

We sought to identify potential overlap between the DERs and genetic loci conferring risk for 13 neurodevelopmental disorders, starting with schizophrenia, specifically the 108 genome-wide significant loci from the latest Psychiatric Genomics Consortium genome-

wide association study (GWAS) of over 150,000 subjects [27]. Specifically, 42 loci (of the 108 loci, 38.9%) overlapped at least 1 DER which was statistically significant via permutation analysis (p=0.0013, see Methods, Table 2). Stratifying the list of DERs by annotation class yielded more significant overlap for exonic ($p=1.2\times10^{-4}$) and intronic ($p=2.9\times10^{-4}$) DERs but non-significant overlap for intergenic DERs (p=0.053). These effects represented odds ratios of approximately 2.0 for all, exonic, and intronic DERs, and 1.8 for intergenic DERs (see Methods).

We also assessed the overlap between the genes containing DERs and a series of pre-defined gene sets for other neurodevelopmental disorders, including autism, intellectual disability, and syndromal neurodevelopmental disorders [3]. There was significant enrichment for genes associated with intellectual disability ($p<10^{-4}$), and marginal association with autism (p=0.017, genes in the SFARI database [28]) and genes associated with syndromal neurodevelopmental disorders (p=0.027) – these associations were in line with a previously published report on genes showing differential expression comparing fetal to postnatal life using microarray data [3]. Overall, these results implicate the genes containing DERs as enriched for diverse neurodevelopmental disorders.

Lastly, we conducted several analogous analysis in other disorders not typically associated with neurodevelopment including brain- (Alzheimer's disease, AD, and Parkinson's disease, PD) and non-brain-related disorders (type 2 diabetes, T2D, see Methods), and identified significant overlap with the age-related DERs and PD [29], marginal overlap with AD [30], and no overlap with T2D [31]. Notably, while only a small fraction of DERs were most highly expressed in adult life or later (8.4%), 4/7 AD and 5/11 PD genetic loci overlapped at least one of these DERs that was most highly expressed in adult life or later ($p = 7.19\times10^{-5}$ and 1.01×10^{-4} respectively), in contrast to schizophrenia and other neurodevelopmental syndromes, in which the enrichment was primarily for DERs highly expressed in fetal life.

Fetal brain has the largest fraction of expressed genome

We utilized the coverage-level RNA-seq data in our 36 discovery brain samples to barcode regions of expression within each age group (essentially a one-group generalization of the *derfinder* procedure) regardless of differential expression signal. After normalizing each sample to an 80 million read library size, we identified contiguous regions where the average within-group expression levels were ≥ 5 reads. While we identified a similar number of expressed sequences across the six age groups, the fetal samples had a larger fraction of the genome expressed (approximately 4%) and had the fewest proportion of expressed sequences overlapping Ensembl-annotated exons (Table 3). Surprisingly, each age group had a very similar proportion of all annotated Ensembl exons and introns covered (55–58%). Lastly, we observe that the majority of PGC risk loci associated with schizophrenia [27] contain expressed sequence in the DLPFC, one of the brain regions most consistently implicated in schizophrenia [32]. We observed similar metrics and inference using a threshold of ≥ 10 reads as a sensitivity analysis. Based on these results, we have created a custom UCSC “Track Hub” 33 called “LIBD Human DLPFC Development” which illustrates the coverage-level sequencing data within each age group, (Figure S8). These data

can allow easy visualization of our data integrated with the diverse functionality of the UCSC Genome Browser.

Expression changes across development associate with a changing neuronal phenotype

Changes in gene expression across the lifespan may reflect a combination of changes within individual cellular populations and composition changes of varying cell types in the underlying brain tissue. In particular, a comparison of fetal frontal cortex, which contains predominantly neurons and neuronal precursors, and adult prefrontal cortex, which contains a mixture of neurons and glia, may reflect primarily these changing cell constituents. We, therefore, performed an *in silico* estimation [34] of neuronal, non-neuronal, and progenitor cell composition using DNA methylation (DNAm) data from our brain samples projected onto publicly-available DNAm data derived from cell lines (Table S6), including ES-derived NPCs [35], and adult cortex tissue flow-sorted into neuronal and non-neuronal components using the NeuN antibody [34,36]. These composition estimates (i.e. the relative proportion of each cell type in each brain sample, Figure S9A–C) quantitatively confirm the proliferation of non-neuronal cells across the lifespan ($p=5.56\times10^{-5}$) and the loss of remaining NPCs at birth ($p=6.01\times10^{-17}$).

We then correlated these cell type proportions with the expression levels across individuals within each DER. The majority of DERs were significantly associated with only the NPC relative composition estimate (92.2% of DERS, $p_{bonf} < 0.05$, Figure S9D) and not the NeuN- estimate (1.6% of DERs, $p_{bonf} < 0.05$). Multivariate statistical modeling incorporating both NPC and NeuN- proportions (which are negatively correlated at $\rho = -0.53$) illustrate that the vast majority of DERs associate only with the loss of NPCs (N=43,917), and very few DERs associate only with NeuN- (N=6). These results suggest that the widespread expression changes in human brain [1,2] at birth are more about a changing neuronal phenotype than a rise in non-neuronal cell types, specifically the differentiation of neural precursor and progenitor cells into mature neurons.

Discussion

We have identified widespread changes in the transcriptomes of the developing human prefrontal cortex, typically involving many genes previously implicated in brain development. However, unlike previous characterizations that rely on existing annotation, we observed extensive age-dependent expression of sequences previously annotated as intronic and intergenic in commonly accessed genomic databases (Ensembl, Gencode, and UCSC). The majority of these differentially expressed regions (DERs) are most highly expressed in the fetal brain, and decrease in expression across the lifespan. These developmental expression changes were largely present in cytosolic RNA from independent brain samples, present in 15 additional brain regions across development, conserved across mouse development using synteny, and showed considerable overlap with differentiating neural progenitor cells. We additionally identified significant enrichment for active chromatin marks and genetic risk for schizophrenia and other neurodevelopmental disorders. Our *in silico* data suggest that the majority of these DERs, regardless of annotation (i.e.

exonic, intronic or intergenic), reflect a changing neuronal phenotype, depicting differentiation and maturation across human brain development.

These developmental expression changes at single base resolution complement recent approaches characterizing the entire brain transcriptome within particular age groups, like fetal [7,37] or postnatal [38], for example, comparing expression changes across brain regions [39]. Based on our integration with BrainSpan data, we identified regions that do not appear to be regionally regulated, and rather appear to be generic developmental switches in brain – this is in contrast to those genes recently reported by Pletikos et al [39] as possibly related to regional parcellation. For example, while the majority of the regionally-associated genes in Pletikos et al [39] were expressed in our data based on gene level measures (i.e. RPKM > 1) – 87.0% of adult, 81.3% of fetal, and 88.2% of infant genes – only a smaller subset were present in the DER-overlapping 5,985 RefSeq genes – 44.4% of adult, 38.2% of fetal, and 29.4% of infant regionally-associated genes. In contrast, those genes overlapped by DERs were not likely to be differentially expressed by region – of the 5,985 genes that overlapped DERs, only 5.1% were present in the adult regional association gene list, 16.3% of fetal, and 0.09% of infant. We therefore hypothesize that genes associated with regional specificity are a separate subset from those associated with overall developmental processes, perhaps reflecting developmental changes arising from shifting cellular phenotypes in the latter case and regional changes representing different underlying cellular connectivities in the former.

The significant enrichment between the age-associated DERs and genetic loci associated with schizophrenia offers support for the neurodevelopmental hypothesis of the disorder [40]. The current state of the art GWAS study of schizophrenia, involving over 150,000 subjects, identified 108 independent loci associated with risk for illness, and these loci contain approximately 340 potential gene candidates. Because many of the candidates that map to these loci are likely not participating in the population level association, a more finely grained analysis of the DERs that map to these loci may help eliminate some of the genes in these loci from the candidate list. Still, the mechanisms by which genes associated with schizophrenia lead to the emergence of the clinical syndrome in early adult life have been increasingly linked to early developmental processes involving both prenatal and postnatal factors[40]. Our evidence from the DER analysis supports this assumption. Similar enrichment of DERs was found for gene sets associated with risk for autism, intellectual disability, and various neurodevelopmental encephalopathy syndromes, all of which involve obvious early developmental clinical phenomena, thus supporting further clinical relevance of the DERs we have identified. Interestingly, while there was enrichment between DERs and loci implicated in neurodegenerative disorders, these genomic loci showed greater enrichment for DERs that reflect increased gene expression in adult life rather than fetal life.

While the age-associated DERs identified using a conservative statistical threshold occupy a relatively small proportion of the genome (8.63 Mb, 0.3% of the genome), we observed a much larger proportion of the genome being expressed across all age groups, particularly among fetal samples (121.8 Mb, 4.0%). As there were extensive differences among these proportions (e.g. 4.0% in fetal brain versus 3.1% in adult brain), our *derfinder* approach applied here depended on differential expression across six age groups, rather than focusing on fetal versus non-fetal expression differences, which are widespread [1,2]. We note these

differences in the proportion of genome expressed could result from the more diverse cellular phenotypes in the fetal brain samples, particularly the residual ES and NPC signatures. We ran *derfinder* with especially conservative parameters (e.g. the single base threshold), sacrificing statistical power in exchange for reducing the number false positive DERs, an important distinction given the extent of identified novel transcriptional activity outside of previously defined exonic sequence. The public availability of our data allow for re-analyses with varying statistical thresholds and post hoc tests, particularly within individual genes of interest. We note that our DERs are, by definition, elements of transcripts, and not full mRNAs. The limitations of relatively short sequence read length makes full transcript assembly challenging, but the DERs provide entry points to explore targeted transcript assembly with other methods. We also note that our RNA capture approach using PolyA pulldown has limitations, particularly with respect to uncovering noncoding RNAs, many of which are not polyadenylated, and observable 3' biases.

Future biological experiments may better characterize the functional roles of these DERs, particularly the intronic and intergenic regions. Earlier RNA-seq characterization in commercially-available fetal and adult brain mRNA also identified widespread intronic expression, which was hypothesized to play a role in co-transcriptional splicing [41]. The generation of additional ChIP-seq based functional histone tail marks in fetal brain can potentially generate more specific activity classes [42]. Additionally, translating ribosome affinity purification (TRAP)-based assays may elucidate potential translation of DERs in particular cellular systems. For example, we find preliminary evidence in the mouse genome that at least 15% of the intronic and intergenic DERs (and almost all exonic DERs) are likely incorporated into translated protein products based on one small dataset consisting of exclusively E14.5 mouse forebrain [43]. The “translatomes” from more diverse cell types in human tissue at various stages of development and cell lines may identify additional functional roles of our DLPFC-identified DERs. Similarly, we find little overlap between the DERs and reported lncRNAs from mouse neural stem cells from the subventricular zone [44] (only 2–3% of DERs, regardless of annotation) suggesting that lncRNA databases may be incomplete for human brain and that specialized subpopulations of cells may have unique transcriptomic signatures difficult to ascertain in homogenate tissue.

This study is the first to our knowledge to quantitatively estimate the influence of cellular composition changes on transcriptome dynamics across brain development, particularly when comparing prenatal and postnatal samples. Our results suggest that many reported differences in expression occurring across birth, and their subsequent association/enrichment in brain disorders [4,6] may be driven principally by changing neuronal phenotypes, rather than by the commonly considered rise of non-neuronal cell types. Importantly, the observation that many DERs result from a shifting cellular landscape cannot fully explain the widespread expression of non-exonic sequences, as a subset of these regions are more highly expressed in non-fetal samples. However, further research will better refine the composition profiles in bulk tissue, particularly in the uniform generation of more numerous replicates (e.g. NPCs) and cell types, for example via the Epigenomics Roadmap Project [25].

We anticipate these data, both processed and raw, will be a useful resource for interrogating expression change across the lifespan. Our custom UCSC track hub can be used to visually

identify novel transcriptional activity in candidate genes, and can be integrated with the other functional genomics tracks. The approach taken here explored one specific question of this rich dataset, and our results underscore the complexity of gene expression and cellular differentiation that occurs during brain development and the incomplete nature of current transcriptome annotation.

Online Methods

Postmortem brain samples

Post-mortem human brain tissue was obtained by autopsy primarily from the Offices of the Chief Medical Examiner of the District of Columbia, and of the Commonwealth of Virginia, Northern District, all with informed consent from the legal next of kin (protocol 90-M-0142 approved by the NIMH/NIH Institutional Review Board). Additional postmortem fetal, infant, child and adolescent brain tissue samples were provided by the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders (http://www.BTBank.org) under contracts NO1-HD-4-3368 and NO1-HD-4-3383. The Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland approved the protocol, and the tissue was donated to the Lieber Institute for Brain Development under the terms of a Material Transfer Agreement. Clinical characterization, diagnoses, and macro- and microscopic neuropathological examinations were performed on all samples using a standardized paradigm. Details of tissue acquisition, handling, processing, dissection, clinical characterization, diagnoses, neuropathological examinations, RNA extraction and quality control measures were described previously in Lipska, et al. [45]. The Brain and Tissue Bank cases were handled in a similar fashion (http://medschool.umaryland.edu/BTBank/ProtocolMethods.html). Toxicological analysis was performed on every case and subjects with evidence of macro- or microscopic neuropathology, drug use, alcohol abuse, or psychiatric illness were excluded.

We selected six samples per age group for our discovery dataset, balancing for sex (4 male, 2 female) and RNA integrity number (RIN, mean = 8 per group), as our larger collection of fetal samples typically have higher RNA quality (eg. in Colantuoni, et al. [46]). Additional demographic information for our discovery dataset is available in Table S1. We then selected an additional 36 samples, also consisting of 6 samples across the 6 age groups as above (fetal, infant, child, teen, adult, and >50) to serve as a replication cohort (Table S8).

RNA extraction and sequencing

Post-mortem tissue homogenates of dorsolateral prefrontal cortex grey matter (DLPFC) approximating BA46/9 in postnatal samples and the corresponding region of PFC in fetal samples were obtained from all subjects. Total RNA was extracted from ~100 mg of tissue using the RNeasy kit (Qiagen) according to the manufacturer's protocol. The poly-A containing RNA molecules were purified from 1 μg DNAse treated total RNA and following purification, fragmented into small pieces using divalent cations under elevated temperature. Reverse transcriptase and random primers were used to copy the cleaved RNA fragments into first strand cDNA, and the second strand cDNA was synthesized using DNA Polymerase I and RNaseH. We performed the sequencing library construction using the

TruSeq© RNA Sample Preparation v2 kit by Illumina. Briefly, cDNA fragments undergo an end repair process using T4 DNA polymerase, T4 PNK and Klenow DNA polymerase with the addition of a single `A' base using Klenow exo (3' to 5' exo minus), and then ligated of the Illumina Paired-end (PE) adapters using T4 DNA Ligase. An index/barcode was inserted into Illumina adapters allowing samples to be multiplexed in one lane of a flow cell. These products were then purified and enriched with PCR to create the final cDNA library for high throughput DNA sequencing using an Illumina HiSeq 2000.

RNA sequencing data processing

The Illumina Real Time Analysis (RTA) module performed image analysis, base calling, and the BCL Converter (CASAVA v1.8.2), generating FASTQ files containing the sequencing reads. These reads were aligned to the human genome (UCSC hg19 build) using the spliced-read mapper TopHat (v2.0.4) using the reference transcriptome to initially guide alignment, based on known transcripts of Ensembl Build GRCh37.67 (the "−G" argument in the software) [47]. The total number of aligned reads across the autosomal and sex chromosomes (dropping reads mapping to the mitochondria chromosome) per sample are provided in Table S1.

derfinder analysis

We implemented the *derfinder* pipeline available from http://bioconductor.org/packages/release/bioc/html/derfinder.html on the 36 discovery samples (Table S1) base-level coverage data (the number of reads crossing each base in the genome) was created from the aligned reads (BAM files). The statistical model was fit at every base (after performing coarse filtering to remove bases without at least 5 reads in at least 1 sample):

$$y_{ij} = \alpha_i + \beta_i \boldsymbol{Group_j} + \gamma_i M_j + \varepsilon_{ij} \quad (1)$$

for coverage y_{ij} at base i for sample j, where $\boldsymbol{Group_j}$ is a categorical indicator variable for the six age groups, and M_j is the scaled and log-transformed total number of mapped reads per sample and adjusts for differences in library size between samples. This model is compared to the null model:

$$y_{ij} = \alpha_i + \gamma_i M_j + \varepsilon_{ij} \quad (2)$$

by constructing an F-statistic F_i which are then thresholded across the genome, and contiguous regions above the threshold form candidate differentially expressed regions (DERs), ranked by their area statistic (average F-statistic times region width), described in Jaffe, et al. [48]. We used the per-base cutoff of F=20.509, which corresponded to a per-base p-value < 10^{-8} for our given statistical model and sample size. Empirical p-values were calculated by permuting the age group variable, keeping the coverage and library size fixed, 1000 times, and rerunning the full procedure within each permuted dataset, and recording the null area statistics. R code is available at: https://github.com/lcolladotor/libd_n36. The family-wise error rate (FWER) for each candidate DER was calculated based on the null distribution of the maximum area statistic within each permutation [49]. We note that our initial F-statistic cutoff was quite conservative: 246/1000 permutations did not result in a

single genome-wide F-statistic greater than the threshold. We retained the 63,135 significant DERs at a FWER ≤ 5%.

We then assessed the DERs in an independent but analogous dataset of 36 samples. Average coverage per DER was calculated within each of these replication samples, and then we calculated one F-statistic per DER using Equations 1 and 2 above (where y_{ij} is now the sample-specific average coverage within the DER). We retained DERs that were at least marginally significant ($p < 0.05$) in this replication dataset, yielding 50,560 (80.1%) genome-wide significant DERs that were also differentially expressed in this independent DLPFC dataset, which were used for the analyses described below. Non-replicated DERs, compared to replicated DERs, were narrower (83.0 bp versus 170.3 bp, $p < 10^{-100}$), had smaller areas (mean 2633.9 versus 7034.9, $p<10^{-100}$) (and therefore lower ranks), and also lower coverage (mean 6.6 reads versus 108.7 reads, $p<10^{-100}$).

Gene annotations

We constructed “genomic state” objects for Ensembl version p12, UCSC build hg19 knownGene, and Gencode v19 for rapid annotation of DERs, which, briefly, assigns a single state (exonic, intronic, or intergenic) to each base in the genome based on the gene annotation. For a given base, we prioritize exon > intron > intergenic, such that any exonic sequence in any transcript, even if other transcripts are annotated as intronic, are assigned the “exon” state. Any intronic sequence not overlapping annotated exons are assigned the “intron” state, and the remaining genome is assigned the “intergenic” state. We required 20 base pairs (bp) of overlap between significant DERs and Ensembl annotation to be considered overlapping. The 100 bp mappability/alignability and Encode-excluded tracks were obtained from the UCSC Track Browser (http://genome.ucsc.edu/cgibin/hgTrackUi?hgsid=141011952&g=wgEncodeMapability). LincRNA and miRNA tracks were obtained from the respective UCSC hg19 tracks as implemented in TxDb.Hsapiens.UCSC.hg19.lincRNAsTranscripts [50] and TxDb.Hsapiens.UCSC.hg19.knownGene [51] R/Bioconductor packages. Pseudogenes were identified from the latest PseudoPipe Human Database, version 61 [52].

Technical exploration of widespread differential expression of novel transcriptional activity

RNA-seq data processing and analysis involves a number of well-documented technical biases [53–56], but we found little evidence for the significant DERs originating from technical or computational artifacts. For example, 93.7% of DERs had average alignability/mappability measurements of 100 bp reads greater than 99%, only 61 and 7 regions were in tracks excluded by the Duke site and Data Analysis Center of the Encode project consisting mainly of “BSR/Beta” satellite repeats, respectively, and only 1.9% of regions mapped to known pseudogenes. We did observe evidence of 3' bias in the entire set of DERs mapping within genes (the average proportion of nearest exon number to the total number of exons was 0.65; 1 means the DER was in the last exon, and 0.5 means the DER was in the middle exon), a well-described aspect of polyA RNA-seq [57]. However, there was substantial variability in this exonic location proportion when stratified by gene – 43.8% of genes had a DER before its middle exon (i.e. the minimum exonic proportion was less than 0.5, by gene)

while 52.3% of genes had a DER at the last exon (i.e. the maximum exonic proportion was 1.0, by gene). Analyzing the sequence composition, the introns containing a DER had only an average 1.4-fold enrichment for polyA ($p=1.58\times10^{-3}$) and polyT ($p=8.61\times10^{-5}$) repeats for almost all run lengths beyond 6 bases compared to sequences of introns that do not contain a differentially expressed region, adjusting for intron length. The average GC content of the exonic DERs was significantly higher than the intronic and intergenic DERs (0.492 in exonic compared to 0.454 and 0.449 in intergenic and intronic respectively, $p < 10^{-100}$), although there was a wide range of values (IQR spanned ~0.15 for each annotation category) and the GC content for all three annotation class was higher than the background genome (~0.42, based on the hg19 build). Only 23 regions cross an annotated micro-RNA (miRNA) but each also overlapped an annotated intron or exon, which is an important negative control given our polyA+ RNA library preparation should not capture these short RNAs. Lastly, of the DERs annotated as intergenic by Ensembl, 12.4% cross a known long non-coding RNA (lincRNA, via the TxDb.Hsapiens.UCSC.hg19.lincRNAsTranscripts database [50]), compared to 3.7% of all DERs ($p<10^{-100}$).

Purification of Cytosolic and Nuclear RNA

We separated total RNA into nuclear and cytosolic fractions using the Cytoplasmic and Nuclear RNA Purification Kit by Norgen (Cat# 21000, 37400) following the manufacturer's protocol with an extra step of DNase I treatment in the cytosolic fraction in three independent adult and three independent fetal samples. Sequencing libraries were constructed as above, using the PolyA protocol, which were then sequenced on one lane of an Illumina HiSeq 2000, generating approximately 25M reads per sample. One sample overclustered in the sequencer, generating ~100M reads, but its expression was highly correlated with the expression of other samples of the same type (after adjusting for library size), and was therefore included in downstream analyses; see Figure 4 and Supplementary Figures 8–9. Additional demographic material for these independent validation samples are provided in Table S9.

BrainSpan RNA-seq analysis

Normalized sample-level RNA seq coverage data was obtained in the bigwig file format (http://download.alleninstitute.org/brainspan/MRF_BigWig_Gencode_v10/) and matched to phenotype data indicating the brain region and age of each sample. Mean coverage levels for each sample within each DER were computed, and $\log_2$ fold changes comparing fetal (age < 0) to postnatal (age > 0) samples were calculated within each of the 16 brain regions that had at least 10 individuals (see Table 1). Principal component analysis (PCA) on the $\log_2$(normalized coverage + 1) matrix was visualized in Figure 2 and Supplementary Figures 6 and 7. Spearman correlation was used to compare fetal versus adult coverage in our DLPFC samples to the fetal versus non-fetal coverage within each brain region.

Mouse RNA-seq analysis

We downloaded raw single end 80 bp sequencing reads in the FASTQ file format from Dillman, et al. [58] available from the Sequence Read Archive (SRA) [59] at accession number SRX172890. Reads were aligned to the mouse genome (build mm10) using TopHat (version 2.0.9) [47], first aligning to the reference transcriptome ("-G" option described above).

Significant differentially expressed regions (DERs) identified in the developing human brain (UCSC hg19) were mapped to the mouse genome (UCSC GRCm38/mm10) using the liftOver tool [60] implemented in the "rtracklayer" R/Bioconductor package [61]. Note that single human regions could result in multiple smaller sub-regions during the liftOver process, which were used to extract coverage-level data from the aligned mouse data, rather than the absolute range of the lifted over region. Log_2 fold changes were calculated as $\log_2$(Mean Adjusted Fetal Coverage + 1) − $\log_2$(Mean Adjusted Adult Coverage + 1), where each sample was normalized by the total number of mapped reads (in millions) and then averaged within each age group. Spearman correlations and directionality concordances were calculated for each human-annotated Ensembl feature comparing the fold changes in mouse and human.

Public RNA-seq data processing

We downloaded raw sequencing reads from the Illumina BodyMap project [62] from SRA at accession ERP000546 in the FASTQ file format. Note that each tissue/sample had one replicate sequenced in a paired end configuration (50 bp reads) and another replicate sequenced using single end reads. Paired end reads were therefore treated as single end reads for alignment with TopHat (using the "–G" option as described above) to obtain base-level coverage estimates (which does not use paired end information), resulting in three measurements per tissue replicate. We note that single and paired end replicates clustered together at the DER and gene count level (Figure 4). Additionally, all samples labeled as "16 tissue mixture" had very low alignment rates (range: 16.4%–40.6 %) which were much higher in the single tissue samples (range: 86.5%–96.0%).

We also downloaded 101 bp paired end raw sequencing reads from the UCSC Epigenome Project on differentiating stem cells [63] from SRA at accession SRP000941, which were aligned to the hg19 genome using TopHat as described above.

Cross-tissue analysis

Gene counts for the Lieber Institute post-mortem brain data and publicly available samples data were computed using the featureCounts program [64] using the Ensembl Homo_sapiens.GRCh37.73 gtf file, which were converted into the reads per kilobase per million mapped (RPKM) normalized count. Both raw and normalized coverage estimates (by total mapped reads) were extracted at the significant replicated brain DERs (N=50,560) and the subset of DERs that did not overlap an Ensembl-annotated exon (N= 20,837). Raw coverage counts were used to confirm coverage of > 5 reads across tissue and cell line group means.

Principal component analysis (PCA) was performed on the normalized coverage levels (scaled with $\log_2$ and an offset of 32) of the total set of DERs (Figure 4A) and the subset of DERs that were non-exonic (Figure 4B). PCA was performed on the gene RPKMs (Data S1), scaled with $\log_2$ with an offset of 1 (Figure 4C). Log_2 fold changes were calculated as above for all samples (our brain data and the publicly available data), relative to our adult (ages 20–50) adjusted coverage levels.

We further performed co-expression analyses within the three expression summarizations (individual DERs, the subset of non-exonic DERs, and the overall gene counts) within the combined cell and tissue type data. To better understand the global patterns described in the main text, we computed fold changes for mean adjusted expression levels for each tissue and cell type relative to the mean of the adult (total RNA) brain samples. The pairwise Spearman correlations and concordances (both invariant to scaling) were computed for each cell and tissue type (Figure S10). Notably, there was high correlation (ρ=0.603) and concordance (κ = 0.738) between the fetal brain sample and neural progenitor cell (NPC) fold changes within the DERs (Figure S11) – which was the only non-brain sample with concordance > 70% (other groups with high concordance were infant brain, and then the cytosolic and nuclear fractions of fetal brains) – conversely these fetal brain samples were explicitly discordant with the other somatic non-brain tissues (all relative to adult brain expression levels). These results are consistent with a recent report by Brennand et al (2014), in which NPCs had significantly correlated gene expression levels measured on microarrays to first, and not second, trimester frontal cortex. The combination of these results suggests that cortex-derived DERs may represent a more general early developmentally conserved feature of the transcriptome.

Enrichment with chromatin marks and disease-associated loci

We downloaded the aligned reads (BED files) from the Epigenome Roadmap Project from the following GEO accession numbers: GSM621393, GSM669625, GSM806937, GSM806945, GSM916061, GSM621410, GSM806938, GSM806946, GSM706850, GSM806934, GSM806942, GSM621457, GSM669624, GSM806935, GSM806943, GSM669623, GSM621427, GSM806936, GSM806944, GSM916054, GSM1027328, GSM530651, GSM595913, GSM595920, GSM595922, GSM595923, GSM595926, GSM595928, GSM665804, GSM665819, GSM878650, GSM878651, GSM878652, GSM669944, GSM706851, GSM806948, GSM817243 which were fetal brain epigenomic data from H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac, H3K9me3, ChromatinAccessibility and input. CisGenome was used to call one set of significant peaks, comparing each set of biological replicates per mark to the inputs using the default settings [65]. We tiled the hg19 genome into 1kb bins, dropping bins in the known gaps (centromeres, telomeres, etc), and then counted how many bins overlapped both a DER and ChIP-seq peak, only a DER, only a ChIP-seq peak, or neither. Each mark therefore generated a 2×2 table that summed to the number of genome-wide bins (N=2,861,069), and we computed the odds ratio of each 2×2 table – significance was assessed with a Chi-squared test.

We performed a similar analysis for the PGC2 schizophrenia GWAS results using the chr:start-end of the 108 genomic loci from the Supplementary Table 3 of that manuscript [66]. First we calculated the observed proportion of 108 genomic loci that overlapped at least one DER. Then, we performed permutation analysis to determine if this overlap was statistically significant – for a given permutation, we sampled 108 regions of the same widths from the genome (after removing the gaps as described above). Performing this permutation procedure 100,000 times resulted in 100,000 null overlap proportions. We then calculated an empirical p-value defined as the number of null proportions greater than the observed

proportion. An R package for this analysis is available from GitHub [67]. The observed proportions were based on a list of a) all DERs, b) exonic DERs, c) intronic DERs, and d) intergenic DERs. The odds ratios for enrichment were calculated as above, using 1kb genomic bins, and counting the number of bins that overlapped PGC loci and DERs.

This analogous procedure was performed on genome-wide significant and replicated rs numbers available from main or supplementary tables for Alzheimer's disease [68], Parkinson's disease [69] and type 2 diabetes [70]. For each list of rs numbers, we used the SNAP tool [71] to find all SNPs with $R^2 > 0.6$ in Caucasian 1000 Genomes samples (mirroring the summary statistics from the schizophrenia associations), and then created a linkage disequilibrium-based loci for each index SNP. These loci were lifted over to hg19 and then used to assess the overlap with the significant DERs, both together and stratified by annotated feature.

Lastly, enrichment for disease-associated genes was calculated by first obtaining gene sets for neurodevelopmental gene sets as defined by Birnbaum, et al. [72] directly from its Supplementary Table 1. We computed the proportion of genes in each gene set that contained at least 1 DER, as assessed the significance of these observed proportions using permutation analysis. Specifically, we defined expressed genes using the featureCounts RPKM output (as described above) greater than 1.0, and resampled the same number of genes per gene set from the expressed genes (by symbol). For each permutated gene set, we calculated the proportion of null genes containing at least 1 DER, and we calculated empirical p-values based on 1,000 permutations (as above).

Expressed sequence analysis

Base-level coverage counts per sample were normalized to an 80 million read library size (by dividing by 80M akin to RPKM) to identify contiguous regions above some coverage level that we defined as “expressed”. Average normalized coverage levels were averaged within each age group, and these mean age group coverages were smoothed using a running mean operation with a window size of 7 bases to improve sensitivity and specificity [48,73] by reducing the number of very short “expressed” regions (unlike the multi-group *derfinder* procedure which did not utilize smoothing). These smoothed age group means were thresholded at a coverage level of 5 reads, a threshold that we previously validated using PCR and corresponds roughly to a one sided p-value < 0.05 for a one sample t-test with a sample size of 6, the number of samples per group here. We used a threshold of 10 reads for sensitivity the analyses displayed in Table S6, which complements Table 2.

Track Hub description

The track hub covers the entire genome at base-level resolution, and display by default: (A) the 50,560 significant DERs in a dense visibility, (B) the F-statistic for group differences, with the cutoff used to determine DERs and (C) the mean expression levels across the six samples in each of the six age groups, adjusted for library size (to 80M reads for easier interpretability, and colored to match Figure 1). Additional tracks are available, but hidden by default, consisting of the average adjusted expression within the fetal and infant nuclear and cytosolic mRNA fractions.

Composition analysis using DNA methylation (DNAm) data

We implemented *in silico* estimation of the relative proportions of three cell types (ES-derived NPCs from culture [74], and adult cortex neuronal and non-neuronal cells from tissue [75]) using epigenome-wide DNAm data using a recently published algorithm [76]. All data was obtained using the Illumina HumanMethylation450 ("450k") microarray platform. After normalizing the publicly available data together using the preprocessQuantile function in the minfi Bioconductor package [77], we picked the cell type-discriminating probes as outlined by Jaffe and Irizarry [78] resulting in 227 unique probes that distinguished the three cell types. We then normalized the DNAm data from our 36 discovery samples, and estimated the composition of our samples from the methylation profiles of the homogenate cell types at the 227 probes using non-linear mixed modeling [76]. Composition estimates were regressed against the normalized and $\log_2$ transformed expression levels within each DER across the 36 samples, and we obtained a moderated T-statistic and corresponding p-value [79] for each cell type and DER. The Bonferroni–adjusted p-value was set at 0.05/50,560, or $p < 9.89 \times 10^{-7}$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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 Richard Straub for helpful criticism of the data analyses. This work was supported by the Lieber Institute for Brain Development. A.E.J. was supported by 1R21MH102791 and L.C.T was supported by CONACyT México [351535].

References

1. Colantuoni C, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature. 2011; 478:519–523. doi:10.1038/nature10524. [PubMed: 22031444]
2. Kang HJ, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011; 478:483–489. doi:10.1038/nature10523. [PubMed: 22031440]
3. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression patterns of genes associated with neuropsychiatric disorders. The American journal of psychiatry. 2014; 171:758–767. doi:10.1176/appi.ajp.2014.13111452. [PubMed: 24874100]
4. Gulsuner S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013; 154:518–529. doi:10.1016/j.cell.2013.06.049. [PubMed: 23911319]
5. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. doi:10.1038/nbt.1621. [PubMed: 20436464]
6. Parikshak NN, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013; 155:1008–1021. doi:10.1016/j.cell.2013.10.031. [PubMed: 24267887]
7. Willsey AJ, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013; 155:997–1007. doi:10.1016/j.cell.2013.10.020. [PubMed: 24267886]
8. Steijger T, et al. Assessment of transcript reconstruction methods for RNA-seq. Nature methods. 2013; 10:1177–1184. doi:10.1038/nmeth.2714. [PubMed: 24185837]

9. Frazee AC, Sabunciyan S, Hansen KD, Irizarry RA, Leek JT. Differential expression analysis of RNA-seq data at single-base resolution. Biostatistics. 2014 doi:10.1093/biostatistics/kxt053.
10. Jaffe AE, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International Journal of Epidemiology. 2012; 41:200–209. doi:10.1093/ije/dyr238. [PubMed: 22422453]
11. Flicek P, et al. Ensembl 2014. Nucleic acids research. 2014; 42:D749–755. doi:10.1093/nar/gkt1196. [PubMed: 24316576]
12. Hinrichs AS, et al. The UCSC Genome Browser Database: update 2006. Nucleic acids research. 2006; 34:D590–598. doi:10.1093/nar/gkj144. [PubMed: 16381938]
13. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome research. 2012; 22:1775–1789. doi:10.1101/gr.132159.111. [PubMed: 22955988]
14. BrainSpan. Atlas of the Developing Human Brain. 2011. <http://developinghumanbrain.org>
15. Dillman AA, et al. mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nature neuroscience. 2013; 16:499–506. doi:10.1038/nn.3332. [PubMed: 23416452]
16. Xie W, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153:1134–1148. doi:10.1016/j.cell.2013.04.022. [PubMed: 23664764]
17. Farrell CM, et al. Current status and new features of the Consensus Coding Sequence database. Nucleic acids research. 2014; 42:D865–872. doi:10.1093/nar/gkt1059. [PubMed: 24217909]
18. Wang Y, Lin L, Lai H, Parada LF, Lei L. Transcription factor Sox11 is essential for both embryonic and adult neurogenesis. Developmental dynamics : an official publication of the American Association of Anatomists. 2013; 242:638–653. doi:10.1002/dvdy.23962. [PubMed: 23483698]
19. Curtis MA, et al. Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension. Science. 2007; 315:1243–1249. doi:10.1126/science.1136281. [PubMed: 17303719]
20. Hyde TM, et al. Expression of GABA signaling molecules KCC2, NKCC1, and GAD1 in cortical development and schizophrenia. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011; 31:11088–11095. doi:10.1523/JNEUROSCI.1234-11.2011. [PubMed: 21795557]
21. Frankland PW, O'Brien C, Ohno M, Kirkwood A, Silva AJ. Alpha-CaMKII-dependent plasticity in the cortex is required for permanent memory. Nature. 2001; 411:309–313. doi:10.1038/35077089. [PubMed: 11357133]
22. Krug A, et al. The effect of neurogranin on neural correlates of episodic memory encoding and retrieval. Schizophrenia bulletin. 2013; 39:141–150. doi:10.1093/schbul/sbr076. [PubMed: 21799211]
23. Morris DW, et al. Confirming RGS4 as a susceptibility gene for schizophrenia. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2004; 125B:50–53. doi:10.1002/ajmg.b.20109.
24. Wang K, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. 2009; 459:528–533. doi:10.1038/nature07999. [PubMed: 19404256]
25. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology. 2010; 28:1045–1048. doi:10.1038/nbt1010-1045.
26. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nature biotechnology. 2008; 26:1293–1300. doi:10.1038/nbt.1505.
27. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. doi:10.1038/nature13595. [PubMed: 25056061]
28. Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. Disease models & mechanisms. 2010; 3:133–135. doi:10.1242/dmm.005439. [PubMed: 20212079]
29. Nalls MA, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature genetics. 2014; 46:989–993. doi:10.1038/ng.3043. [PubMed: 25064009]

30. Lambert JC, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013; 45:1452–1458. doi:10.1038/ng.2802. [PubMed: 24162737]
31. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics. 2012; 44:981–990. doi:10.1038/ng.2383. [PubMed: 22885922]
32. Callicott JH, et al. Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. The American journal of psychiatry. 2003; 160:2209–2215. [PubMed: 14638592]
33. Raney BJ, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics. 2014; 30:1003–1005. doi:10.1093/bioinformatics/btt637. [PubMed: 24227676]
34. Houseman EA, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012; 13:86. doi:10.1186/1471-2105-13-86. [PubMed: 22568884]
35. Kim M, et al. Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. Human molecular genetics. 2014; 23:657–667. doi:10.1093/hmg/ddt453. [PubMed: 24087792]
36. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics : official journal of the DNA Methylation Society. 2013; 8:290–302. doi:10.4161/epi.23924.
37. Miller JA, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014; 508:199–206. doi:10.1038/nature13185. [PubMed: 24695229]
38. He Z, Bammann H, Han D, Xie G, Khaitovich P. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. Rna. 2014; 20:1103–1111. doi:10.1261/rna.043075.113. [PubMed: 24847104]
39. Pletikos M, et al. Temporal specification and bilaterality of human neocortical topographic gene expression. Neuron. 2014; 81:321–332. doi:10.1016/j.neuron.2013.11.018. [PubMed: 24373884]
40. Kleinman JE, et al. Genetic neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem human brains. Biological psychiatry. 2011; 69:140–145. doi:10.1016/j.biopsych.2010.10.032. [PubMed: 21183009]
41. Ameur A, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nature structural & molecular biology. 2011; 18:1435–1440. doi:10.1038/nsmb.2143.
42. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. doi:10.1038/nature09906. [PubMed: 21441907]
43. Hupe M, Li MX, Gertow Gillner K, Adams RH, Stenman JM. Evaluation of TRAP-sequencing technology with a versatile conditional mouse model. Nucleic acids research. 2014; 42:e14. doi:10.1093/nar/gkt995. [PubMed: 24165879]
44. Ramos AD, et al. Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell stem cell. 2013; 12:616–628. doi:10.1016/j.stem.2013.03.003. [PubMed: 23583100]
45. Lipska BK, et al. Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. Biol Psychiatry. 2006; 60:650–658. doi:10.1016/j.biopsych.2006.06.019. [PubMed: 16997002]
46. Colantuoni C, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature. 2011; 478:519–523. doi:10.1038/nature10524. [PubMed: 22031444]
47. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. doi:10.1186/gb-2013-14-4-r36. [PubMed: 23618408]
48. Jaffe AE, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International journal of epidemiology. 2012; 41:200–209. doi:10.1093/ije/dyr238. [PubMed: 22422453]
49. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994; 138:963–971. [PubMed: 7851788]

50. Carlson, M. 2014.
51. Carlson, M. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TranscriptDb object(s). 2014.
52. Zhang Z, et al. PseudoPipe: an automated pseudogene identification pipeline. Bioinformatics. 2006; 22:1437–1439. doi:10.1093/bioinformatics/btl116. [PubMed: 16574694]
53. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic acids research. 2010; 38:e131. doi:10.1093/nar/gkq224. [PubMed: 20395217]
54. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic acids research. 2008; 36:e105. doi:10.1093/nar/gkn425. [PubMed: 18660515]
55. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464:768–772. doi:10.1038/nature08872. [PubMed: 20220758]
56. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology. 2011; 12:R22. doi:10.1186/gb-2011-12-3-r22. [PubMed: 21410973]
57. Levin JZ, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nature methods. 2010; 7:709–715. doi:10.1038/nmeth.1491. [PubMed: 20711195]
58. Dillman AA, et al. mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nature neuroscience. 2013; 16:499–506. doi:10.1038/nn.3332. [PubMed: 23416452]
59. Wheeler DL, et al. Database resources of the National Center for Biotechnology Information. Nucleic acids research. 2008; 36:D13–21. doi:10.1093/nar/gkm1000. [PubMed: 18045790]
60. Hinrichs AS, et al. The UCSC Genome Browser Database: update 2006. Nucleic acids research. 2006; 34:D590–598. doi:10.1093/nar/gkj144. [PubMed: 16381938]
61. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009; 25:1841–1842. doi:10.1093/bioinformatics/btp328. [PubMed: 19468054]
62. Farrell CM, et al. Current status and new features of the Consensus Coding Sequence database. Nucleic acids research. 2014; 42:D865–872. doi:10.1093/nar/gkt1059. [PubMed: 24217909]
63. Xie W, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153:1134–1148. doi:10.1016/j.cell.2013.04.022. [PubMed: 23664764]
64. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013 doi:10.1093/bioinformatics/btt656.
65. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nature biotechnology. 2008; 26:1293–1300. doi:10.1038/nbt.1505.
66. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. doi:10.1038/nature13595. [PubMed: 25056061]
67. enrichedRanges: Identify enrichment between two sets of genomic ranges v. 0.0.1. 2014. https://github.com/lcolladotor/enrichedRanges
68. Lambert JC, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013; 45:1452–1458. doi:10.1038/ng.2802. [PubMed: 24162737]
69. Nalls MA, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature genetics. 2014; 46:989–993. doi:10.1038/ng.3043. [PubMed: 25064009]
70. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics. 2012; 44:981–990. doi:10.1038/ng.2383. [PubMed: 22885922]
71. Johnson AD, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24:2938–2939. doi:10.1093/bioinformatics/btn564. [PubMed: 18974171]

72. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression patterns of genes associated with neuropsychiatric disorders. The American journal of psychiatry. 2014; 171:758–767. doi:10.1176/appi.ajp.2014.13111452. [PubMed: 24874100]

73. Aryee MJ, et al. Accurate genome-scale percentage DNA methylation estimates from microarray data. Biostatistics. 2011; 12:197–210. doi:10.1093/biostatistics/kxq055. [PubMed: 20858772]

74. Kim M, et al. Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. Human molecular genetics. 2014; 23:657–667. doi:10.1093/hmg/ddt453. [PubMed: 24087792]

75. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics : official journal of the DNA Methylation Society. 2013; 8:290–302. doi:10.4161/epi.23924.

76. Houseman EA, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012; 13:86. doi:10.1186/1471-2105-13-86. [PubMed: 22568884]

77. MJ A, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. Bioinformatics. 2014 In press.

78. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome biology. 2014; 15:R31. doi:10.1186/gb-2014-15-2-r31. [PubMed: 24495553]

79. Smyth, GK. in Bioinformatics and Computational Biology Solutions Using R and Bioconductor. In: Robert, Gentleman, et al., editors. Ch. Statistics for Biology and Health. Springer; New York: 2005. p. 397-420.

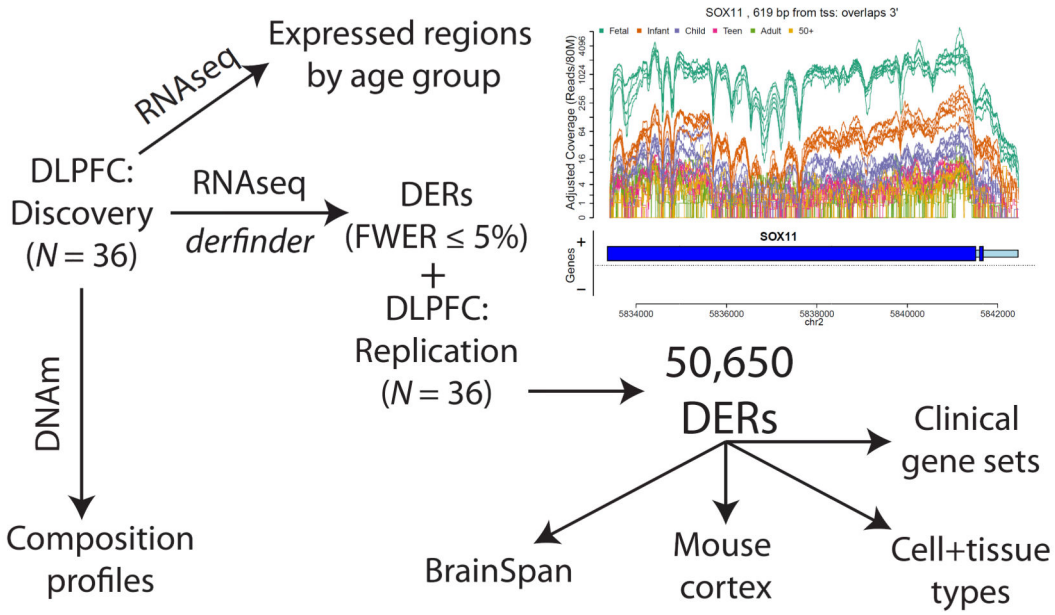

Figure 1.

Schematic design of the project. We performed RNA sequencing (RNA-seq) on 36 DLPFC samples from across the lifespan, and implemented the *derfinder* method to identify "differentially expressed regions" (DERs). These DERs were replicated in an independent DLPFC sample, and explored across other brain regions, in the developing mouse cortex, in diverse cell and tissue types, and in the context of disease-associated gene sets. An example of a DER is shown in the top right corner (see legend of Figure S1 for a detailed description). We additionally quantified the cell composition of these DLPFC samples and defined regions of expression across the genome by age group.

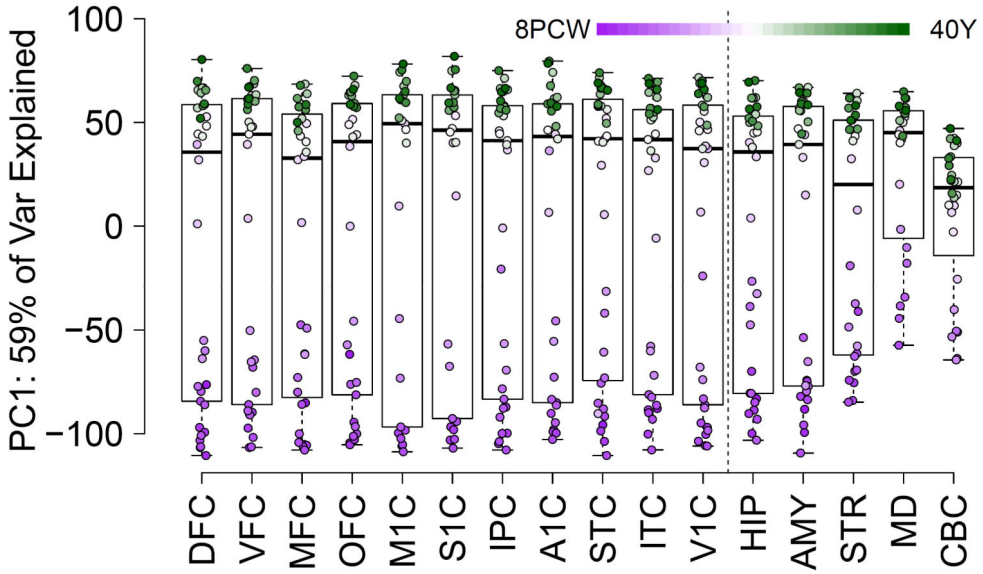

Figure 2.

Age-associated differentially expressed region (DER) expression patterns across multiple brain regions. Principal component analysis (PCA) was performed on normalized coverage estimates across all DERs using all BrainSpan samples. Each point is a sample colored by age (purple: prenatal and green: postnatal), where white corresponds to birth.

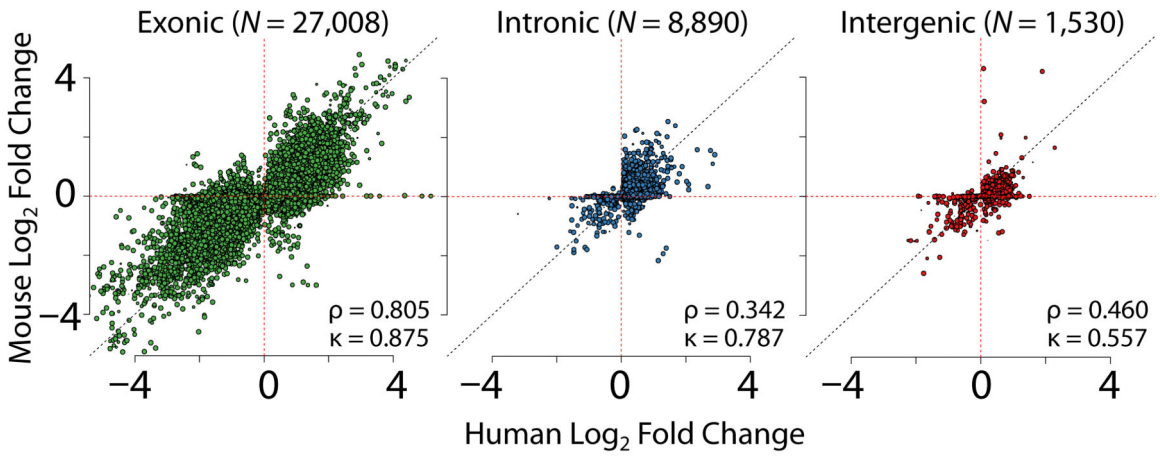

Figure 3.

Cross-species comparison of differentially expressed regions (DERs). Significant DERs were lifted over to the mouse genome mm10 and RNA-seq coverage was extracted from the reprocessed Dillman et al 2013 study comparing E17 to adult C57BL/6 mice. Log_2 fold changes comparing depth-adjusted mean differences between fetal and adult human samples are highly correlated with E17 versus adult mouse samples within each DER, stratified by human-annotated (A) exonic, (B) intronic, and (C) intergenic sequence, such that any DER with both exonic and intronic sequence was classified as exonic. Each point represents a single DER, where the size indicates the proportion of the DERs width that was successfully lifted over. ρ = Spearman correlation, κ = directionality concordance (e.g. higher or lower expression in fetal relative to adult in both species).

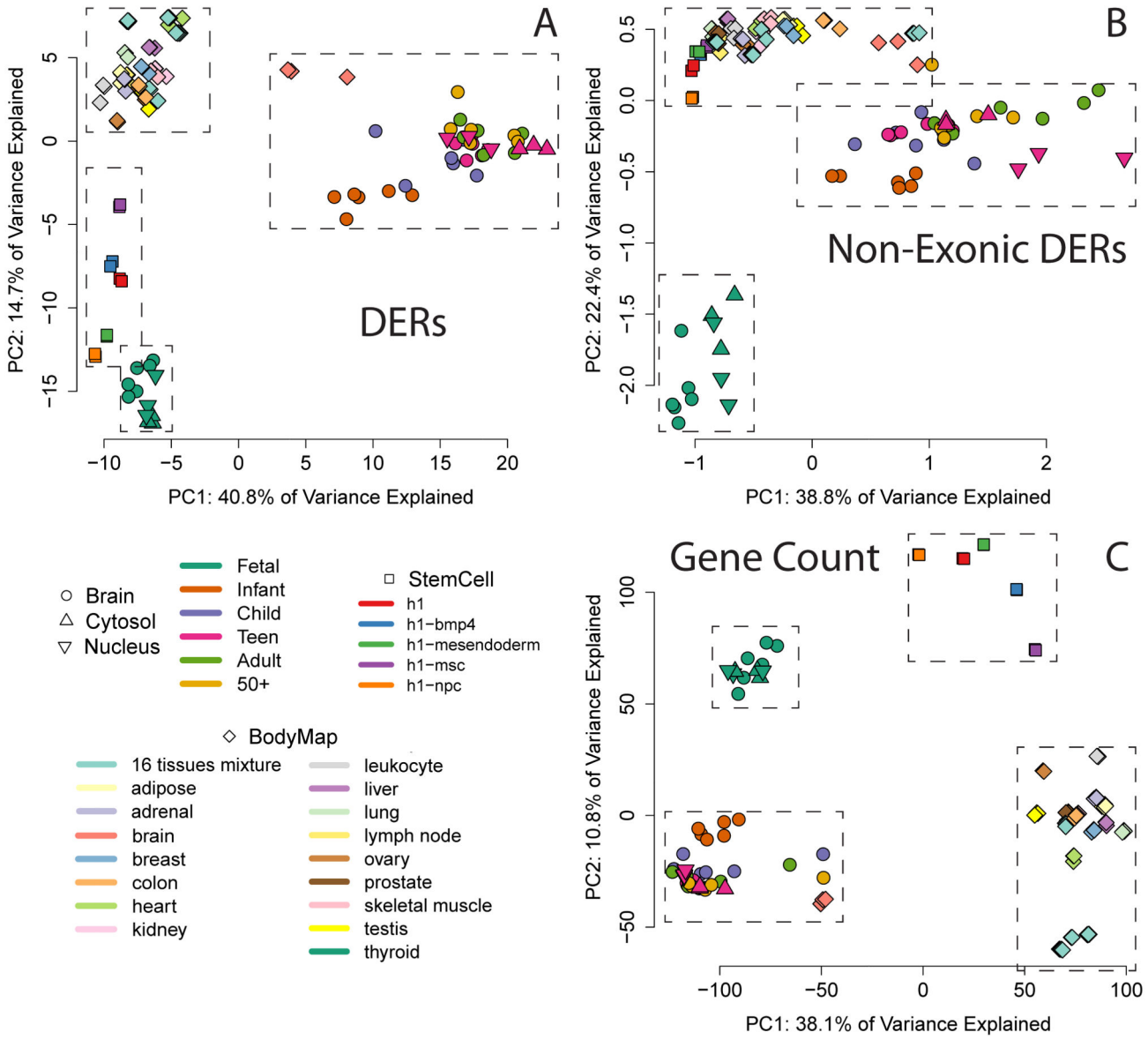

Figure 4.

Clustering analysis of differentially expressed regions (DERs). Principal component analysis (PCA) of (A) all significant DERs, (B) non-exonic sequence within the DERs and (C) gene counts from Ensembl annotation. PCA was performed on $\log_2$ adjusted coverage estimates across multiple datasets including our human brain samples along with publicly available differentiating stem cell and somatic tissue data. Colors and shapes for each point represent dataset and condition (see legend).

Table 1

Correlation of fetal versus adult fold changes across brain regions within differentially expressed regions (DERs]. Spearman correlation coefficients were calculated between $\log_2$ fold changes comparing fetal versus postnatal expression levels within the DLPFC discovery dataset and each brain region in the BrainSpan database across the DERs [All], and within the DERs annotated to specific Ensembl features.

BrainSpan Region	All (N=50,560)	Intragenic (N=4,221)	Intronic (N=16,616)	Exonic (N=29,813)
DFC	0.863	0.702	0.49	0.895
VFC	0.851	0.684	0.429	0.888
MFC	0.858	0.705	0.485	0.891
OFC	0.845	0.674	0.36	0.891
M1C	0.841	0.675	0.388	0.882
S1C	0.83	0.657	0.326	0.878
IPC	0.849	0.681	0.464	0.882
A1C	0.86	0.698	0.517	0.888
STC	0.871	0.72	0.576	0.894
ITC	0.852	0.694	0.473	0.881
V1C	0.867	0.701	0.534	0.894
HIP	0.828	0.66	0.397	0.862
AMY	0.845	0.677	0.444	0.872
STR	0.788	0.607	0.428	0.816
MD	0.699	0.528	0.266	0.731
CBC	0.627	0.434	0.23	0.673

DFC: Dorsolateral prefrontal cortex; VFC: Ventrolateral prefrontal cortex; MFC: Anterior (rostal] cingulate (medial frontal cortex]; OFC: Orbital frontal cortex; MIC: Primary motor cortex (area M1, area 4]; S1C: Primary somatosensort cortex (area S1, areas 3,1,2]; IPC: Posteroinferior (ventral] parietal cortex; A1C: Primary auditory cortex (core]; STC: Posterior (caudal] superior temportal cortex (area Tac]; ITC: Inferolateral temportal cortex (area Tev, area 20]; V1C: Primary visual cortex (striate cortex, area V1/17]; HIP: Hippocampus (hippocampal formation]; AMY: Amygaloid complex; STR: Striatum; MD: Mediodorsal nucleus of thalamus; CBC: Cerebellar cortex.

Table 2

Enrichment of DERs among GWAS-positive regions. Shown are p-values assessing significant overlap between DERs and locations of GWAS-positive loci for schizophrenia, Alzheimer's disease, Parkinson's disease, and type 2 diabetes.

Trait	All	Exon	Intron	Intergenic
Schizophrenia	0.0013	0.0001	0.0003	0.0530
Alzheimer's Disease	0.0385	0.2778	0.0117	0.6016
Parkinson's Disease	0.0039	0.0100	0.0035	0.0882
Type 2 Diabetes	0.2500	0.1029	0.4307	0.1200

Table 3

Expressed sequences/regions by age group defined by 5 or more adjusted reads across consecutive bases (adjusted for library size]. MB: megabases; exonic/intronic/intergenic: the percentages of the expressed regions overlapping annotated features; exons/introns: the converse, being the proportion of all Ensembl features (313,836 unique exons and 266,102 unique introns] covered by expressed sequences in– each age group. “108 PGC2 for SZ” – the number of PGC2 loci overlapping at least 1 expressed sequence in DLPFC. Lastly, we show the percent of expressed regions when defined using 10 or more adjusted reads, as a sensitivity analysis.

	Age Group					
	Fetal	**Infant**	**Child**	**Teen**	**Adult**	**50+**
#of Regions	459,426	481,029	413,202	365,903	437,935	420,294
# in DERs	46,813	37,618	33,958	31,818	32,849	31,563
Coverage (MB)	121.8	107.5	97.1	90.5	92.9	91.4
Genome Covered	4.1%	3.6%	3.2%	3.0%	3.1%	3.0%
Exonic	44.0%	46.8%	54.0%	58.8%	53.1%	54.1%
Intronic	77.1%	72.8%	71.1%	70.2%	69.9%	68.9%
Intergenic	11.9%	13.3%	12.9%	12.5%	12.9%	13.4%
Exons (Ensembl)	55.2%	56.8%	56.9%	55.3%	56.5%	55.8%
Introns (Ensembl)	57.6%	58.1%	57.7%	55.4%	57.2%	56.0%
108 PGC2 for SZ	83	84	83	82	83	88
Intronic ≥10 reads	73.2%	65.6%	64.6%	64.4%	63.7%	62.4%