Identifying age-specific gene signatures of the human cerebral cortex with joint analysis of transcriptomes and functional connectomes

Xingzhong Zhao†, Jingqi Chen†, Peipei Xiao, Jianfeng Feng, Qing Nie and Xing-Ming Zhao

Abstract

The human cerebral cortex undergoes profound structural and functional dynamic variations across the lifespan, whereas the underlying molecular mechanisms remain unclear. Here, with a novel method transcriptome-connectome correlation analysis (TCA), which integrates the brain functional magnetic resonance images and region-specific transcriptomes, we identify age-specific cortex (ASC) gene signatures for adolescence, early adulthood and late adulthood. The ASC gene signatures are significantly correlated with the cortical thickness ($P$-value $< 2.00 \times 10^{-3}$) and myelination ($P$-value $< 1.00 \times 10^{-3}$), two key brain structural features that vary in accordance with brain development. In addition to the molecular underpinning of age-related brain functions, the ASC gene signatures allow delineation of the molecular mechanisms of neuropsychiatric disorders, such as the regulation between ARNT2 and its target gene ETF1 involved in Schizophrenia. We further validate the ASC gene signatures with published gene sets associated with the adult cortex, and confirm the robustness of TCA on other brain image datasets. Availability: All scripts are written in R. Scripts for the TCA method and related statistics result can be freely accessed at https://github.com/Soulnature/TCA. Additional data related to this paper may be requested from the authors.

Key words: brain development; neuropsychiatric disorders; MRI; transcriptome; connectome

Xingzhong Zhao is currently working as a postgraduate student at Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, China. She is interested in bioinformatics and image genetics.

Jingqi Chen is currently working as a young associate professor at Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, China. She mainly studies the genome in neuropsychiatric disorders, by integrating multi-omics data and designing new tools.

Peipei Xiao is a graduate of the Department of Electronic and Information Engineering, Tongji University, China. She is interested in data mining and bioinformatics.

Jianfeng Feng is currently working as a professor with the Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, China. His current research interests include big data analysis and mining for neuroscience and brain diseases.

Qing Nie is currently working as a professor with the Department of Biomedical Engineering, University of California, Irvine, USA. His current research interests include computational systems biology, cell fates, multiscale biology and tools for analyzing single-cell RNA-seq data.

Xing-Ming Zhao is currently working as a professor affiliated with ISTBI, RIICS, Fudan University, and MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, and Frontiers Center for Brain Science, China. His current research interests include data mining and computational systems biology.

Submitted: 23 September 2020; Received (in revised form): 10 November 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
Introduction

The cerebral cortex, making up over three quarters of the human brain volume, plays key roles in the cognition system [1]. During maturation and aging, the cerebral cortex was suggested to exhibit dynamic functional connectivity [2, 3]. For example, with aging, the long-range functional connectivity decreases among medial prefrontal cortex (mPFC), posterior cingulate (pC) and lateral parietal cortex (LP), whereas the connectivity among the left supplementary motor area, the right inferior temporal gyrus and the left temporal pole increase significantly, moreover, this regionally heterogeneous age effects were mainly detected in several functional subnetworks [e.g. default-mode network (DMN)] [3, 4]. The aberrance of either functional connectivity was found to lead to neuropsychiatric disorders [5]. For example, children with autism spectrum disorder (ASD) show reduced short- and long-distance connectivity, especially in default and higher-order visual areas [6].

Much effort has been made to explore the molecular mechanisms underlying the structural or functional variations of the cerebral cortex, and revealed that it is useful and important to interpret the molecular mechanisms of brain complexity through transcriptional architecture [7]. Using transcriptomic analysis, the glucocorticoid receptor gene NR3C1 has been found correlated with cortical thinning during adolescence, indicating its important role in cortical maturation [8]. In addition to being correlated with structural variations, using the correlation between the dynamic patterns of gene expression and spatial topography of group difference in some imaging phenotype could provide molecular mechanisms that underlie functional connectivity of brain cortex, such as during the process of memory encoding [9, 10]. By analyzing gene expressions using Allen Human Brain Atlas, Hawrylycz et al. [11] identified a set of 136 genes whose co-expressions were strongly correlated with the functional connectome built over healthy adolescents, implying their roles of neuron projection and axon guidance. Anderson et al. [12] found strong associations between another set of genes and the limbic functional networks based on transcriptome data, and showed that some of them were enriched in the inhibitory interneuron marker somatostatin receptor pathway and they were potential risk genes for psychiatric disorders. Nevertheless, existing studies often focus on a specific circuitry or an age group, the molecular dynamics or changes across the lifespan for cerebral cortex remain largely unknown.

Here we present a method, transcriptome-connectome correlation analysis (TCA), to identify gene signatures associated with cortex development and functionality across lifespan. By integrating age-specific brain connectomes with gene expression data, we identify gene signatures for three distinct age groups, i.e. adolescence (aged 8–20), early adulthood (aged 20–40) and late adulthood (aged 40–82). We show that these gene signatures can explain age-specific functional dynamics of cerebral cortex during maturation and aging, and may help understand the molecular mechanisms of neuropsychiatric disorders.

Materials and methods

Preprocessing of magnetic resonance imaging data

The functional and structural magnetic resonance imaging (fMRI) and structural MRI data used in this study are downloaded from the Human Connectome Project (HCP) Lifespan Pilot Project (http://lifespan.humanconnectome.org). The dataset contains both functional and structural MRI images covering the whole brain, acquired from 27 subjects (aged 8–75). The subjects are divided into 3 age groups (8–20, 20–40 and 40–75; see details in Table S5). Both the structural and functional MRI images are preprocessed by following the HCP pipelines with the default parameters (http://www.humanconnectome.org/documentation/HCP-pipelines).

Then, we adopt the surface-based cortical parcellations atlas of the Brodmann Area Map (Brodmann lh, colin.R via pals_R-to-fs_LR) to divide the left-hemisphere cortex into 41 spatially contiguous parcels as in the VDG11b parcellation of the Brodmann Area Map [13, 14]. The HCP software, called Workbench Command (https://www.humanconnectome.org/software/workbench-command), is used to extract for each parcel of the Blood oxygenation level dependent (BOLD) time series signal across 420 time points and the structural metrics, including cortical volume, cortical surface area, cortical thickness, curvature, myelination and sulcus.

Preprocessing of gene expression data

The gene expression data for the human cerebral cortex used in this study are downloaded from the NCBI Gene Expression Omnibus under the accession number GSE25219 (Affymetrix Human Exon 1.0 ST Array) [15]. This microarray data contain 17 565 probes across 11 areas of the neocortex (NCX, Table S6), with a total of 1340 cortical and subcortical tissue samples ranging from prenatal stages to late adulthood. We preprocessed the microarray data as described by Kang et al. [15]. First, a stringent criterion (log2-transformed signal intensity ≥6 in at least one sample, and mean detection-above-background P-value ≤0.01 in at least one NCX region during at least one period) is used to define ‘expressed’ probes that results in 12 837 (73.1%) probes for further analysis. Second, we map the 12 837 expressed probes to 17 650 protein-coding genes based on the annotation of Affymetrix Human Exon 1.0 ST Array (as of 18 February 2019). When multiple probes are mapped to the same gene, the mean value of the probes is used as the expression value of the gene.

Finally, we extract the preprocessed gene expression profiles for subjects aged 8–78 (232 samples) and re-group them into 3 age groups (8–20, 20–40 and 40–78), in order to match the MRI datasets.

Known gene signatures

Two gene sets, named as the RichSet [16] and MichSet [11], that have been reported as gene signatures of adult cortex are used for validation in this study. Both gene signatures are defined based on the gene expression data from the Allen Human Brain Atlas of 6 adults (20–60 years old). The RichSet containing 136 genes was generated by integrative analysis of gene expression data with connectome derived from 15 healthy right-handed subjects [16]. The MichSet (75 genes) was generated based on gene expression with HCP S500 connectome data [11], where a subset of the genes whose expression were highly associated with functional connectivity are selected.

Test image datasets

We download the connectome data from IMAGEN (https://imagen-europe.com) and HCP S900 project (www.humanconnectome.org/study/hcp-young-adult). The participants of the IMAGEN project are healthy adolescents aged 14 or 19, as described in Schumann et al. [17]. We use the SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) to process the IMAGEN resting-state functional magnetic resonance (rsfMRI) images with realignment, slice-timing correction, movement...
correction, non-linear warping into MNI space using a custom EPI template and Gaussian-smoothing at 5 mm full-width half-maximum. The HCP S900 project has 900 healthy participants aged from 22 to 35. We preprocess the HCP S900 dataset by following the HCP pipelines with default parameters (http://www.humanconnectome.org/documentation/HCP-pipelines). The parcellation of the brain is carried out as described above. We then extract the 11 cortical regions of left-hemisphere used in this study with the mean value of voxels in a ROI as the value for the region of interest (ROI).

Connectome–transcriptome correlation analysis
First, we design a score GS\text{score}, to describe the stability of gene expression among individuals. For each gene in each age group, the GS\text{score} is defined as the average of the Pearson’s correlation coefficients (PCC) between all pairs of subjects in this age group based on their gene expression profiles over 11 cortical regions as follows.

\[
GS_{\text{score}} = \frac{\sum r_{xy}}{n^2 - n}
\]

where \( \sum r_{xy} \) is the summary of pairwise PCCs between subjects \( (x, y; x \neq y) \) for a specific gene in each age group, \( n \) is the total number of subjects within this age group, and \( n^2 - n \) is the number of combinations counted for each gene at each age group. For each age group, we rank all the genes based on their GS\text{score}, and select the top 10\% genes as the stable genes for further analysis. Second, for each stable gene in each age group, we construct a gene network where each node is a cortical region, and the weight accompanying each edge is the normalized co-expression correlation coefficient between a pair of regions across all the subjects in the age group as defined below.

\[
z_{xy} = \frac{1}{2} \ln \left( \frac{1 + r_{mk}}{1 - r_{mk}} \right)
\]

where \( r_{mk} \) is the co-expression correlation coefficient between cortical region \( m \) and \( k \). Third, for each subject, a functional connectome is constructed from the BOLD time series data across 420 time points, where each node is a cortical region and the weight of each edge is the normalized correlation coefficient between a pair of cortex regions as defined in Equation (2). Then, the average of the connectomes of all subjects in each age group is used as the resultant functional connectome for the age group. Finally, for each age group, a list of genes whose connectivity are significantly correlated (\( P\)-value \( \leq 0.05 \)) with the functional connectome of the same age group are kept for further analysis. To avoid false positives, the permutation test is adopted by randomly permuting sample labels, and the above-mentioned computation process is run for 1000 times. For each gene, the \( P\)-value is defined as the ratio of the number of times among 1000 times that the gene has a higher correlation coefficient with the functional connectome compared with the original one. Consequently, for each age group, the list of genes with \( P\)-value \( < 0.05 \) is defined as the age-specific cortex (ASC) gene signature for the corresponding age group.

Risk genes of neuropsychiatric disorders
The attention-deficit hyperactivity disorder (ADHD) related genes are acquired from the ADHDgene database (http://adhd.psych.ac.cn/), and we select the genes supported by at least 60\% of all studies included in the database [18]. The ASD-related genes are downloaded from the AutDB database (http://autism.mindspec.org/autdb/) and a latest published work [19, 20]. Alzheimer’s disease related genes used in this paper are downloaded from the ALZGene database (http://www.alzgene.org/) [21]. Schizophrenia risk genes are downloaded from the SZGene database (http://www.szgene.org/) and Wang et al. [22, 23]. Bipolar risk genes are downloaded from DisGeNet [24]. Parkinson’s disease associated genes are downloaded from the PDGene database website (http://www.pdgene.org) [25]. Depression associated genes are downloaded from the website (http://www.polygenicpathways.co.uk/depression.htm). All these risk genes can be found in https://github.com/Soulnature/TCA.

Gene function annotation
The functional enrichment analysis on ASC genes is performed using the ToppGene portal (https://toppgene.cchmc.org/, latest update on 21 August 2019) [26]. The ToppGene portal incorporates a comprehensive list of gene function annotation databases, and the \( P\)-value for a query gene list is obtained with a hypergeometric test, with various choices for multi-test correction. From the databases, we select Gene Ontology (GO) annotations (biological process, cellular component and molecular function), pathway annotations (KEEG and REACTOME) and disease gene annotations (OMIM and GWAS), for functional enrichment analysis of the ASC gene sets.

Differentially expressed genes
One-way analysis of variance (ANOVA) is adopted to identify the differentially expressed (DE) genes among the 3 age groups, and do the post-hoc analysis by Tukey’s test to select gene sets that are specific in one group.

R functions used in the analysis
The analysis is performed in R with the following default functions: Pearson correlation coefficient—corr(); ANOVA— aov(); Tukey’s test—TukeyHSD(); Fisher’s exact test—fisher.test(); Principal component analysis (PCA)—prcomp(); nonlinear fitting for BOLD time series signals across all time points—loss(). Functional similarity between two sets of genes is carried out with functions from the R package ‘GOSemSim’ [27].

Identifying the potential neuropsychiatric disorders
mechanism of transcription factors
Neuropsychiatric disorders associated single nucleotide polymorphisms (SNPs, \( P\)-value \( \leq 5\times10^{-6} \)) are collected from the summary statistics of Genome-Wide Association Studies (GWAS) hosted in the PGC data portal (http://www.med.unc.edu/pgc/) and the GWAS Catalog database (https://www.ebi.ac.uk/gwas/). We use the online tool LiftOver (http://genome.ucsc.edu/cgi-bin/hg/hgLiftOver) to convert SNP locations from genome versions hg18 and hg19 to genome version hg38. Next, we download the genome annotation file from NCBI (hg38; ftp://ftp.ncbi.nih.gov/ genomes/H_sapiens/GFF/), and extract the gene position information. We download the motif files of all human transcription factors (TFs) from Jaspar (http://jaspar.genereg.net/) [28], and use FIMO (http://meme-suite.org/tools/fimo) to get TF binding motif information across the whole genome. Furthermore, we use BEDTools [29] to match the locations of risk SNPs, gene promoter regions (\( \pm 10\) KB), and the transcription factor binding motifs.
As shown in Figure 1, we presented an approach, TCA (connectome–transcriptome correlation analysis), to identify the ASC genes based on the integration of fMRI connectomes and gene connectivity maps that derived from the fMRI and transcriptome data across multiple cortical regions for each subject. The mean values of the connectomes and gene connectivity maps for each age group were used to perform the correlation analysis (section ‘Materials and methods’). As a result, 3 sets of ASC genes were identified consisting of 155, 114 and 136 genes for adolescence (aged 8–20), early adulthood (aged 20–40) and late adulthood (aged 40–82), respectively (Tables S1 and S2). Most of them (90.5%) only showed up in one signature, indicating the age-specificity of these signatures (Figure S1). One gene signature, MYLK (encoding myosin light chain kinase), was observed in all the three age groups. MYLK has been reported to induce retraction of mature oligodendrocyte processes, and this process plays an important role in myelin formation and maturation in the central nervous system [30].

By examining their expression dynamics, we noticed that the signatures demonstrated brain region-specific and age-specific expression patterns. Compared with other genes in the genome, the genes from the three signatures were significantly highly expressed (T-test P-values = 1.96e-10, 2.55e-10 and 4.97e-9 for adolescence, early adulthood and late adulthood, respectively), where the ASC genes for early adulthood were most highly expressed, followed by those for adolescence and for late adulthood (Figures 2A and S2). We then identified the brain regions in which each signature was highly expressed, and referred them as ‘highlighted regions’ for the corresponding age group (Figure 2B, section ‘Materials and methods’). Three brain regions that were found to be highlighted for adolescence, including dorsolateral prefrontal cortex (DFC), primary somatosensory cortex (S1C) and ventrolateral prefrontal cortex (VFC). Previously, both DFC and VFC have been reported to be involved in the process of ‘executive’ functions (e.g. working memory and response inhibition) that were highly demanded during adolescence [31, 32]. For the early adulthood group, the highlighted brain regions include MFC, orbital prefrontal cortex (OFC) and inferior temporal cortex (ITC) that were reported to be correlated with cognitive, behavioral and emotional control [33]. For the late adulthood group, the functional decline in some basic cognitive areas of the frontal lobe has been reported [34]. Consistent with this prior knowledge, the regions VFC, MFC, ITC and OFC were highlighted by the ASC genes for late adulthood, and these brain regions were known to be responsible for memory, information processing and the ability for language [35–37].

Together, the identified age-specific gene signatures showed specific spatiotemporal patterns, highlighting brain regions that are important for brain functions during the corresponding age periods.

Age-specific gene signatures characterize age-relevant cortical functions

Next, we investigated the biological functions of the three ASC gene signatures based on the annotations from Gene Ontology (GO, http://geneontology.org). We studied a selection of significantly enriched biological processes for each age group (Figure 2C, FDR ≤ 0.05, section ‘Materials and methods’). As expected, some basic functions important for neuron development and cortex structure were enriched across all three age groups, e.g. chemical synaptic transmission, signal release and calcium ion binding. For each age group, the enriched functions of the corresponding ASC genes were generally consistent with prior knowledge of cortex development and functionality around that age period (Figure 2C). For example, the ASC adolescent genes were enriched with the processes related to neuron and synapse development, e.g. neurotransmitter receptor activity (GO:0030594) and learning or memory (GO:0007611), which were obviously important for the adolescent brain. Specifically, the ASC adolescent genes PPFBP1 and ARL15 have been reported to participate in the guidance and development of axons, and GRIA1, a subunit of the AMPA receptor, have been found to be associated with mature synaptic function [38, 39]. The ASC genes for early
adulthood were enriched with processes involving motion control and cell differentiation, e.g. limbic system development (GO:0005157) and cerebral cortex cell migration (GO:0021795), and those for late adulthood were enriched with immune cells and inflammation-related biological processes such as macrophage colony-stimulating factor receptor (GO:0005157).
Figure 3. Age-specific cortical structural features significantly associated with the ASC genes. Correlations between the expression levels of the ASC genes and cortical thickness (top) and myelination (bottom) were calculated as the Pearson correlation coefficient ($r^2$) between the first principal component (1st PC) of the expression of the ASC genes and the cortical structural measures (for thickness or myelination) for each age group. $P$ denotes the $P$-value for the Pearson’s correlation test.

Conventionally, age-specific genes have been identified with differential expression analysis between distinct age groups [15, 40, 41]. Compared with genes differentially expressed between the three age groups considered here, some of the ASC genes were also differentially expressed genes (DEGs) (20.6%, 6.1% and 15.4% for adolescence, early adulthood and late adulthood, respectively; Figure 2E). The functional enrichment analysis shows that both DEGs and ASC genes were enriched in common neuronal functions, e.g. synaptic signaling (GO:0099536). However, the ASC genes not differential expressed were found enriched with age-specific advanced cognitive functions that were missed by DEGs (Figure 2D). For example, the non-DE ASC genes for adolescence were enriched with learning, memory and cognition, whereas the non-DE ASC genes for early adulthood were enriched with forebrain development. Therefore, our identified age-specific gene signatures were able to characterize the ASC functions better than DEGs.

Age-specific gene signatures are implicated in cortical structural dynamics

In previous studies, it has been indicated that morphological features of the cerebral cortex varied with age [42]. For example, the cortical volume and thickness would change as cortex matures from childhood to adolescence, and these dynamic structural variations were closely associated with cognitive development [43, 44]. Here, with the age-specific signatures, we wanted to see whether these ASC genes can provide an extra layer of explanation for age-related brain structural dynamics. For this purpose, we extracted structural features from structure MRIs of the same samples used in this study, including cortical volume, cortical surface area, cortical thickness, curvature and myelination. With the first principal component of the expression profile of each ASC gene signature, we investigated the correlation between each pair of structural features and ASC signature across the 11 cortical regions. As shown in Figure 3, the expression of ASC genes was only significantly correlated with cortical thickness ($P$-values ranging from 3.00e-4 to 2.00e-3) and myelination ($P$-values ranging from 2.00e-4 to 1.00e-3), both of which were important for cortical development and cognitive abilities [45, 46]. Take GRIA1 as an example, it was one of the ASC genes for adolescence, and its expression was significantly correlated with cortical thickness ($P$-value = 8.14e-4). GRIA1 encoded a subunit of the AMPA-selective glutamate receptor 1, whereas the metabolism of glutamate has been reported to play important roles in cortical thickness [47]. Furthermore, we wanted to
check whether our ASC genes overlap with gene sets previously reported to be associated with brain structural features. For example, a set of 62 genes have been reported significantly associated with cortical volume based on their expression, and we found that our ASC genes were significantly enriched in (P-value = 9.46e-8, Fisher’s exact test) and highly functionally similar to (GOSemSim [27] = 0.86) those 62 genes [48]. These findings indicate that our gene signatures may provide molecular insights into structural variations during cortical maturation and aging.

Age-specific gene signatures play central roles in age-specific brain transcriptional-regulatory networks

Recently, Pearl et al. [49] have constructed a brain-specific transcriptional-regulatory network (TRN) by integrating sequence features and DNase footprinting data of the brain, further filtered by gene expression information in multiple brain regions. The full network consisted of 741 TFs and 11,092 target genes, where 17 of our ASC genes were found to be among the 741 TFs. Taking advantage of this brain TRN, for each age group, we first extracted a sub-TRN where the expression levels of each pair of nodes (a TF and a target gene) connected by an edge were significantly correlated in that age group (Pearson correlation test, FDR ≤ 0.05). As a result, the sub-TRN for adolescence consisted of 479 TFs and 9137 target genes, the sub-TRN for early adulthood consisted of 477 TFs and 8953 target genes, and the sub-TRN for late adulthood consisted of 480 TFs and 8941 target genes. In the three age-specific sub-TRNs, we examined the network-based properties of our ASC genes. Interestingly, in all three sub-TRNs, the degrees of ASC genes were significantly higher than other genes (P-values < 2e-16, Figure S3), indicating that the ASC genes might have more central transcriptional regulatory roles in the brain. For example, NFIB, which was an ASC TF for adolescence and regulated 60 target genes in the sub-TRN of adolescence, has been reported essential for late fetal forebrain development and loss of NFIB gene leads to defects in basilar pons formation and hippocampus development [50]. On the other hand, the target genes regulated by ASC genes in those sub-TRNs were significantly enriched in genes associated with neuropsychiatric disorders (Figure S4). For adolescence, the target genes were significantly enriched in ASDs (P-value = 1.66e-4) and schizophrenia (SCZ, P-value = 3.12e-4). The target genes of early adulthood were significantly enriched in depression (P-value = 4.79e-3) and bipolar (P-value = 3.15e-5). For late adulthood, the target genes were not significantly enriched in any disorders.

We analyzed the TFs that targeted the ASC genes, especially those which target genes were significantly overrepresented with ASC genes (Fisher’s exact test, P-value ≤ 0.05 after multistest corrections). Five, six and eight TFs were identified for adolescence, early adulthood and late adulthood, respectively (Table S3). Some of the TFs found in adolescence and early adulthood were known to be associated with brain development, such as the SOX gene family [51]. In mouse, Sox2 and Sox3 can help maintain the state of stem cells in the hippocampus [52], Sox21 promotes hippocampal adult neurogenesis [53], and Sox9 was essential for the establishment and maintenance of multipotent neural stem cells in central nervous system development [54].

Together, the ASC genes were found to play central regulatory roles in the brain transcriptional regulatory network, which may explain their important functional roles in cortex maturation and aging.

Age-specific gene signatures can interpret the mechanism of neuropsychiatric disorders

The etiologies of many neuropsychiatric disorders have been reported related to the cerebral cortex, but the underlying molecular mechanisms remain to be uncovered [55, 56]. It has been known that different neuropsychiatric disorders have different onset times—for example, ADHD tends to start in childhood, depression in adulthood and Alzheimer’s disease (AD) and Parkinson’s diseases (PD) in old age. Here, we attempted to use the ASC gene signatures to uncover the relationships between ASC functions and neuropsychiatric disorders. We found that the ASC genes for each age group tended to be enriched in risk genes of neuropsychiatric disorders with typical onset in the age group (Figure S5). Specifically, we find that the ADHD risk genes were significantly enriched in adolescent ASC genes (P-value = 9.00e-3), Depression risk genes were significantly enriched in early adulthood ASC genes (P-value = 2.90e-2) and AD risk genes were significantly enriched in late adulthood ASC genes (P-value = 1.30e-2). In addition, we found that for each neuropsychiatric disorder, the ASC risk genes were often highly expressed in brain regions known to be related to the disorder. For example, 11 AD risk genes were among the ASC genes for late adulthood, and they were highly expressed in the inferior posterior parietal cortex, where the reduction of cortex thickness has been reported in AD patients [57]. Five Depression risk genes among the ASC genes for early adulthood were highly expressed in the medial and orbital prefrontal cortex (MPFC & OFPC), and these regions were responsible for motion and autonomic responses to stress, which were also known to be related to depression [58, 59]. The consistency between ASC signatures and the onset time of neuropsychiatric disorders indicates that our identified age-specific gene signatures can be used to interpret the molecular mechanisms underlying those disorders.

Further to explore how the ASC genes underscore the disease-related molecular details, we used the age-specific brain sub-TRN to extract the ASC-disease sub-TRNs, by requiring that each edge has one node that is an ASC gene and the other one is a disease gene. The resulted ASC-disease sub-TRNs showed topologies that varied with age (Figure 4A–C, Table S4). Among all 3 ASC-disease sub-TRNs, 9 of the 17 ASC TF genes were disease genes, such as POUSF2—a key TF for schizophrenia [49]. We investigated those disease genes regulated by the ASC TFs with an assumption that the dysregulation of disease genes may lead to disorders. By screening the disease genes regulated by ASC TFs, we found two cases where the TF binding sites were disrupted by disease-associated SNPs, a strong evidence supporting the important roles of ASC TFs in diseases. In one case, as shown (Figure 4E), the SNP rs187653, which locates in the binding motif of ARNT2 (an ASC TF) near the transcriptional starting site of the Schizophrenia risk gene ETF1, has been reported to be significantly associated with Schizophrenia with P-value = 3.11e-8 [60]. The ARNT2 gene played important roles in the translational termination process and transcriptional regulation, and has been reported to be associated with verbal fluency, a common neuropsychological deficit in Schizophrenia and Bipolar disorder [61, 62]. This analysis suggested that the dysregulation between ARNT2 and ETF1 genes involved in SZ might be caused by the disruption of binding by the mutation of rs187653. In the other case, the SNP rs7916271 (weakly associated with Bipolar disorder, with P-value = 5.72e-7 [63]) was found to disrupt the binding of TF KLF16 (an ASC TF for late adulthood) through its location at the binding motif of the TF (Figure 4E). In
Figure 4. The role of ASC genes in neuropsychiatric disorders. (A–C) The disease sub-TNs for adolescence, early adulthood and late adulthood, featuring only the regulatory relationships between the ASC genes and neuropsychiatric disease genes. Nodes with black labels are the ASC genes, and nodes with red labels are disease genes. If a node is both an ASC gene and a disease gene, it is labeled in blue. The size of each node represents its degree in the sub network. (D) The ASC genes regulated by their upstream regulator—the TF NR3C1 (in the middle, red labeled, indicating it was a disease gene for ADHD), as well as the other TFs regulated ASC gene, in the sub-TTN for adolescence. The nodes with blue label are the ASC genes, other TFs of the ASC genes that are disease genes for ADHD are labeled in red, whereas others are labeled in black. (E) Disruption of the KLF16 binding motifs by SNP rs7916271, and disruption of the ARNT2 binding motifs by SNP rs187653.

Age-specific gene signatures are validated with independent datasets

In order to validate our ASC gene signatures, we compared our ASC gene signatures with two previously published gene sets for the adult brain—the RichSet [16] and MichSet [11] (section ‘Materials and methods’). The RichSet contains 136 genes significantly correlated with cortex connectome of adults aged between 20 and 60, and the MichSet contains 75 genes defined in the similar way as RichSet for adults aged between 20 and 60. From Table 1, both RichSet and MichSet were significantly enriched in the ASC genes for age groups after adolescence, with P-values of 5.73e-12 and 7.83e-9, respectively (Fisher’s exact test). In addition to analyzing the overlap between gene sets, we examined the functional similarity between gene sets with the assumption that functional similar gene sets may have equivalent functional roles. Our ASC signatures were found to be highly functionally similar to both RichSet and MichSet, and the functional similarities between two gene sets were quantified by GOSemSim [27] (section ‘Materials and methods’).

We further tested the robustness of our TCA method, by replicating the computational process on the same transcriptome but two test image datasets—the resting-state fMRI data from
the IMAGEN project [17] and the HCP S900 dataset [69] (section ‘Materials and methods’). The two datasets were respectively collected from adolescent and early adult subjects (see Table 1).

As a result, 46 genes (the IMAGEN dataset) and 129 genes (the HCP S900 dataset) were obtained. We noticed that these two gene sets were significantly enriched with the union of our ASC signatures for adolescence and early adulthood (P-value = 1.20e-6 and P-value = 2.2e-16). By examining further the functional similarities between the two gene sets with our ASC signatures, we found that both gene sets were functionally similar to the ASC signature with the similarity value larger than 0.70 (by GOSemSim [27]). The analysis demonstrates the robustness of our method, and indicates that TCA can be applied to datasets for other biological systems or contexts.

**Discussion**

Both the transcriptome and the connectome of the human cerebral cortex undergo extensive changes across the whole lifespan. Here, with a novel method named connectome–transcriptome correlation analysis (TCA), we identify ASC gene signatures for three different age groups, i.e., adolescence, early adulthood and late adulthood. These ASC genes are functionally enriched with cortex-specific biological processes, and their expression is significantly correlated with chronological structural changes of the cortex. These genes also play central roles in the transcriptional-regulatory network of the brain, and the disrupted regulations of them may contribute to the etiology of neuropsychiatric disorders. Taken together, through integratively analyzing transcriptome and connectome, we have successfully identified three age-specific gene signatures, which recapitulate the structural and functional dynamics during cortical maturation and aging, and pinpoint the molecular mechanisms underlying neuropsychiatric disorders.

Our study provides new insights regarding developmental functions of the cortex, whereas some ASC genes have been confirmed by published studies. Krienen et al. [70] found 19 genes associated with functions of the cerebral cortex, which were significantly functionally similar to the ASC genes (Fisher’s exact test, P-value = 3.14e-05 and GOSemSim = 0.59). The overlapped genes were mainly for the first age group (8–20 years), and were associated with the visual and somato/motor functions. Wang et al. [71] identified 38 genes related to the fractional amplitude of low-frequency fluctuations of the cortex, and these genes were also functionally significantly similar to the ASC genes (Fisher’s exact test, P-value = 1.78e-13, GOSemSim = 0.81). The overlapped genes mainly participated in some basic pathways for cortex development, e.g., the neurotransmitter release and the activities of excitatory synapses. These comparisons confirm the reliability of the ASC genes, and at the same time show that the ASC genes cover more age stages and are functionally more comprehensive.

Although the ASC gene signatures are identified with functional connectome instead of structural MRIs, the expression of our identified ASC genes is significantly correlated with cortical thickness and myelination, the two brain structural features known to vary with age. In fact, some genes reported to be correlated with brain structure are also contained in our ASC genes. For example, SLIT3 has been reported to be associated with cortical volume [48] and WNT3 with myelination [72], whereas both genes belong to the ASC genes, implying that the ASC genes are indeed correlated with brain structures. Therefore, the ASC gene signatures may help better capture the molecular mechanisms involved with the interplay between brain structure and function.

Recently, a growing body of evidence has suggested that many neuropsychiatric disorders are associated with certain genetic mutations. However, it is largely unclear how these genetic mutations contribute to the disorders. Moreover, these mutations themselves only account for part of the heritability of the neuropsychiatric disorders [73]. Understanding the biology underlying the normal trajectories of cortical maturation and aging, as a result, may provide a global picture for how the disorder-associated mutations fit into the pathological process. Thus, our ASC genes may help interpret the mechanisms of neuropsychiatric disorders in at least the following three ways: (i) given that the ASC genes depict age-specific biological functions of the cortex, the regulatory relationships between the ASC genes and other genes that are disrupted by mutations may explain the functional impact of the mutations, as we have shown in this manuscript; (ii) given the age-specific association between the ASC genes and the cortical functional and structural MRI features, disrupting the regulations of the ASC genes by the disorder-associated mutations can be mapped to a network of regional neuro-circuitries, such as the DMN implicated in multiple neuropsychiatric disorders [74] and (iii) the ASC genes can also be presented as a set of candidate biomarkers whose disruptions may lead to disorder status.

In this work, we only focus on 11 regions of the cerebral cortex due to limited availability of the brain regional transcriptome data. These 11 regions may not precisely depict the functional details of cortex, as they are relatively broad regions. In addition, the relatively small sample size used may reduce the statistical power of our approach. With more and more brain transcriptome and imaging data released in the future, we may be able to obtain results with better reliability and higher resolution. Nevertheless, we have partially validated our results with previously identified gene sets, proving the robustness of our approach to some extent. It is also possible to incorporate other types of brain datasets into the TCA method, so as to gain insights from additional levels.

We are aware that the level of mRNA is not a complete representation of protein level, and the mechanisms of translation control are complicated. Some ASC genes, for example, PCBDI
and ARL15, have been proven relevant for cognitive trajectory at the protein level [75, 76], but for most other ASC genes there have not been much evidence at the protein level. In the future, if proteomics data for matched age groups become available, our method is flexible to accommodate such data, so as to gain a more comprehensive view on cortex development and aging. As such, we present our method and the ASC genes to help understand the mechanisms underlying cortical maturation and aging. We hope our findings will be of value for the study of both the normal and the disease states of the brain.

Key Points

- We propose a novel algorithm, named transcriptome-connectome correlation analysis (TCA), for integrating the brain functional magnetic resonance images and region-specific transcriptomes.
- We apply TCA to identify age-specific cortex (ASC) gene signatures during different age stages. And, we find that the ASC genes are associated with morphological features and age-specific functions of the cortex.
- The results of TCA are robust, and allow delineation of the molecular mechanisms of neuropsychiatric disorders (such as schizophrenia).

Supplementary data

Supplementary data are available online at Briefings in Bioinformatics.

Acknowledgment

We thank Dr Tianye Jia for his suggestions on the differential gene analysis.

Funding

This work was partly supported by National Key R&D Program of China (2020YFA0712403, 2018YFC0910500), National Natural Science Foundation of China (61932008, 61772368), Shanghai Science and Technology Innovation Fund (19511101404) and Shanghai Municipal Science and Technology Major Project (2018SHZDZX01).

Conflict of interest

None declared.

References