

Cortical gene expression architecture links healthy neurodevelopment to the imaging, transcriptomics and genetics of autism and schizophrenia

Received: 1 October 2022

Accepted: 18 March 2024

Published online: 22 April 2024

 Check for updates

Richard Dear¹✉, Konrad Wagstyl², Jakob Seidlitz^{3,4,5}, Ross D. Markello⁶, Aurina Arnatkevičiūtė⁷, Kevin M. Anderson⁸, Richard A. I. Bethlehem¹, Lifespan Brain Chart Consortium*, Armin Raznahan⁹, Edward T. Bullmore¹ & Petra E. Vértes¹

Human brain organization involves the coordinated expression of thousands of genes. For example, the first principal component (C1) of cortical transcription identifies a hierarchy from sensorimotor to association regions. In this study, optimized processing of the Allen Human Brain Atlas revealed two new components of cortical gene expression architecture, C2 and C3, which are distinctively enriched for neuronal, metabolic and immune processes, specific cell types and cytoarchitectonics, and genetic variants associated with intelligence. Using additional datasets (PsychENCODE, Allen Cell Atlas and BrainSpan), we found that C1–C3 represent generalizable transcriptional programs that are coordinated within cells and differentially phased during fetal and postnatal development. Autism spectrum disorder and schizophrenia were specifically associated with C1/C2 and C3, respectively, across neuroimaging, differential expression and genome-wide association studies. Evidence converged especially in support of C3 as a normative transcriptional program for adolescent brain development, which can lead to atypical supragranular cortical connectivity in people at high genetic risk for schizophrenia.

How does the complex anatomical and functional organization of the human brain develop from the expression of over 20,000 genes¹, and how does this process go awry in neurodevelopmental disorders? In the past 10 years, whole-brain, whole-genome transcriptional atlases, such

as the Allen Human Brain Atlas (AHBA)², have suggested that healthy brain organization may depend on ‘transcriptional programs’ representing the coordinated expression of large numbers of genes over development^{3–7}.

¹Department of Psychiatry, University of Cambridge, Cambridge, UK. ²Wellcome Centre for Human Neuroimaging, London, UK. ³Lifespan Brain Institute, Children’s Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, USA. ⁴Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children’s Hospital of Philadelphia, Philadelphia, PA, USA. ⁵Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. ⁶McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada. ⁷Turner Institute for Brain and Mental Health, Monash University, Melbourne, VIC, Australia. ⁸Department of Psychology, Yale University, New Haven, CT, USA. ⁹Section on Developmental Neurogenetics, National Institute of Mental Health, Bethesda, MD, USA. *A full list of members and their affiliations appears in the Supplementary Information.

✉e-mail: raj22@cam.ac.uk

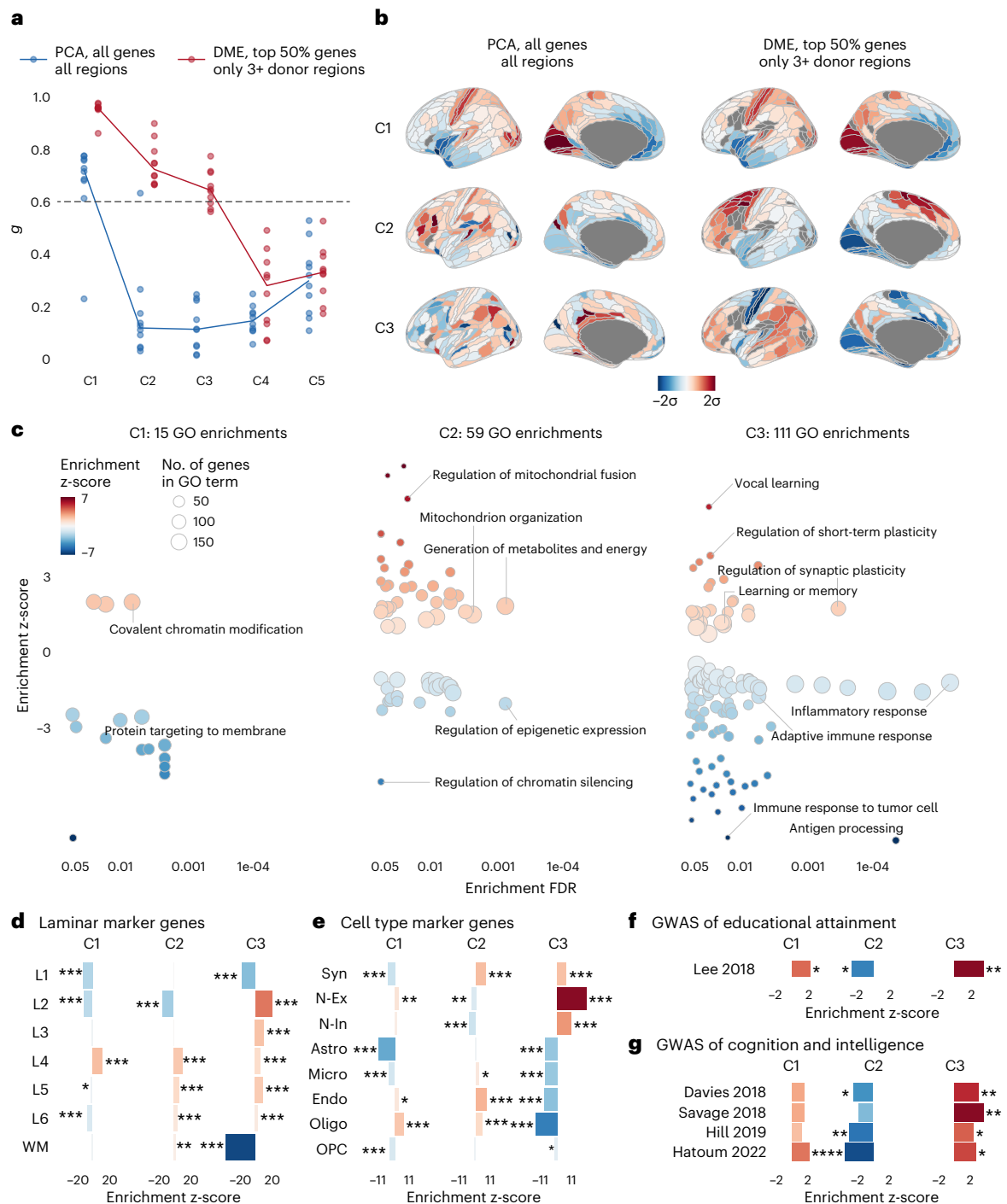


Fig. 1 | Three generalizable components of human cortical gene expression were enriched for biological processes, cytoarchitecture and cognitive capacity.

a, To identify robust components of cortical gene expression, we split the six-brain AHBA dataset into two disjoint triplets of three brains, applied PCA to each triplet and correlated the resulting matched components (C1, C2, C3...) (Methods). For each component, the median absolute correlation over all 10 permutations of triplet pairs was a proxy for its generalizability, g . Using PCA and previously published best practices for processing the AHBA dataset^{33,34}, generalizability decreased markedly beyond the first component: $g_{C1} = 0.78$, $g_{C2} = 0.09$, $g_{C3} = 0.14$. Using DME with the top 50% most stable genes, and the 137 regions with data available from at least three brains, the generalizability of the first three components substantially increased: $g_{C1} = 0.97$, $g_{C2} = 0.72$, $g_{C3} = 0.65$. **b**, Cortical maps of brain regional scores of components C1–C3 estimated by DME on the filtered AHBA dataset displayed smooth spatial gradients (right; Moran's I 0.48, 0.58 and 0.21 for C1–C3, respectively), unlike those of PCA on the unfiltered

data (left; Moran's I 0.50, 0.09 and 0.07). **c**, GO biological process enrichments for C1–C3 showed that the number of significant enrichments was greater for higher-order components, illustrating that they were more biologically specific. C2-positive genes were enriched for metabolism, whereas C2-negative genes were enriched for regulatory processes. C3-positive genes were enriched for synaptic plasticity and learning, whereas C3-negative genes were enriched for immune processes. **d**, C1–C3 were distinctively enriched for marker genes of six cortical layers and white matter (WM)³⁷. **e**, C1–C3 were also distinctively enriched for marker genes of cell types and synapses⁴⁴. **f**, All three components were significantly enriched for genes mapped to common variants associated with educational attainment in previous GWAS data³⁹. **g**, C2 and C3 (but not C1) were significantly enriched for genes mapped to common variation in intelligence and cognition across four independent GWAS studies^{40–43}. For **d–g**, significance was computed by two-sided permutation tests (Methods) and FDR-corrected across all tests in each panel; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

In 2012, Hawrylycz et al.² showed that principal components of the AHBA dataset capture distinct features of cortical anatomy. In 2018, Burt et al. argued that the first principal component of cortical gene expression (C1) reflects an anterior-to-posterior ‘neuronal hierarchy’, defined in macaque tract-tracing data by feedforward and feedback axonal connections between cortical areas^{8–10} and indexed in humans by the ratio of T1-weighted and T2-weighted (T1w/T2w) magnetic resonance imaging (MRI) signals, a putative marker of cortical myelination⁸. These discoveries echoed previous findings from studies of embryonic development of chick, mouse and human brains where spatially patterned transcriptional gradients were shown to organize neurodevelopmental processes, such as areal differentiation, axonal projection and cortical lamination^{6,11–13}. Single-cell RNA sequencing (RNA-seq) data have also revealed an anterior-to-posterior gradient in the gene expression of inhibitory interneurons, which is conserved across multiple species, including humans¹⁴. It is, therefore, likely that the principal component of gene expression in the adult human cortex represents a transcriptional program key to its normative development.

However, it is not clear that C1 is the only component of spatially patterned and neurodevelopmentally coordinated gene expression in the human brain. Hawrylycz et al.² suggested that principal component analysis (PCA) of a restricted set of 1,000 genes in one of the six brains of the AHBA dataset revealed multiple biologically relevant components (Supplementary Fig. 1). Later, Goyal et al.¹⁵ used nonlinear dimension reduction across whole-genome spatial expression, again from only one of the six AHBA brains, to show that aerobic glycolysis was associated with a second transcriptional component. To our knowledge, more recent studies using all available AHBA data have reliably found only C1 (refs. 8,16). This first component has been linked to a general ‘sensorimotor-association axis’ (S-A axis) of brain organization¹⁰ derived from several macroscale brain phenotypes, including, among others, the principal gradient of functional connectivity¹⁷, maps of brain metabolism and blood flow¹⁵ and the map of human cortical expansion compared to other primates¹⁸. Although it is parsimonious to assume that such diverse brain phenotypes could all be determined by a single transcriptional program, it seems more realistic to expect that multiple transcriptional programs are important for human brain development, as is generally the case for brain development in other species¹⁹.

Here we present two higher-order components of human cortical gene expression, C2 and C3, that likely represent additional transcriptional programs distinct from the C1 component already reliably described⁸. These higher-order components emerged only when optimized data-filtering and dimension-reduction methods were applied to the AHBA dataset. We found that C2 and C3 are each specifically enriched for biologically relevant gene sets and spatially co-located with distinct clusters of neuroimaging phenotypes or macroscale brain maps. Leveraging independent RNA-seq datasets on single-cell and developmental gene expression, we further demonstrate that all three components are generalizable to other datasets, representative of coordinated transcription within cells of the same class, and dynamically differentiated over the course of fetal, childhood and adolescent brain development. Finally, by triangulating evidence across case-control neuroimaging, differential gene expression and genome-wide association studies (GWASs), we demonstrate that components C1 and C2 are specifically associated with autism spectrum disorder (ASD) and C3 with schizophrenia. Although previous studies used the AHBA to derive gene sets correlated with disorder-related MRI phenotypes^{20–25}, this disorder-first, ‘imaging transcriptomics’^{26–28} approach is susceptible to identifying genes whose co-location with MRI phenotypes reflects secondary associations or consequences of a disorder, such as behavioral changes (for example, smoking and alcohol use), physical health disorders (for example, obesity and diabetes) or pharmacological treatment^{29–31}. What is of most interest for neurodevelopmental disorders is to understand the pathogenic provenance of a clinically diagnosable disorder—to ask ‘what developed differently?’

rather than merely ‘what is different?’. Our approach sought to distinctively address the question of what ‘develops differently’ based on an understanding of ‘normal development’, by linking genetic risks and atypical phenotypes to a generalizable transcriptional architecture of healthy brain development.

Results

Three components pattern cortical gene expression

We first applied PCA to the entire AHBA dataset of six adult brains². Microarray measurements of relative mRNA levels were processed to represent mean expression of approximately 16,000 genes at each of the 180 regions of the left hemispheric cortex defined by the HCP-MMP parcellation^{32–34} (Methods). We initially found that higher-order components (C2 and C3) estimated by PCA of the resulting {180 × 16,000} data matrix were not robust to sampling variation of the six donor brains, with low generalizability, g , compared to C1: $g_{C1} = 0.78$, $g_{C2} = 0.09$, $g_{C3} = 0.14$ (Methods). However, two data processing improvements were found to enhance the generalizability of higher-order components. First, we optimized the tradeoff involved in excluding noisy data—by filtering spatially inconsistent genes (with low differential stability³⁵) and under-sampled brain regions—while seeking to maximize the anatomic and genomic scope of the data matrix (Extended Data Fig. 1). Second, we used the nonlinear dimension reduction technique of diffusion map embedding (DME), instead of linear PCA, to identify coordinated gene expression patterns from the matrix. DME is robust to noise and more biologically plausible than PCA in this context because of its less strict orthogonality constraints (Methods). We found that, although PCA and DME both identified the same components from the filtered gene expression matrix (Extended Data Fig. 1d), using DME was necessary to achieve high generalizability, g , while also retaining sufficient genes for downstream enrichment analyses.

We applied DME to the {137 × 7,937} filtered AHBA data matrix comprising the expression of the 50% most stable genes measured in the 137 cortical areas with data available from at least three brains. The generalizability of the first three components was substantially increased— $g_{C1} = 0.97$, $g_{C2} = 0.72$, $g_{C3} = 0.65$ —whereas the generalizability of even higher-order components remained low—for example, $g_{C4} = 0.28$ (Fig. 1a). We found that the cortical maps of C2 and C3 derived from DME on filtered data were more spatially smooth than the corresponding PCA-derived maps on unfiltered data (Fig. 1b), consistent with the interpretation that higher generalizability indicates less contamination by spatially random noise. C1–C3 were also robust to variations in parameters for processing the AHBA, including choice of parcellation template (Extended Data Fig. 2). Finally, the transcriptional patterns represented by C1–C3 in the AHBA dataset were reproducible in an independent PsychENCODE dataset comprising bulk RNA-seq measurements of gene expression at 11 cortical regions from $n = 54$ healthy controls³⁶ (regional correlation: $r_{C1} = 0.85$, $r_{C2} = 0.75$, $r_{C3} = 0.73$; Extended Data Fig. 3 and Supplementary Table 5).

The first three DME components, C1–C3, explained 38%, 10% and 6.5%, respectively, of the total variance of the filtered AHBA dataset (Methods). The proportion of variance explained was related to the number of genes that were strongly weighted (absolute correlation $|r| \geq 0.5$) on each component: 4,867 genes (61%) were strongly weighted on C1, 967 genes (12%) on C2 and 437 genes (5.5%) on C3 (Supplementary Fig. 2). The three components also had distinct axial alignments in anatomical space, and the co-expression network of cortical regions displayed clear anatomical structure even when the highest-variance C1 component was regressed out (Extended Data Fig. 4). These findings demonstrate that these three expression patterns shared across hundreds to thousands of genes are likely to be biologically relevant.

To interpret the DME-derived components in more detail, we first used enrichment analyses of the weights of the 7,973 genes on each component (Methods). Many more Gene Ontology (GO) biological process terms were significantly enriched (with false discovery rate

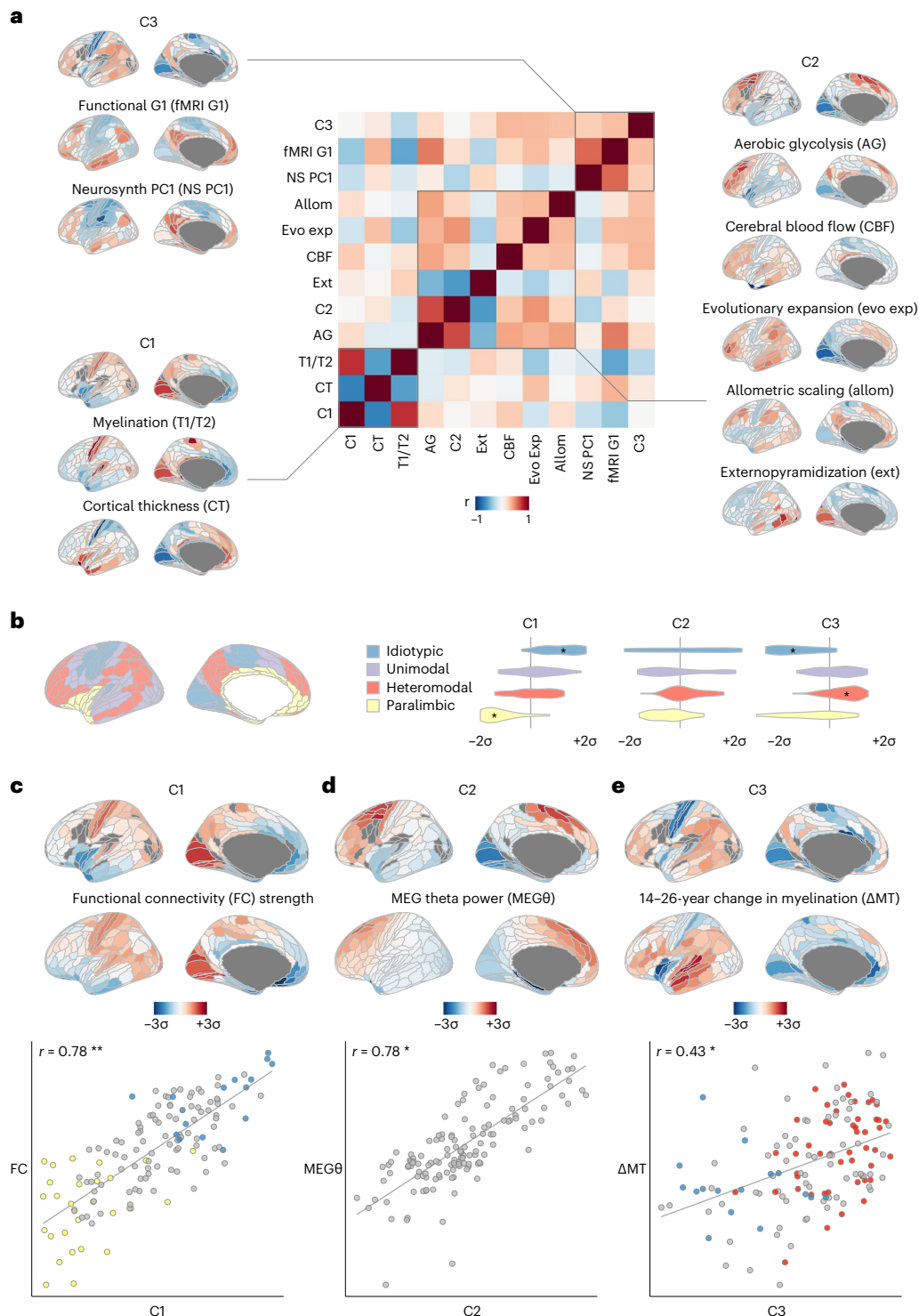


Fig. 2 | Neuroimaging and macroscale maps of brain structure, function and development were distinctively co-located with three components of cortical gene expression. a, Correlation matrix of intrinsic transcriptional components C1–C3 together with the nine neuroimaging-derived and physiologically derived maps that Sydner et al.¹⁰ combined with C1 to define S–A axis of brain organization. Many of the maps were not highly correlated to each other (median $|r| = 0.31$), and data-driven clustering of the matrix revealed three distinct clusters around each of the mutually orthogonal transcriptional components C1–C3, demonstrating that all three components are relevant for understanding macroscale brain organization. **b**, Distributions of regional scores of C1–C3 in histologically defined regions of laminar cytoarchitecture⁵¹. C1 distinguished idiotypic ($P = 0.005$) and paralimbic ($P = 0.002$) regions,

whereas C3 distinguished idiotypic ($P = 0.002$) and heteromodal ($P = 0.01$) regions. $^*P < 0.05$, FDR-adjusted two-sided permutation test as the percentile of the mean z-score relative to null spin permutations, with adjustment for multiple comparisons across all 12 tests. **c**, Degree of fMRI connectivity^{52,53} was significantly aligned to C1 ($r = 0.78$, $P_{\text{spin}} < 0.001$). Blue/yellow highlighted points correspond to idiotypic/paralimbic cytoarchitectural regions as in **b**. **d**, MEG-derived theta power⁵⁴ was significantly aligned to C2 ($r = 0.78$, $P_{\text{spin}} = 0.002$). **e**, Regional change in myelination over adolescence^{55,56} was significantly aligned to C3 ($r = 0.43$, $P_{\text{spin}} = 0.009$). Blue/red highlighted points correspond to idiotypic/heteromodal cytoarchitectural regions as in **b**. In **c** and **d**, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, FDR-corrected two-sided spin-permutation test, with corrections for multiple comparisons of all maps in **c** and **d** being compared with all of C1–C3.

(FDR = 5%) for C2 (59 GO terms) and C3 (111 GO terms) than for C1 (15 GO terms) (Fig. 1c).

Although C1 was enriched for relatively few, functionally general biological processes, it precisely matched the first principal component previously reported ($r = 0.96$)⁸. The same interneuron marker genes (*SST*, *PVALB*, *VIP* and *CCK*) and glutamatergic neuronal genes (*GRIN* and *GABRA*) were strongly weighted with opposite signs (positive or negative) on C1 (Supplementary Fig. 3).

For genes positively weighted on C2, 23 of 36 enrichments were for metabolic processes, and, for negatively weighted genes, 19 of 23 enrichments were for epigenetic processes (Fig. 1c and Supplementary Table 2). Whereas, for genes positively weighted on C3, 19 of 27 enrichments were related to synaptic plasticity or learning, and, for negatively weighted genes, 33 of 84 enrichments involved the immune system. We further analyzed enrichment for genes identified as markers of specific cortical layers³⁷ (Fig. 1e) and cell types³⁸ (Fig. 1f) and, in each case, observed distinct enrichment profiles for C1–C3. For example, genes positively weighted on C3 were enriched for marker genes of neurons, synapses and cortical layers 2 and 3 (L2 and L3), whereas genes negatively weighted on C3 were enriched for glial (especially oligodendroglial) marker genes.

We also explored the biological relevance of the three components by enrichment tests for genes associated with variation in adult cognitive capacity. We found that all three components, C1–C3, were enriched for genes significantly associated with educational attainment³⁹ (Fig. 1f). Across four independent GWASs of intelligence and cognition^{40–43}, genes strongly weighted on C1 were not significantly enriched, but genes negatively weighted on C2 were enriched for genetic variants associated with intelligence in three of the four studies, and genes positively weighted on C3 were enriched for genes identified by all four previous GWASs of intelligence (Fig. 1g).

Neuroimaging maps align to three transcriptional components

Previous work linked gene transcription to a multimodal S-A axis¹⁰ of brain organization, defined as the composite of 10 brain maps, comprising C1 and nine other MRI or positron emission tomography (PET) neuroimaging maps that were selected to differentiate sensorimotor and association cortices. We first aimed to build on this work by analyzing the correlation matrix of the same set of nine brain maps together with the three transcriptional components derived from DME of the filtered AHBA dataset. Data-driven cluster analysis of this {12 × 12} correlation matrix identified three clusters, each including one of the orthogonal transcriptional components (Fig. 2a and Methods). C1 was clustered together with two MRI maps: the T1w/T2w myelination marker⁴⁴ and cortical thickness⁴⁵; C2 was clustered with five maps: aerobic glycolysis⁴⁶, cerebral blood flow⁴⁷, cortical expansion in humans relative to non-human primates¹⁸, inter-areal allometric scaling⁴⁸ and external pyramidal cell density⁴⁹; and C3 was clustered with two maps: the principal gradient of functional MRI (fMRI) connectivity¹⁷ and first principal component of cognitive terms meta-analyzed by Neurosynth⁵⁰. Although some maps were specifically aligned to one component—for example, aerobic glycolysis $r_{C2} = 0.66$ ($P_{\text{spin}} = 0.004$, FDR < 5%)—others were moderately correlated with multiple transcriptional components—for example, for cerebral blood flow: $r_{C1} = 0.25$, $r_{C2} = 0.28$, $r_{C3} = 0.33$. This clustering analysis suggests that it is overly parsimonious to align all nine neuroimaging phenotypes with just one transcriptional component (C1) as part of a singular S-A cortical axis.

We also found that the three transcriptional components were associated with a wider range of cellular, functional and developmental phenotypes than the nine neuroimaging maps above and that these associations were again distinct for the three components. For example, at cellular scale, histologically defined regions of laminar cytoarchitectural differentiation⁵¹ were co-located with C1 and C3 but not with C2 (ANOVA, $P < 0.001$; Fig. 2b). In fMRI and

magnetoencephalography (MEG) data, we found that weighted nodal degree of cortical regions in an fMRI network^{52,53} was strongly correlated with C1 ($r_{C1} = 0.78$, $P_{\text{spin}} < 0.001$, FDR = 5%; Fig. 2c) but not with C2 or C3 ($r_{C2} = -0.01$, $r_{C3} = 0.00$); across all canonical frequency intervals of MEG data⁵⁴, an FDR-significant association was observed between theta band (4–7 Hz) oscillations and C2 ($r_{C2} = 0.78$, $P_{\text{spin}} = 0.002$, FDR = 5%; Fig. 2d) but not C1 or C3 ($r_{C2} = -0.18$, $r_{C3} = -0.02$) (see Extended Data Fig. 5 for other MEG results). In addition, in support of the hypothetical prediction that adult brain transcriptional programs are neurodevelopmentally relevant, we found that a previous map of adolescent cortical myelination, as measured by change in magnetization transfer between 14 years and 24 years (ΔMT)^{55,56}, was significantly co-located with C3 ($r_{C3} = 0.43$, $P_{\text{spin}} = 0.009$; Fig. 2e) but not with C1 or C2 ($r_{C2} = 0.17$, $r_{C3} = 0.15$).

C1–C3 are distinctly developing intracellular programs

We next used two additional RNA-seq datasets to investigate the consistency of AHBA-derived components with gene co-expression in single cells—for example, neurons or glia—and to explore the developmental phasing of gene transcription programs represented by C1–C3.

First, for single-cell RNA-seq data comprising 50,000 nuclei sampled from five cortical regions of three donor brains⁵⁷, the total weighted expression of the C1–C3 gene weights in each sample was computed separately for genes positively and negatively weighted in each component (Methods). We reasoned that if the components derived from bulk tissue microarray measurements in the AHBA dataset were merely reflective of regional differences in cellular composition—for example, neuron–glia ratio—then genes weighted positively and negatively on each component should not have anti-correlated expression across cells of the same class. However, we observed that genes weighted positively and negatively on the same component had strongly anti-correlated expression at the single-cell level (Fig. 3a), whereas genes that were positively and negatively weighted on different components were not anti-correlated (Supplementary Fig. 5). The anti-correlation of genes positively and negatively weighted on C1 or C2 was stronger within each class of cells than across multiple cell classes, and was even stronger when the single-cell data were stratified by subclasses of cells in specific cortical layers—for example, L2 VIP-expressing interneurons (Fig. 3a, inset). By contrast, for C3, the anti-correlation of positively and negatively weighted genes was stronger across cell classes than within each class, although there was still evidence for significantly coupled expression across cells of the same class or subclass.

Second, to explore the developmental trajectories of the transcriptional components, we used BrainSpan, an independent dataset where gene expression was measured by RNA-seq of bulk tissue samples from 4–14 cortical regions for each of 35 donor brains ranging in age from –0.5 years (mid-gestation) to 40 postnatal years⁶. We first asked if the gene weights for each of the components derived from the AHBA dataset would exhibit similar spatial patterns in the BrainSpan dataset. We projected the C1–C3 gene weights from the AHBA onto the subset of adult brains (18–40 years, $n = 8$) in BrainSpan (Fig. 3b and Methods) and found that the resulting cortical maps of component scores in the BrainSpan data were highly correlated with the corresponding cortical maps derived from the AHBA dataset ($r_{C1} = 0.96$, $r_{C2} = 0.88$, $r_{C3} = 0.84$; Fig. 1d). This indicated that the three components defined in the AHBA were generalizable to the adult brains in the BrainSpan dataset (for a full replication of C1–C3 in independent data, see Extended Data Fig. 3). We then similarly compared the cortical component maps derived from the AHBA dataset to the corresponding maps calculated for subsets of the BrainSpan cohort from two earlier developmental stages (prebirth, $n = 20$, and birth to 13 years, $n = 14$). We observed that, for C1 and C2, AHBA component scores were almost as highly correlated with BrainSpan component scores in fetal (prebirth) and childhood (birth to 13 years) brains as in

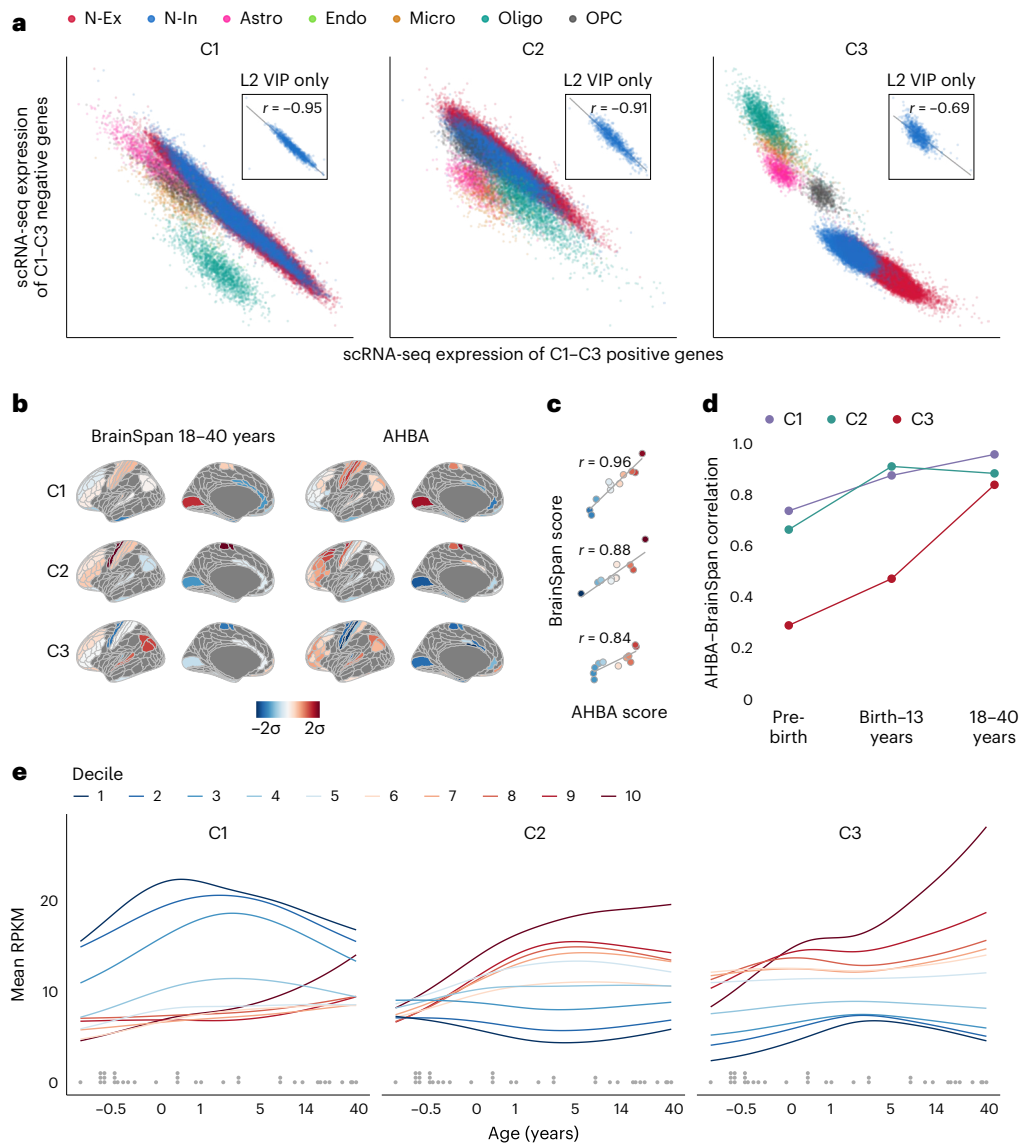


Fig. 3 | Transcriptional components represent intracellular coordination of gene expression programs with distinct developmental trajectories. **a**, For each of approximately 50,000 single-cell RNA-seq samples, the weighted average expression of the negatively weighted genes of each AHBA component C1–C3 is plotted against that of the positively weighted genes (Methods). Samples are colored by cell type, demonstrating that genes positively and negatively weighted on C1–C3 have correlated expression within each major class of brain cells. Astro, astrocytes; Endo, endothelial cells; Micro, microglia; N-Ex, excitatory neurons; N-In, inhibitory neurons; Oligo, oligodendrocytes; OPC, oligodendrocyte precursor cell. Inset, a subset of samples from L2 VIP interneurons, illustrating that C1–C3 weighted genes were transcriptionally coupled even within a fine-grained, homogeneous group of cells. **b**, Cortical maps representing the regional scores of components C1–C3 for each of 11 regions with transcriptional data available in the BrainSpan cohort of adult brains (left) and C1–C3 component scores for the matching subset of regions in the AHBA (right). **c**, Scatter plots of matched regional C1–C3 scores from **b**, demonstrating that the three transcriptional components defined in the AHBA

had consistent spatial expression in BrainSpan. **d**, Correlations between AHBA C1–C3 scores and BrainSpan C1–C3 scores (as in **c**) for each of three age-defined subsets of the BrainSpan dataset. C1 and C2 component scores were strongly correlated between datasets for all age subsets, whereas C3 component scores were strongly correlated between datasets only for the 18–40-year subset of BrainSpan. This indicates that C1 and C2 components were expressed in nearly adult forms from the earliest measured phases of brain development, whereas C3 was not expressed in adult form until after adolescence. **e**, Developmental trajectories of brain gene expression as a function of age (–0.5 years to 40 years; x axis, log scale) were estimated for each gene (Methods) and then averaged within each decile of gene weights for each of C1–C3; fitted lines are color-coded by decile. Genes weighted positively on C3 were most strongly expressed during adolescence, whereas genes weighted strongly on C1 or C2 were most expressed in the first 5 years of life. Dots above the x axis represent the postmortem ages of the donor brains used to compute the curves. RPKM, reads per kilobase per million mapped reads.

the adult (18–40 years) brains (birth to 13 years, $r_{C1} = 0.87$, $r_{C2} = 0.91$; prebirth, $r_{C1} = 0.74$, $r_{C2} = 0.66$; Fig. 3d). However, C3 scores in the AHBA dataset were not so strongly correlated with C3 scores in the fetal and childhood subsets of the BrainSpan dataset (prebirth, $r_{C3} = 0.29$; birth to 13 years, $r_{C3} = 0.47$). These results suggest that C3 may only emerge developmentally during adolescence, whereas the C1 and C2 have nearly-adult expression from the first years of life.

We tested this hypothesis by further analysis of the BrainSpan dataset, modeling the nonlinear developmental trajectories of each gene over the age range of –0.5 years to 40 years (Methods) and then averaging trajectories over all genes in each decile of the distributions of gene weights on each of the three components. We found that genes in the top few deciles of C3 gene weights became more strongly expressed during and after adolescence, whereas genes in the top

few (C2) or bottom few (C1) deciles of gene weights on the other two components were most strongly expressed in the first 5 years of life and then declined or plateaued during adolescence and early adulthood (Fig. 3e). These results confirmed that components C1–C3 have distinct neurodevelopmental trajectories, with genes positively weighted on C3 becoming strongly expressed after the first postnatal decade.

Autism and schizophrenia have specific links to C1/C2 and C3

Finally, we explored the clinical relevance of C1–C3 by analysis of previous neuroimaging, differential gene expression and GWAS associations for ASD, major depressive disorder (MDD) and schizophrenia.

First, we leveraged the BrainChart neuroimaging dataset of more than 125,000 MRI scans⁵⁸, in which atypical deviation of regional cortical volumes in psychiatric cases was quantified by centile scores relative to the median growth trajectories of normative brain development over the lifecycle (Fig. 4a). Using the Desikan–Killiany parcellation of 34 cortical regions necessitated by alignment with this dataset (Methods), we found that cortical shrinkage in ASD was significantly associated with both C1 and C2 ($r_{C1} = 0.49$, $P_{\text{spin}} = 0.0002$, $\text{FDR} < 5\%$; $r_{C2} = -0.28$, $P_{\text{spin}} = 0.0006$, $\text{FDR} < 5\%$), whereas shrinkage in schizophrenia was specifically associated with C3 ($r_{C3} = 0.43$, $P_{\text{spin}} = 0.008$, $\text{FDR} < 5\%$) (Fig. 4b).

Second, we compiled consensus lists of differentially expressed genes (DEGs) from RNA-seq measurements of dorsolateral prefrontal cortex (DLPFC) tissue in independent studies of ASD^{36,59,60}, MDD⁶¹ and schizophrenia^{60,62–65} (Methods). We found that genes differentially expressed in ASD were specifically enriched in both C1 and C2 (but not in C3), whereas genes differentially expressed in schizophrenia were enriched in C3 (but not in C1 or C2), and genes differentially enriched in MDD were enriched only in C1 (Fig. 4b). Corroborating the enrichments of ASD DEGs, case–control differences in expression at 11 cortical regions for ASD cases compared to healthy controls showed that the positively weighted genes on C1 and C2 were significantly less strongly expressed in ASD cases than in controls (Extended Data Fig. 3).

Third, using data from the most recent GWASs of ASD⁶⁶, MDD⁶⁷ and schizophrenia⁶⁸, we found that genetic variants significantly associated with ASD were enriched in both C1 and C2 (but not in C3), whereas genes associated with schizophrenia were enriched in C3 (but not in C1 or C2) (Fig. 4d). Genes associated with MDD were not significantly enriched in any transcriptional component. These associations were replicated when using alternative methods (MAGMA⁶⁹ and H-MAGMA⁷⁰) to test the association between GWAS-derived P values for the association of each gene with ASD, MDD or schizophrenia and the C1–C3 gene weights without requiring an explicit prioritization of GWAS-associated genes (Supplementary Fig. 6). This pattern of results for autism and schizophrenia GWAS associations evenly mirrored the pattern of previous results from analysis of case–control neuroimaging (Fig. 4b) and differential gene expression studies (Fig. 4c), with ASD consistently linked to components C1 and C2 and schizophrenia consistently linked to C3.

Notably, this consistency of association between disorders and specific transcriptional components was observed despite minimal overlap between the DEGs and GWAS risk genes identified as significant by the primary studies of each disorder⁷¹ (Fig. 4e). However, motivated by the association of C3 with regions of greatest laminar differentiation (Fig. 2b), we found that the subsets of the schizophrenia-associated DEG and GWAS gene sets that were positively weighted on C3 were both significantly enriched for marker genes of L2 and L3 (Fig. 4g and Extended Data Fig. 6). These shared laminar associations between the non-overlapping DEG and GWAS gene sets were present only when sub-setting to C3-positive genes and were specific to schizophrenia (that is, C3-positive subsets of ASD and MDD genes did not show the same L2/L3 enrichments). Convergent with C3 revealing an L2/L3 association in schizophrenia-associated genes from DEG and GWAS gene sets, we found that the cortical map of C3 was significantly co-located with an MRI-derived map of specifically supragranular, L2/L3-predominant thinning in schizophrenia⁷² ($r_{C3} = 0.55$, $P_{\text{spin}} = 0.002$, $\text{FDR} < 1\%$; Fig. 4g).

Discussion

Our results offer a new perspective on how the brain's macroscale organization develops from the microscale transcription of the human genome. Through optimized processing of the AHBA and replication in PsychENCODE, we show that the transcriptional architecture of the human cortex comprises at least three generalizable components of coordinated gene expression. The two higher-order components (C2 and C3) had not previously been robustly demonstrated, although the initial AHBA paper identified similar components to C1 and C2 by applying PCA to one of the six AHBA brains and filtering for only 1,000 genes² (Supplementary Fig. 1). In the present study, we derived C2 and C3 from all six AHBA brains and show here that they each represent the coordinated expression of hundreds of genes (Supplementary Fig. 2). Broadly, the C2 genes were enriched for 'metabolic' and 'epigenetic' processes, whereas the C3 genes were enriched for 'synaptic' and 'immune' processes (Fig. 1c). Both higher-order components were significantly enriched for genes associated with intelligence and educational attainment (Fig. 1f,g), indicating their relevance to the brain's ultimate purpose of generating adaptive behavior. The brain maps corresponding to each of the components were also distinctively co-located with multiple neuroimaging or other macroscale brain phenotypes (Fig. 2). These co-locations were often convergent with the gene enrichment results, triangulating evidence for C2 as a metabolically specialized component and for C3 as a component specialized for synaptic and immune processes underpinning adolescent plasticity (Table 1). Together, these convergent results expand on the proposal of a single S-A axis^{10,73} by demonstrating that macroscale brain organization emerges from multiple biologically relevant transcriptional components.

The discovery of these biologically relevant, higher-order transcriptional components in the AHBA dataset raised further questions. (1) Do the components reflect coordinated gene expression within cells or only variation in cell composition? (2) When do the components emerge during brain development? (3) How do they intersect with neurodevelopmental disorders? We addressed these questions using additional RNA-seq datasets (Supplementary Table 5). First, we found that genes positively or negatively weighted on the components derived from the AHBA bulk tissue samples had consistently coupled co-expression across RNA-seq measurements in single cells—for example, individual neurons and glia (Fig. 3a). This indicated that C1–C3 represent transcriptional programs coordinated at the intracellular level, not merely regional variation in the proportion of different cell types. Second, we found that C1–C3 have differentially phased developmental trajectories of expression—for example, that the positive pole of C3 becomes strongly expressed only during adolescence, convergent with its spatial co-location with a map of adolescent cortical myelination (Fig. 3b,c). Finally, we established that these transcriptional programs are not only critical for healthy brain development but, as might be expected, are also implicated in the pathogenesis of neurodevelopmental disorders (Fig. 4).

The pattern of results for disorders was strikingly convergent across multiple data modalities: C1 and C2 were both enriched for genes implicated by both GWAS and DEG data on ASD, whereas C3 was specifically enriched for genes implicated by both GWAS and DEG data on schizophrenia (Table 1). We observed a similar pattern of significant co-location between C1–C3 maps and MRI phenotypes: developmentally normalized scores on reduced cortical volume in ASD were correlated with maps of C1 and C2 and, for schizophrenia, with the map of C3 (Fig. 4a,b). In contrast, there was no evidence for enrichment of C1–C3 by genes associated with risk of Alzheimer's disease⁷⁴ (Supplementary Fig. 6). An intuitive generalisation of these results is that the developmental processes that give rise to these three components of gene expression in the healthy adult brain are pathogenically more relevant for neurodevelopmental disorders than for neurodegenerative disease.

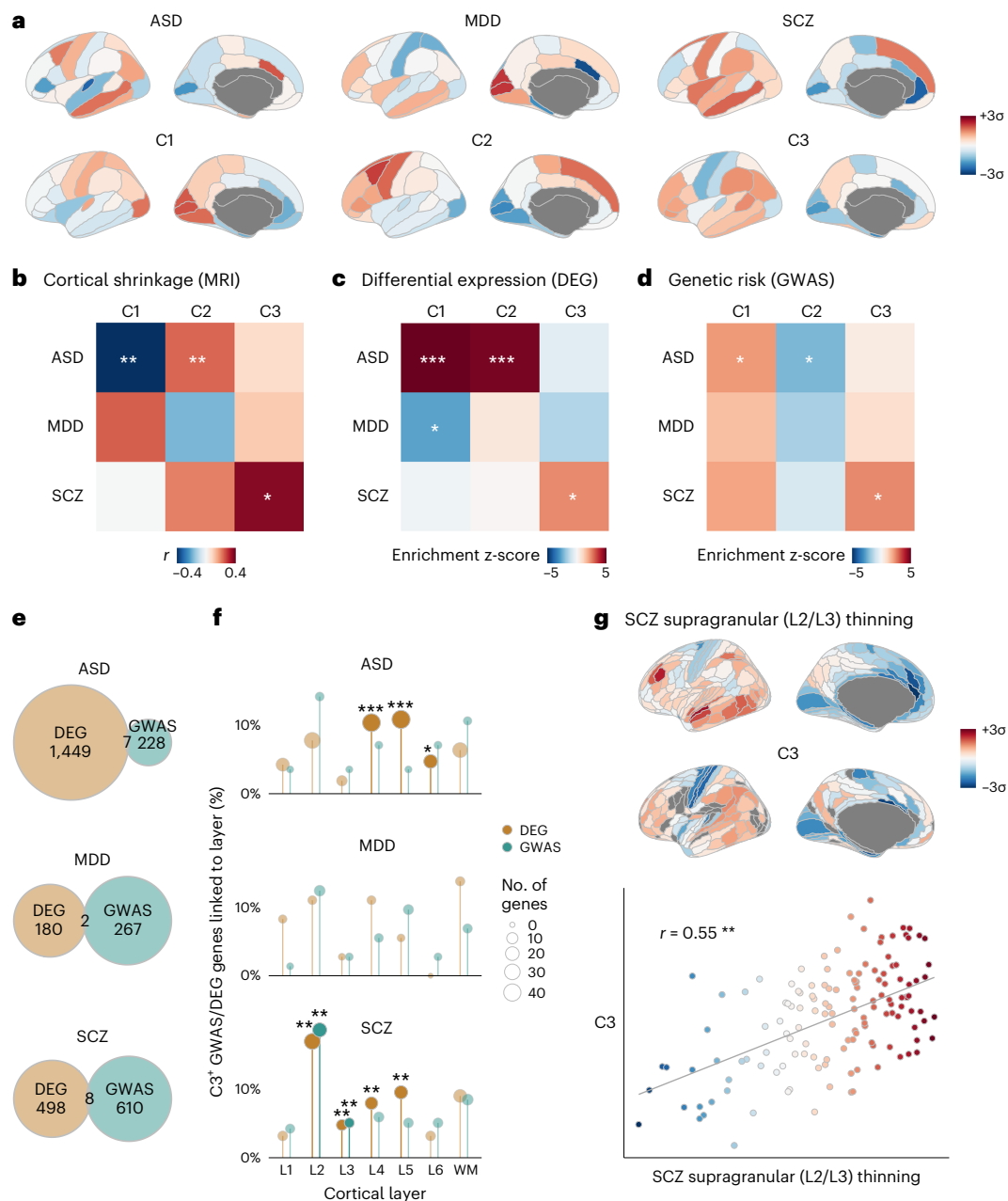


Fig. 4 | Genetics, transcriptomics and neuroimaging of autism and schizophrenia were consistently and specifically linked to normative transcriptional programs. **a**, First row: cortical volume shrinkage in ASD, MDD and schizophrenia (SCZ) cases. Red indicates greater shrinkage, computed as z-scores of centiles from normative modeling of more than 125,000 MRI scans. Second row: AHBA components projected into the same Desikan–Killiany parcellation. **b**, Spatial correlations between volume changes and AHBA components, C1–C3. Significance was tested by two-sided FDR-adjusted spatially autocorrelated spin permutations and corrected for multiple comparisons. **c**, Enrichments in C1–C3 for consensus lists of DEGs in postmortem brain tissue of donors with ASD, MDD and SCZ compared to healthy controls (Methods). Significance was assessed as percentile of mean weight of DEGs in each component relative to randomly permuted gene weights and corrected for multiple comparisons; two-sided FDR-adjusted *P* values. **d**, Enrichment in C1–C3 for GWAS risk genes for ASD⁶⁶, MDD⁶⁷ and SCZ⁶⁸, tested for significance as in **c**,

demonstrating alignment with both spatial associations to volume changes and enrichments for DEGs. **e**, Venn diagrams showing the lack of overlap of DEGs and GWAS risk genes reported by the primary studies summarized in **c** and **d**. **f**, DEGs and GWAS risk genes for each disorder were filtered for only C3-positive genes and then tested for enrichment with marker genes for each cortical layer³⁷. Significance was tested by one-sided Fisher’s exact test and corrected for multiple comparisons across all 42 tests. C3-positive DEGs and GWAS genes for SCZ (but not ASD or MDD) were both enriched for L2 and L3 marker genes, despite the DEGs and GWAS gene sets having nearly no overlap for each disorder (see Extended Data Fig. 6 for more detail). **g**, Convergent with L2/L3 enrichment in the C3-positive SCZ-associated DEGs and GWAS genes, a cortical map of supragranular-specific cortical thinning in SCZ⁷² was significantly and specifically co-located with C3 ($r = 0.55$, two-sided spin-permutation $P = 0.002$); each point is a region, and color represents C3 score. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Overall, our results were strongly supportive of the motivating hypothesis that the transcriptional architecture of the human cortex represents developmental programs crucial both to the brain’s healthy

organization and to the emergence of neurodevelopmental disorders. For example, when interpreting C3 as a transcriptional program mediating adolescent plasticity (Table 1), our finding that C3 represents

Table 1 | Summary of convergent results on the biological and clinical relevance of three human brain transcriptional programs

	C1: neuronal hierarchy	C2: cognitive metabolism	C3: adolescent plasticity
Biological processes (Fig. 1c)	<i>Most genes are aligned, especially PVALB, SST</i>	<i>Metabolism Epigenetics</i>	<i>Synaptic plasticity Learning/memory Immunity</i>
Architectonics (Fig. 1d)	<i>L4 L1, L2, L6</i>	<i>L4, L5, L6 L2</i>	<i>L2, L3, L4, L5, L6 L1, white matter</i>
Cell types (Fig. 1e)	<i>Oligodendrocytes Astrocytes</i>	<i>Synapses Endothelial cells</i>	<i>Synapses, Neurons Oligodendrocytes, Microglia</i>
GWAS (Fig. 1f,g)	<i>Educational attainment</i>	<i>Intelligence/cognition Educational attainment</i>	<i>Intelligence/cognition Educational attainment</i>
Imaging (Fig. 2)	<i>fMRI degree T1w/T2w Cortical thickness</i>	<i>MEG theta power Aerobic glycolysis</i>	<i>Adolescent change in myelination</i>
Development (Fig. 3b,c)	<i>Prenatal, greatest expression at birth</i>	<i>Prenatal, greatest expression in first decade</i>	<i>Adolescence, greatest expression in adulthood</i>
Atypical	Imaging (Fig. 4a,b,g)	ASD volume shrinkage	ASD volume shrinkage
	RNA-seq of brain tissue (Fig. 4c,f)	ASD DEGs	ASD DEGs
	GWAS (Fig. 4d,f)	ASD risk genes	ASD risk genes
			SCZ volume shrinkage and L2/L3-specific thinning
			SCZ DEGs, with L2, L3 enrichment
			SCZ risk genes, with L2, L3 enrichment

Each of three components of normative human brain gene expression (C1–C3; table columns) was biologically validated by testing for enrichment of gene weights on each component and for co-location of regional component scores with neuroimaging or other macroscale brain phenotypes, in healthy brain samples (normative) and in studies of neurodevelopmental disorders (atypical). Each row summarizes results for a distinct gene enrichment analysis (italicized) or spatial co-location analysis (plain font). Based on previous knowledge that theta oscillations are linked to intelligence and cognition⁸⁰ as well as to glucose metabolism⁸³, the spatial alignments between C2 and maps of MEG theta power (Fig. 2d) and aerobic glycolysis (Fig. 2a) were convergent with the enrichment of C2 for genes linked to cognitive capacity (Fig. 1f,g) and metabolism (Fig. 1c). Similarly, previous knowledge implicates microglia and oligodendrocytes in the immune-mediated synaptic pruning and myelination that, over adolescence, gives rise to adult cognitive capacity^{84,85}, such that the spatial alignment between C3 and the map of adolescent myelination (Fig. 2d) was convergent with the enrichments of C3 for genes related to immunity, synaptic development and learning (Fig. 1c); oligodendrocytes, microglia and synapses (Fig. 1e); and cognitive capacity (Fig. 1f,g), among which one GWAS explicitly linked intelligence to myelination⁴¹. SCZ, schizophrenia.

coupled transcription of synapse-related and immune-related genes within cells (Fig. 3a) is consistent with previous work indicating that the neuronal expression of immune-related, typically glial genes can play a mechanistic role in synaptic pruning⁷⁵ and, vice versa, that neuronal genes associated with synapse and circuit development can also be expressed in glial cells⁷⁶. Although atypical synaptic pruning has long been hypothesized to be a mechanistic cause of schizophrenia^{77–79}, previous results on the biology of schizophrenia have shown limited consistency, both among the primary data modalities of GWAS, postmortem expression and neuroimaging^{80,81} and even among DEG studies⁷¹. Here we demonstrate that the C3 transcriptional program offers a unifying link between these disparate previous results. When parsed by the C3-positive genes, the otherwise non-overlapping GWAS and DEG gene sets for schizophrenia display a shared enrichment for supragranular marker genes (Fig. 4e,f), and, convergently, C3 was spatially associated with supragranular-specific thinning in schizophrenia (Fig. 4g). Supragranular layers have dense cortico-cortical connections⁸² and are expanded in humans relative to other species^{83–85}, mature latest in development⁸⁶, have been linked to intelligence⁸⁷ and have previously been linked to schizophrenia^{88–90}. This triangulation of evidence strongly suggests that the third component of the brain's gene expression architecture represents the transcriptional program coordinating the normative, neuro-immune processes of synaptic pruning and myelination in adolescence⁵⁵, such that atypical expression of C3 genes due to schizophrenia genetic risk variants can result in atypical development of supragranular cortical connectivity, leading to the clinical emergence of schizophrenia.

Clearly, there are limits to what can be learned from RNA measurements of bulk tissue samples from six healthy adult brains. In the present study, we explicitly identified the limits of the AHBA dataset by optimizing data processing against an unbiased measure of generalizability, *g*, which yielded three components. The architecture of human

brain gene expression likely involves more than three components; however, our analysis suggests that their discovery will rely on additional high-granularity transcriptional data. In particular, gene expression varies with sex, age, genetics and environment⁹¹, so we expect that future data will reveal additional components that are more individually variable and demographically diverse than the three that we characterize here. In addition, the code and data that supplement our results can help future research to leverage our work with the unique AHBA resource.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-024-01624-4>.

References

- van den Heuvel, M. P. & Yeo, B. T. T. A spotlight on bridging microscale and macroscale human brain architecture. *Neuron* **93**, 1248–1251 (2017).
- Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
- Oldham, M. C. et al. Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).
- Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
- Ayoub, A. E. et al. Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc. Natl Acad. Sci. USA* **108**, 14950–14955 (2011).
- Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).

7. Bakken, T. E. et al. A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016).
8. Burt, J. B. et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
9. Fulcher, B. D., Murray, J. D., Zerbi, V. & Wang, X.-J. Multimodal gradients across mouse cortex. *Proc. Natl Acad. Sci. USA* **116**, 4689–4695 (2019).
10. Sydnor, V. J. et al. Neurodevelopment of the association cortices: patterns, mechanisms, and implications for psychopathology. *Neuron* **109**, 2820–2846 (2021).
11. Rayon, T., Maizels, R. J., Barrington, C. & Briscoe, J. Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic programme with human-specific features. *Development* **148**, dev199711 (2021).
12. Kudo, L. C., Karsten, S. L., Chen, J., Levitt, P. & Geschwind, D. H. Genetic analysis of anterior posterior expression gradients in the developing mammalian forebrain. *Cereb. Cortex* **17**, 2108–2122 (2007).
13. Polleux, F., Ince-Dunn, G. & Ghosh, A. Transcriptional regulation of vertebrate axon guidance and synapse formation. *Nat. Rev. Neurosci.* **8**, 331–340 (2007).
14. Krienen, F. M. et al. Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).
15. Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. & Raichle, M. E. Aerobic glycolysis in the human brain is associated with development and neotenus gene expression. *Cell Metab.* **19**, 49–57 (2014).
16. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
17. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl Acad. Sci. USA* **113**, 12574–12579 (2016).
18. Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. *Proc. Natl Acad. Sci. USA* **107**, 13135–13140 (2010).
19. Goulas, A., Margulies, D. S., Bezgin, G. & Hilgetag, C. C. The architecture of mammalian cortical connectomes in light of the theory of the dual origin of the cerebral cortex. *Cortex* **118**, 244–261 (2019).
20. Morgan, S. E. et al. Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes. *Proc. Natl Acad. Sci. USA* **116**, 9604–9609 (2019).
21. Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S. & Bethlehem, R. A. I. Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Mol. Psychiatry* **24**, 1053–1064 (2019).
22. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).
23. Romero-Garcia, R. et al. Schizotypy-related magnetization of cortex in healthy adolescence is colocalized with expression of schizophrenia-related genes. *Biol. Psychiatry* **88**, 248–259 (2020).
24. Anderson, K. M. et al. Transcriptional and imaging-genetic association of cortical interneurons, brain function, and schizophrenia risk. *Nat. Commun.* **11**, 2889 (2020).
25. Anderson, K. M. et al. Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. *Proc. Natl Acad. Sci. USA* **117**, 25138–25149 (2020).
26. Fornito, A., Arnatkeviciūtė, A. & Fulcher, B. D. Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).
27. Arnatkeviciute, A., Fulcher, B. D., Bellgrove, M. A. & Fornito, A. Imaging transcriptomics of brain disorders. *Biol. Psychiatry Glob. Open Sci.* **2**, 319–331 (2021).
28. Martins, D. et al. Imaging transcriptomics: convergent cellular, transcriptomic, and molecular neuroimaging signatures in the healthy adult human brain. *Cell Rep.* **37**, 110173 (2021).
29. Voineskos, A. N. et al. Effects of antipsychotic medication on brain structure in patients with major depressive disorder and psychotic features: neuroimaging findings in the context of a randomized placebo-controlled clinical trial. *JAMA Psychiatry* **77**, 674–683 (2020).
30. Perzel Mandell, K. A. et al. Molecular phenotypes associated with antipsychotic drugs in the human caudate nucleus. *Mol. Psychiatry* **27**, 2061–2067 (2022).
31. Schulmann, A. et al. Antipsychotic drug use complicates assessment of gene expression changes associated with schizophrenia. *Transl. Psychiatry* **13**, 93 (2023).
32. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
33. Arnatkeviciūtė, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).
34. Markello, R. D. et al. Standardizing workflows in imaging transcriptomics with the abagen toolbox. *eLife* **10**, e72129 (2021).
35. Hawrylycz, M. et al. Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).
36. Gandal, M. J. et al. Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
37. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
38. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
39. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
40. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
41. Hill, W. D. et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol. Psychiatry* **24**, 169–181 (2019).
42. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
43. Hatoum, A. S. et al. Genome-wide association study shows that executive functioning is influenced by GABAergic processes and is a neurocognitive genetic correlate of psychiatric disorders. *Biol. Psychiatry* **93**, 59–70 (2023).
44. Glasser, M. F. & Van Essen, D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
45. Van Essen, D. C. et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
46. Vaishnavi, S. N. et al. Regional aerobic glycolysis in the human brain. *Proc. Natl Acad. Sci. USA* **107**, 17757–17762 (2010).
47. Satterthwaite, T. D. et al. Impact of puberty on the evolution of cerebral perfusion during adolescence. *Proc. Natl Acad. Sci. USA* **111**, 8643–8648 (2014).
48. Reardon, P. K. et al. Normative brain size variation and brain shape diversity in humans. *Science* **360**, 1222–1227 (2018).
49. Paquola, C. et al. A multi-scale cortical wiring space links cellular architecture and functional dynamics in the human brain. *PLoS Biol.* **18**, e3000979 (2020).

50. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
51. Mesulam, M.-M. *Principles of Behavioral Neurology* (Oxford Univ. Press, 1985).
52. Glasser, M. F. et al. The Human Connectome Project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187 (2016).
53. Paquola, C. et al. Shifts in myeloarchitecture characterise adolescent development of cortical gradients. *eLife* **8**, e50482 (2019).
54. Larson-Prior, L. J. et al. Adding dynamics to the Human Connectome Project with MEG. *Neuroimage* **80**, 190–201 (2013).
55. Whitaker, K. J. et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl Acad. Sci. USA* **113**, 9105–9110 (2016).
56. Váša, F. et al. Conservative and disruptive modes of adolescent change in human brain functional connectivity. *Proc. Natl Acad. Sci. USA* **117**, 3248–3253 (2020).
57. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
58. Bethlehem, R. A. I. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
59. Parikshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
60. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
61. Jaffe, A. E. et al. Decoding shared versus divergent transcriptomic signatures across cortico-amygdala circuitry in PTSD and depressive disorders. *Am. J. Psychiatry* **179**, 673–686 (2022).
62. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
63. Birnbaum, R. et al. Investigating the neuroimmunogenic architecture of schizophrenia. *Mol. Psychiatry* **23**, 1251–1260 (2018).
64. Collado-Torres, L. et al. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron* **103**, 203–216 (2019).
65. Jaffe, A. E. et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
66. Matoba, N. et al. Common genetic risk variants identified in the SPARK cohort support *DDHD2* as a candidate risk gene for autism. *Transl. Psychiatry* **10**, 265 (2020).
67. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
68. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
69. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
70. Sey, N. Y. A. et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).
71. Merikangas, A. K. et al. What genes are differentially expressed in individuals with schizophrenia? A systematic review. *Mol. Psychiatry* **27**, 1373–1383 (2022).
72. Wagstyl, K. et al. Multiple markers of cortical morphology reveal evidence of supragranular thinning in schizophrenia. *Transl. Psychiatry* **6**, e780 (2016).
73. Hilgetag, C. C., Goulas, A. & Changeux, J.-P. A natural cortical axis connecting the outside and inside of the human brain. *Netw. Neurosci.* **6**, 950–959 (2022).
74. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
75. Faust, T. E., Gunner, G. & Schafer, D. P. Mechanisms governing activity-dependent synaptic pruning in the developing mammalian CNS. *Nat. Rev. Neurosci.* **22**, 657–673 (2021).
76. Kalish, B. T. et al. Single-cell transcriptomics of the developing lateral geniculate nucleus reveals insights into circuit assembly and refinement. *Proc. Natl Acad. Sci. USA* **115**, E1051–E1060 (2018).
77. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17**, 319–334 (1982).
78. Johnson, M. B. & Hyman, S. E. A critical perspective on the synaptic pruning hypothesis of schizophrenia pathogenesis. *Biol. Psychiatry* **92**, 440–442 (2022).
79. Howes, O. D. & Onwordi, E. C. The synaptic hypothesis of schizophrenia version III: a master mechanism. *Mol. Psychiatry* **28**, 1843–1856 (2023).
80. Harrison, P. J. & Weinberger, D. R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* **10**, 40–68 (2005).
81. Kleinman, J. E. et al. Genetic neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem human brains. *Biol. Psychiatry* **69**, 140–145 (2011).
82. Romero-Garcia, R. et al. Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. *Neuroimage* **171**, 256–267 (2018).
83. Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K. & Van Hoesen, G. W. Prefrontal cortex in humans and apes: a comparative study of area 10. *Am. J. Phys. Anthropol.* **114**, 224–241 (2001).
84. Hutsler, J. J., Lee, D.-G. & Porter, K. K. Comparative analysis of cortical layering and supragranular layer enlargement in rodent carnivore and primate species. *Brain Res.* **1052**, 71–81 (2005).
85. Galakhova, A. A. et al. Evolution of cortical neurons supporting human cognition. *Trends Cogn. Sci.* **26**, 909–922 (2022).
86. Petanjek, Z. et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl Acad. Sci. USA* **108**, 13281–13286 (2011).
87. Heyer, D. B. et al. Verbal and general IQ associate with supragranular layer thickness and cell properties of the left temporal cortex. *Cereb. Cortex* **32**, 2343–2357 (2022).
88. Chance, S. A., Walker, M. & Crow, T. J. Reduced density of calbindin-immunoreactive interneurons in the planum temporale in schizophrenia. *Brain Res.* **1046**, 32–37 (2005).
89. Cullen, T. J. et al. Anomalies of asymmetry of pyramidal cell density and structure in dorsolateral prefrontal cortex in schizophrenia. *Br. J. Psychiatry* **188**, 26–31 (2006).
90. Batiuk, M. Y. et al. Upper cortical layer-driven network impairment in schizophrenia. *Sci. Adv.* **8**, eabn8367 (2022).
91. Viñuela, A. et al. Age-dependent changes in mean and variance of gene expression across tissues in a twin cohort. *Hum. Mol. Genet.* **27**, 732–741 (2018).
92. Senoussi, M. et al. Theta oscillations shift towards optimal frequency for cognitive control. *Nat. Hum. Behav.* **6**, 1000–1013 (2022).
93. Pizzagalli, D. A., Oakes, T. R. & Davidson, R. J. Coupling of theta activity and glucose metabolism in the human rostral anterior cingulate cortex: an EEG/PET study of normal and depressed subjects. *Psychophysiology* **40**, 939–949 (2003).

94. Choudhury, S., Charman, T. & Blakemore, S.-J. Development of the teenage brain. *Mind Brain Educ.* **2**, 142–147 (2008).
95. Brenhouse, H. C. & Schwarz, J. M. Immunoadolescence: neuroimmune development and adolescent behavior. *Neurosci. Biobehav. Rev.* **70**, 288–299 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Methods

AHBA data and donor-level parcellation images

Probe-level gene expression data with associated spatial coordinates were obtained from the Allen Institute website (<https://human.brain-map.org>), which collected the data after obtaining informed consent from the deceased's next of kin. HCP-MMP1.0 parcellation images matched to the individual native MRI space of each donor brain ($n = 6$) were obtained from Arnatkevičiūtė et al. (<https://figshare.com/articles/dataset/AHBAdata/6852911>)³³. The use of native donor parcellation images (rather than a standard parcellation image with sample coordinates mapped to MNI space) was chosen as it optimized the triplet generalizability metric (see the following).

AHBA processing parameters

To correctly match AHBA samples to regions in native donor space parcellation images using published processing pipelines, we recommend the use of either (1) abagen version 0.1.3 or greater (for Python)³⁴ or (2) the version of the AHBA processing pipeline updated in June 2021 or later (for MATLAB)³³.

In the present study, we processed the AHBA with the abagen package, with one modification: we filtered the AHBA samples for only those annotated as cortical samples before subsequent processing steps. This was done such that subcortical and brainstem samples did not influence the intensity filter and probe aggregation steps. This modification was chosen as it optimized the triplet generalizability metric (see the following). The code used to apply the modification is available in the code/processing_helpers.py file at https://github.com/richardajdear/AHBA_gradients.

Other than this modification, abagen was run using the following parameters, which follow published recommendations³³ unless otherwise specified:

- Hemisphere: The right hemisphere samples that are present for two of the six donors were reflected along the midline and processed together with the left hemisphere samples of those donor datasets to increase sample coverage.
- Intensity-based filter: Probes were filtered to retain only those exceeding background noise (as defined by the binary flag provided with the data by the Allen Institute) in at least 50% of the samples³³.
- Probe aggregation: Probes were aggregated to genes by differential stability, meaning that, for each gene, the probe with the highest mean correlation across donor pairs was used.
- Distance threshold: Samples were matched to regions with a tolerance threshold of 2 mm, using the voxel-mass algorithm in the abagen package.
- Sample normalization: Before aggregating over donors, samples were normalized across all genes, using the scaled robust sigmoid method (a sigmoid transformation that is robust to outliers³³).
- Gene normalization: Before aggregating over donors, genes were normalized across all samples, again using the scaled robust sigmoid.

To ensure robustness, the primary analysis of computing components of the AHBA was repeated in a series of sensitivity analyses varying all of the processing parameters above—for example, not mirroring right hemisphere samples to the left hemisphere, different or no intensity filter for genes and different methods for aggregating and normalizing probes. Sensitivity analyses also included running the pipeline with alternative parcellation templates: HCP-MMP1.0 (ref. 32), Schaefer-400 (ref. 96) Desterieux³⁷ and Desikan–Killiany⁹⁸ (Extended Data Fig. 2).

Gene filtering by differential stability

Genes were filtered for those that showed more similar spatial patterns of expression across the six donors using the metric of differential

stability as previously described by Hawrylycz et al.³⁵ For each gene, differential stability was calculated as the average correlation of that gene's regional expression vector between each donor pair (15 pairs with all six brains or three pairs in the triplets analysis; see below). Genes were ranked by differential stability, and then only the top 50% percent of genes were retained. The 50% threshold was chosen on the basis of a grid search (in combination with the region filter to optimize for generalizability) where the threshold for differential stability was varied between 10% and 100% (Extended Data Fig. 1).

Filtering regions by donors represented

Regions were filtered for those that included samples from at least three of the six AHBA donor brains, which, in the HCP-MMP1.0 parcellation, retained 137 of 180 regions. Note that, in the triplets analysis (see below), this means that only brain regions with samples from all three donors in the triplet were retained. The choice to filter for representation of three of the six donors was made on the basis of a grid search in combination with the differential stability gene filter to optimize for generalizability (Extended Data Fig. 1).

Triplets analysis: disjoint triplet correlation as a proxy for generalizability

To test for generalizability, we separated the six AHBA brains into pairs of disjoint triplets (for example, donor brains 1,2,3 in one triplet and 4,5,6 in another). We applied our full analysis pipeline (including all processing steps—for example, probe aggregation, normalization and filters) independently to each of the 20 possible combinations of triplets and correlated the regional scores for each DME or PCA component between each of the 10 disjoint pairs (Pearson's r). When filtering for consistently sampled regions, the retained regions were different for each triplet of donor brains, so correlations were performed on only the intersection of regions retained in both triplets of each pair.

As the order of principal components can vary across different triplets, we used a matching algorithm in which the full correlation matrix was computed among the top five principal components of both triplets (for example, C1 from triplet A was correlated with each of C1–C5 of triplet B). The highest absolute correlation value in the matrix was then identified as representing two matched components and removed from the matrix, with the process repeated until all components were matched. The components were then ranked by the mean variance explained in each matched pair.

The median absolute correlation across all 10 disjoint triplet pairs represented the generalizability, g , of the AHBA components processed using the given set of parameters. Processing parameters, in particular the filters for regions and donors, were optimized so as to maximize g while retaining as many genes and regions as possible (Extended Data Fig. 1).

Dimension reduction methods

Dimension reduction was performed using both PCA and DME, the latter having been described for use in spatial gradient analysis of brain imaging data by Margulies et al.¹⁷ For DME, the normalized cosine function was used as the kernel for the affinity matrix. No sparsity was added, and the alpha parameter was set at 1. These parameters were chosen as they optimized the inter-triplet correlation metric for generalizability. Both PCA and DME methods were implemented using the BrainSpace package⁹⁹. See Supplementary Methods for further explanation on DME and its benefits over PCA and other alternatives (for example, independent component analysis).

Component gene weights

For each component, gene weights were computed as the Pearson's correlation of each gene's individual spatial expression vector with the regional scores of the component. For PCA, these correlations are equivalent to the PCA loadings (eigenvectors) multiplied by the square root of the variance explained by the component (eigenvalues).

Variance explained

For PCA, variance explained is given directly by the squared eigenvalues of the singular value decomposition. For DME, eigenvalues do not represent variance explained as the gene expression matrix is first converted to an affinity matrix using a kernel (here, the normalized cosine). Therefore, variance explained was calculated as the difference in the total variance of the region-by-gene expression matrix before and after regressing the matrix on each component's region scores.

That is, defining the residual regional expression vector of gene g after regressing out i components as $\mathbf{e}_{g,i}$, the total variance V_i of the residualized region-by-gene expression matrix is

$$V_i = \sum_g \text{Var}(\mathbf{e}_{g,i})$$

and, for each component C_i , variance explained VE_i is given by

$$VE_i = V_{i-1} - V_i$$

GO enrichment analysis for biological processes

Biological process enrichments of the gene weights for each component were computed using the 'proteins with values/ranks' function of the online software STRING¹⁰⁰, which tests whether the mean weight of each annotated gene list is significantly higher or lower than random permutations of the same gene weights (the 'aggregate fold change' method^{100,101}) and includes a Benjamini–Hochberg adjustment of the FDR.

The aggregate fold change method was chosen as it does not require thresholding the gene weights of the components to define 'target' versus 'background' gene lists (as in, for example, Fisher's exact test). That is, rather than setting a threshold for which genes are 'in' or 'out' of each component, we took the weighted gene list where all genes can have some contribution to each component and, for each component, tested whether each GO gene list was, in aggregate, more positively or negatively weighted than chance.

Layer and cell type enrichment analyses

The gene lists for cortical layer marker genes were obtained from published analyses of laminar enrichment in spatial transcriptomic data from human postmortem tissue in the DLPFC³⁷ (columns Q–W of supplementary table 4b in Maynard et al.³⁷).

Cell type gene lists were obtained from Seidlitz et al.²², who compiled lists of significantly differentially expressed genes from five independent single-cell RNA-seq studies^{38,102–105}. The gene list for synaptic marker genes was the unfiltered gene list from SynaptomeDB¹⁰⁶.

All enrichments for layers and cell types were computed by the same aggregate fold change method¹⁰¹ as in the STRING software¹⁰⁰, whereby the mean gene weight of each gene list was computed for both the true set of gene weights of each component and for 5,000 random permutations of the weights. The z-scores and permutation test P values for significance testing of enrichment were corrected for multiple comparisons with the Benjamini–Hochberg FDR.

GWAS enrichment analyses for educational attainment and intelligence

Genes associated with cognitive capacity by GWAS were obtained from:

- Lee et al.³⁹, supplementary table 7 (educational attainment)
- Davies et al.⁴⁰, supplementary table 6
- Savage et al.⁴², supplementary table 15
- Hill et al.⁴¹, supplementary table 5
- Hatoum et al.⁴³, supplementary table 16

Enrichment tests were performed by the aggregate fold change method¹⁰¹, as above.

Neuroimaging and other macroscale brain maps (Fig. 2)

Neuroimaging and other macroscale maps were obtained as follows:

- The nine neuroimaging and macroscale maps in the clustering analysis (Fig. 2a) were obtained from the neuromaps package¹⁰⁷ and are also available in Sydnor et al.¹⁰.
- The regions of cytoarchitectural differentiation (Fig. 2b) were obtained from Paquola et al.¹⁰⁸ and averaged into the HCP-MMP parcellation using the neuromaps package¹⁰⁷.
- The map of fMRI degree (Fig. 2c) was obtained from Paquola et al.⁴⁹ and was originally computed from the HCP S900 release¹⁰⁹.
- The maps of MEG power bands (Fig. 2d and Extended Data Fig. 5) were obtained from the neuromaps package¹⁰⁷.
- The map of adolescent change in cortical myelination was obtained from Váša et al.⁵⁶.

All maps were aggregated into the HCP-MMP parcellation and are provided in Supplementary Table 3.

Spatial associations between maps and the transcriptional components were computed by Pearson's correlations and tested for significance using spin permutation tests (5,000 spins) by the Cornblath method¹¹⁰, leveraging tools from neuromaps¹⁰⁷, and tested for significance with FDR correction for multiple testing.

For the regions of cytoarchitectural differentiation, the mean component scores in each architectonic class were tested for differences between class mean scores using ANOVA against spin-permuted null models, followed by correction for FDR. The associations between individual cytoarchitectural regions and each component were computed by the z-score of the mean component score in each region normalized by a spin permutation distribution of the regional mean component score with significance testing corrected for FDR.

Single-cell co-variation analysis (Fig. 3a)

Single-cell RNA-seq data were obtained from the Allen Cell Types Database (<https://portal.brain-map.org/atlas-and-data/rnaseq>)⁵⁷.

Single-cell gene expression was filtered for the 7,873 genes in the optimally filtered AHBA dataset. To perform the analysis in Fig. 3a, the positive and negative gene weights were separated for each of C1–C3, and the dot product was taken with the gene expression matrix of single-cell samples. This produced a vector of six numbers, representing the weighted total expression of C1⁺, C1[−], C2⁺, C2[−], C3⁺ and C3[−] genes, respectively, for each of the 50,000 single-cell samples.

That is, given the gene expression vector s_j of each single-cell sample j , we computed the total weighted positive and negative expression $s^+_{j,Ci}$ and $s^-_{j,Ci}$ from the C1–C3 gene weights as:

$$s^+_{j,Ci} = s_j \cdot u^+_{Ci} \quad \text{and} \quad s^-_{j,Ci} = s_j \cdot u^-_{Ci}$$

where $u^+_{Ci} = \max\{u_{Ci}, 0\}$ and $u^-_{Ci} = \min\{u_{Ci}, 0\}$.

BrainSpan developmental gene expression processing (Fig. 3b–d)

BrainSpan data were obtained directly from the Allen Institute website⁶ (<http://brainspan.org>) and processed as follows:

1. The 11 cortical regions in the BrainSpan data were manually matched to the HCP-MMP1.0 parcellation regions according to the descriptions in the BrainSpan documentation. This mapping is provided at https://github.com/richardajdear/AHBA_gradients.
2. Exon-level expression data were filtered for only the matched BrainSpan regions.
3. Donor brains from which fewer than four regions were sampled were dropped.
4. Within each donor, expression of each gene was z-normalized over regions.

- Donors were aggregated into three age ranges (pre-birth, birth to 13 years and 18–40 years), and expression was averaged for each gene.

AHBA-BrainSpan developmental consistency analysis (Fig. 3b–d)

Consistency between the AHBA components and BrainSpan was evaluated as follows:

- Processed BrainSpan data were filtered for only the 7,973 genes retained in the filtered AHBA dataset (top 50% by differential stability; see above).
- The dot product of the gene weights for C1–C3 were taken against the BrainSpan data, resulting in ‘BrainSpan scores’ for each of C1–C3, for each of the 11 BrainSpan regions, at each age range (pre-birth, birth to 13 years and 18–40 years).
- In each of the 11 BrainSpan regions, ‘AHBA scores’ were computed as the mean of the matching HCP-MMP region scores from the original C1–C3 maps derived from the AHBA.
- The ‘BrainSpan scores’ and ‘AHBA scores’ were correlated over the 11 BrainSpan regions (Pearson’s r) for each of C1–C3 and for each age bucket of the BrainSpan data.

As further clarification: given gene weights u_i for AHBA component C_i and the vector of expression over genes b_j for each BrainSpan sample j (with a given age and region), the ‘BrainSpan score’ is

$$y_{j,i} = b_j \cdot u_i$$

and the consistency was tested as the correlation across the matched regions of the AHBA scores x and the mean of the BrainSpan scores y of BrainSpan donors in each age range.

BrainSpan developmental trajectory modeling (Fig. 3e)

The developmental trajectories of each decile of C1–C3 were computed as follows:

- The ages in the BrainSpan data were converted to post-conception days on a \log_{10} scale.
- For each gene, a generalized additive model was fitted using the GLMGam function in the statsmodels Python package with $\alpha = 1$ and 12 3rd-degree basis splines as a smoothing function ($df = 12$, $degree = 3$ in the BSplines function). Sex and brain region were included as covariates.
- Developmental curves were plotted from the fitted models for each gene, sex and region and then averaged by decile of gene weight for each of C1–C3.

Disorder spatial associations (Fig. 4a,b)

Maps of the regional centile score differences in cortical volume for ASD, MDD and schizophrenia were obtained from the BrainCharts project by Bethlehem et al.⁵⁸, in which normative models were computed for multiple brain phenotypes across the human lifespan from a harmonized dataset of more than 125,000 total MRI scans ($n_{\text{controls}} = 38,839$, $n_{\text{ASD}} = 381$, $n_{\text{MDD}} = 3,861$, $n_{\text{SCZ}} = 315$). As these data were in the Desikan–Killiany parcellation, the AHBA components in the HCP-MMP parcellation were mapped to a vertex-level surface map (FreeSurfer’s 41k fsaverage atlas) and then re-averaged into the Desikan–Killiany parcellation. Pearson’s correlations with cortical maps of C1–C3 scores were computed, and significance was assessed by spin permutation tests and corrected for FDR across all nine tests (three disorders by three components).

These disorder maps are provided in Supplementary Table 4.

Disorder DEG associations (Fig. 4c)

DEGs (FDR < 5%) from RNA-seq of postmortem brain tissue were obtained from the following case–control studies for each of ASD, MDD and schizophrenia:

- ASD:
 - Gandal et al.³⁶, supplementary table 3, WholeCortex_ASD_FDR < 0.05
 - Gandal et al.⁶⁰, supplementary table 1, ASD.fdr < 0.05
 - Parikshak et al.⁵⁹, supplementary table 2, FDR-adjusted P value, ASD versus CTL < 0.05
- MDD
 - Jaffe et al.⁶¹, supplementary table 2, Cortex_adjPVal_MDD < 0.05
- Schizophrenia
 - Fromer et al.⁶², supplementary table 16, FDR estimate < 0.05
 - Gandal et al.⁶⁰, supplementary table 1, SCZ.fdr < 0.05
 - Jaffe et al.⁶⁵, supplementary table 9, fdr_qsva < 0.05
 - Collado-Torres et al.⁶⁴, supplementary table 11, adj.P.Val < 0.05 & region == ‘DLPFC’

A consensus list of DEGs was compiled for each disorder (except MDD where only one study was included) by including only those genes identified in at least two studies.

Enrichments for these gene sets in each disorder were computed by the aggregate fold change method¹⁰¹—that is, computing the percentile of the mean weight of the DEGs in C1–C3 relative to the 5,000 random permutations of the gene labels.

Disorder-associated genes from GWAS (Fig. 4d)

Genes significantly associated with ASD, MDD and schizophrenia by GWAS were obtained from:

- ASD: Matoba et al.⁶⁶, supplementary table 7
- MDD: Howard et al.⁶⁷, supplementary table 9
- Schizophrenia: Trubetskoy et al.⁶⁸, extended GWAS: <https://figshare.com/articles/dataset/scz2022/19426775?file=35775617>

Associations with GWAS were calculated using three methods (Supplementary Fig. 6):

- Enrichment of the prioritized genes identified in each of the specific studies, using the aggregate fold change method¹⁰¹ as described above.
- MAGMA⁶⁹, a regression technique that tests for association between each of the components C1–C3 and the P values for each gene’s association with ASD, MDD or schizophrenia (from corresponding primary GWASs) without requiring a threshold to be applied to the GWAS-derived P values to define a prioritized subset of genes for enrichment analysis. MAGMA additionally accounts for gene length and gene–gene correlations. The COVAR function of MAGMA was used to test for association of the GWAS P values with the C1–C3 gene weights as a continuous variable. For standard MAGMA, a single-nucleotide polymorphism (SNP)-to-gene mapping window of +35 kb to –10 kb was used.
- H-MAGMA⁷⁰, an extension of MAGMA where SNP-to-gene mapping is performed using Hi-C chromatin measurements from postmortem brain tissue so as to capture *trans*-regulatory effects. We used the Hi-C mapping from adult brain DLPFC, available online from the original H-MAGMA authors.

Laminar enrichments shared across DEG and GWAS gene sets (Fig. 4f)

Enrichments for the marker genes of each cortical layer³⁷ were computed for the disorder-associated gene lists from DEGs and GWASs using Fisher’s exact test. These enrichments were computed both with and without filtering for only genes with positive C3 weights.

Schizophrenia supragranular-specific cortical thinning (Fig. 4g)

The MRI-derived map of supragranular cortical thinning in schizophrenia was obtained from Wagstyl et al.⁷² ($n = 90$ subjects, $n = 46$ cases) and parcellated using HCP-MMP1.0 parcellation. Pearson’s correlations

were computed with C1–C3, and significance was assessed by spin permutation tests, corrected for FDR.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Regional scores and gene weights for the transcriptional components C1–C3 are provided in Supplementary Table 1. Gene expression datasets used are all publicly available. The Allen Human Brain Atlas is available at <http://human.brain-map.org> and individual donor HCP-MMP parcellation images at <https://figshare.com/articles/dataset/AHBAdata/6852911>. The BrainSpan Atlas is available at <https://www.brainspan.org/>. The Allen Human Cell Atlas is available at <https://portal.brain-map.org/atlasses-and-data/rnaseq>. The PsychENCODE dataset is available at <https://github.com/dhglab/Broad-transcriptomic-dysregulation-across-the-cerebral-cortex-in-AS-D>. Neuroimaging maps of healthy brain features are available in the neuromaps package (<https://github.com/netneurolab/neuromaps>). For convenience, all brain maps used are provided in Supplementary Tables 3 and 4. Gene lists used for enrichment analyses were all obtained from previous publications as detailed in the Methods.

Code availability

Analyses were performed with Python version 3.10.5 and R version 2.2. Key Python packages include abagen version 0.1.3, brainspace version 0.1.10 and neuromaps version 0.0.3. Full details of all packages, a Dockerfile and links to Docker images and all code used for these analyses are publicly available at https://github.com/richardajdear/AHBA_gradients.

References

- Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
- Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
- Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
- Vos de Wael, R. et al. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun. Biol.* **3**, 103 (2020).
- Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2018).
- Yu, C. et al. A strategy for evaluating pathway analysis methods. *BMC Bioinformatics* **18**, 453 (2017).
- Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
- Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
- Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
- Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
- Pirooznia, M. et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* **28**, 897–899 (2012).
- Markello, R. D. et al. neuromaps: structural and functional interpretation of brain maps. *Nat. Methods* **19**, 1472–1479 (2022).
- Paquola, C. et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol.* **17**, e3000284 (2019).
- Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
- Markello, R. D. & Misic, B. Comparing spatial null models for brain maps. *Neuroimage* **236**, 118052 (2021).

Acknowledgements

The authors acknowledge and thank S. Valk and V. Warriar for their helpful comments on the manuscript. R.D. was supported by the Gates Cambridge Trust (OPP1144). J.S. was supported by National Institute of Mental Health (NIMH) T32MH019112-29 and K08MH120564. A.A. was funded by a grant from the Australian Research Council under its Linkage Project scheme (LP160101592). R.A.I.B. was supported by the Autism Research Trust. K.S.W. was supported by the Wellcome Trust (215901/Z/19/Z). E.T.B. was supported by a National Institute for Health and Care Research (NIHR) Senior Investigator award and the Wellcome Trust collaborative award for the Neuroscience in Psychiatry Network. A.R. was supported by the NIMH Intramural Research Program (National Institutes of Health (NIH) annual report number: 1ZIAMH002949-04). P.E.V. is a fellow of MQ: Transforming Mental Health (MQF17_24). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Data were curated and analyzed using a computational facility funded by a Medical Research Council research infrastructure award (MR/M009041/1) to the School of Clinical Medicine, University of Cambridge. All research in the Department of Psychiatry at the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and the NIHR Applied Research Centre. The views expressed are those of the authors and not necessarily those of the NIH, the National Health Service, the NIHR or the Department of Health and Social Care. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

P.E.V., R.D., J.S., K.W. and A.R. designed the research. J.S., A.A., K.W. and R.B. contributed data. R.D. and P.E.V. performed the research. E.T.B., K.A., K.W., A.R., J.S., R.M. and A.A. helped interpret results. R.D., P.E.V. and E.T.B. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

K.M.A. is an employee of Neumora Therapeutics. R.D.M. is an employee of Octave Biosciences. E.T.B. has consulted for Boehringer Ingelheim, SR One, GlaxoSmithKline, Sosei Heptares and Monument Therapeutics. All other authors have no disclosures to make.

Additional information

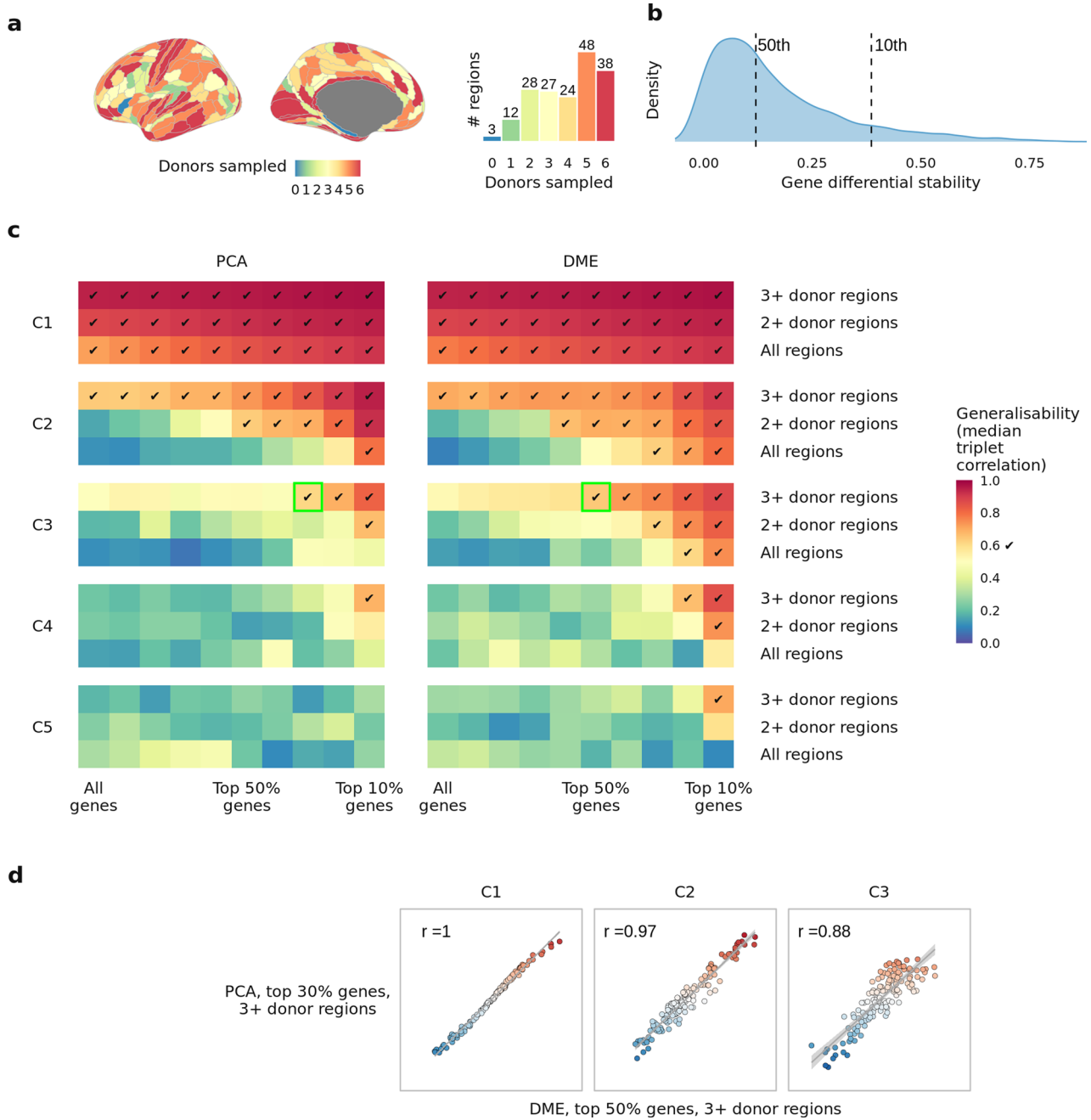
Extended data is available for this paper at <https://doi.org/10.1038/s41593-024-01624-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-024-01624-4>.

Correspondence and requests for materials should be addressed to Richard Dear.

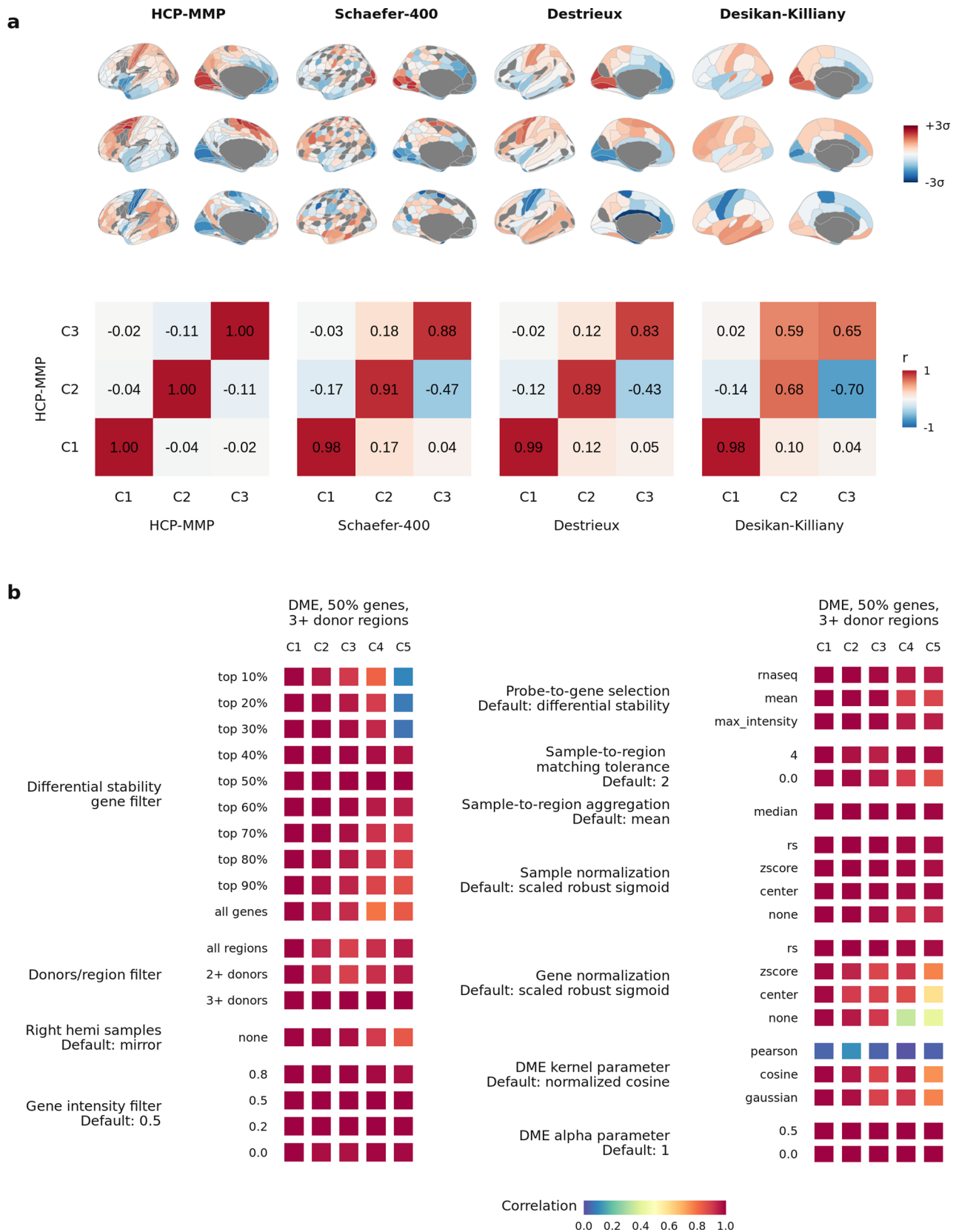
Peer review information *Nature Neuroscience* thanks Michael Hawrylycz, Neelroop Parikshak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



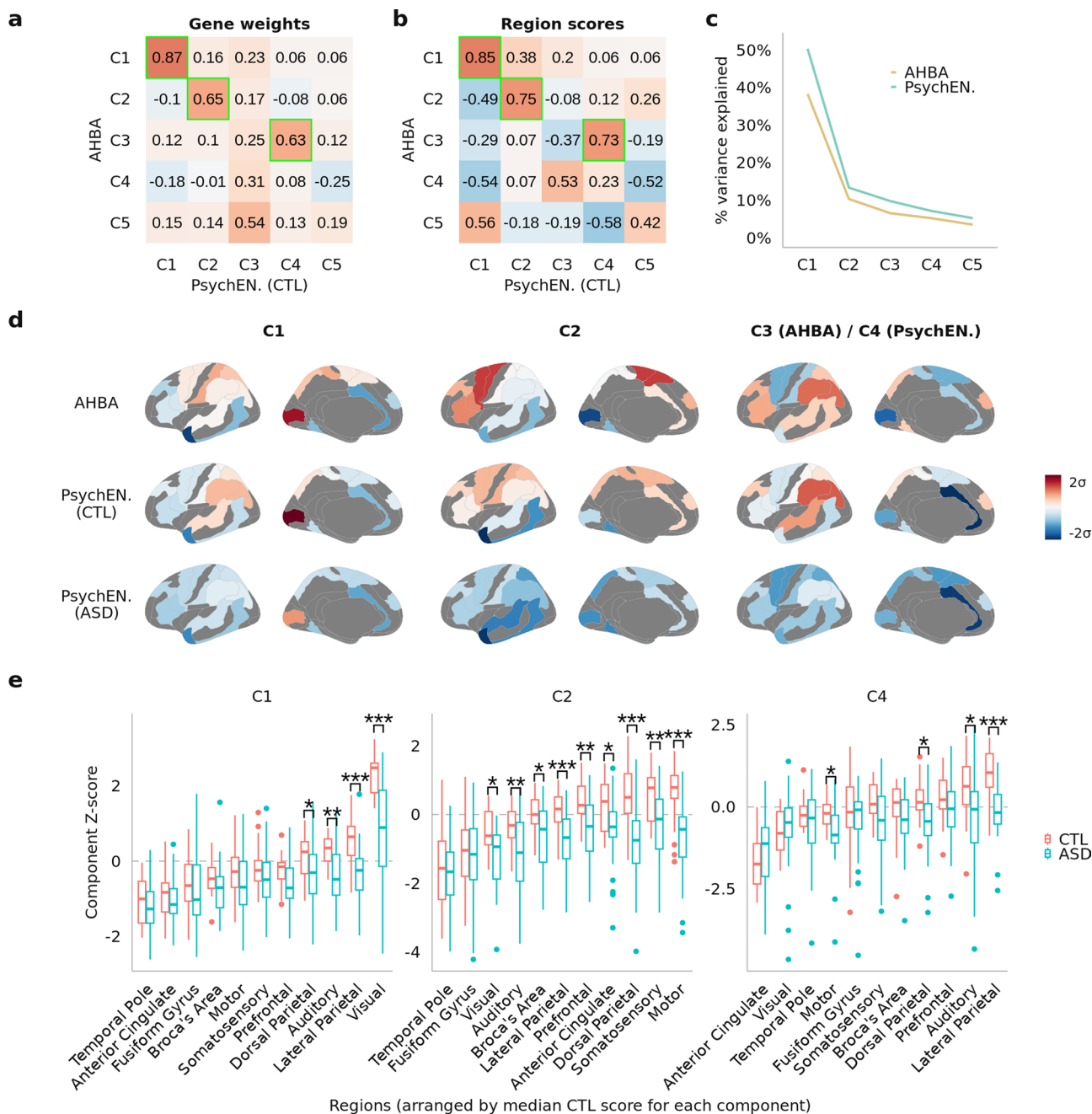
Extended Data Fig. 1 | Optimised processing of the AHBA identified three generalisable components. a, In the HCP-MMP parcellation, 43/180 regions are matched to samples representing less than 3 of the 6 AHBA donors. **b**, Distribution of differential stability of genes measured in the AHBA dataset processed in the HCP-MMP parcellation. **c**, Generalisability of first five components of the AHBA dataset computed with either principal components analysis (PCA) or diffusion map embedding (DME). Color represents generalisability g , defined as the median absolute correlation between matched components computed across all 10 disjoint triplet pairs (Methods); x-axis

represents variation in the proportion of genes filtered out by differential stability prior to PCA/DME; y-axis represents variation in which regions are filtered out prior to PCA/DME. Tick mark indicates parameter combinations that exceed generalisability $g > 0.6$. Green highlights for C3 indicate the best parameter option with PCA and DME respectively, showing that switching to DME achieves similar generalisability while retaining more genes. **d**, Scatter plots of regional scores for AHBA components computed using the best PCA/DME options, demonstrating that PCA and DME derive spatially equivalent components.



Extended Data Fig. 2 | Transcriptional components were robust to parcellation and processing. Transcriptional components were computed in four different parcellation templates (Methods). For each parcellation, the gene weights for the first three components were correlated with the weights obtained from the HCP-MMP parcellation used throughout. Gene weights were highly consistent, although in the less-granular (34-regions/hemisphere) Desikan-

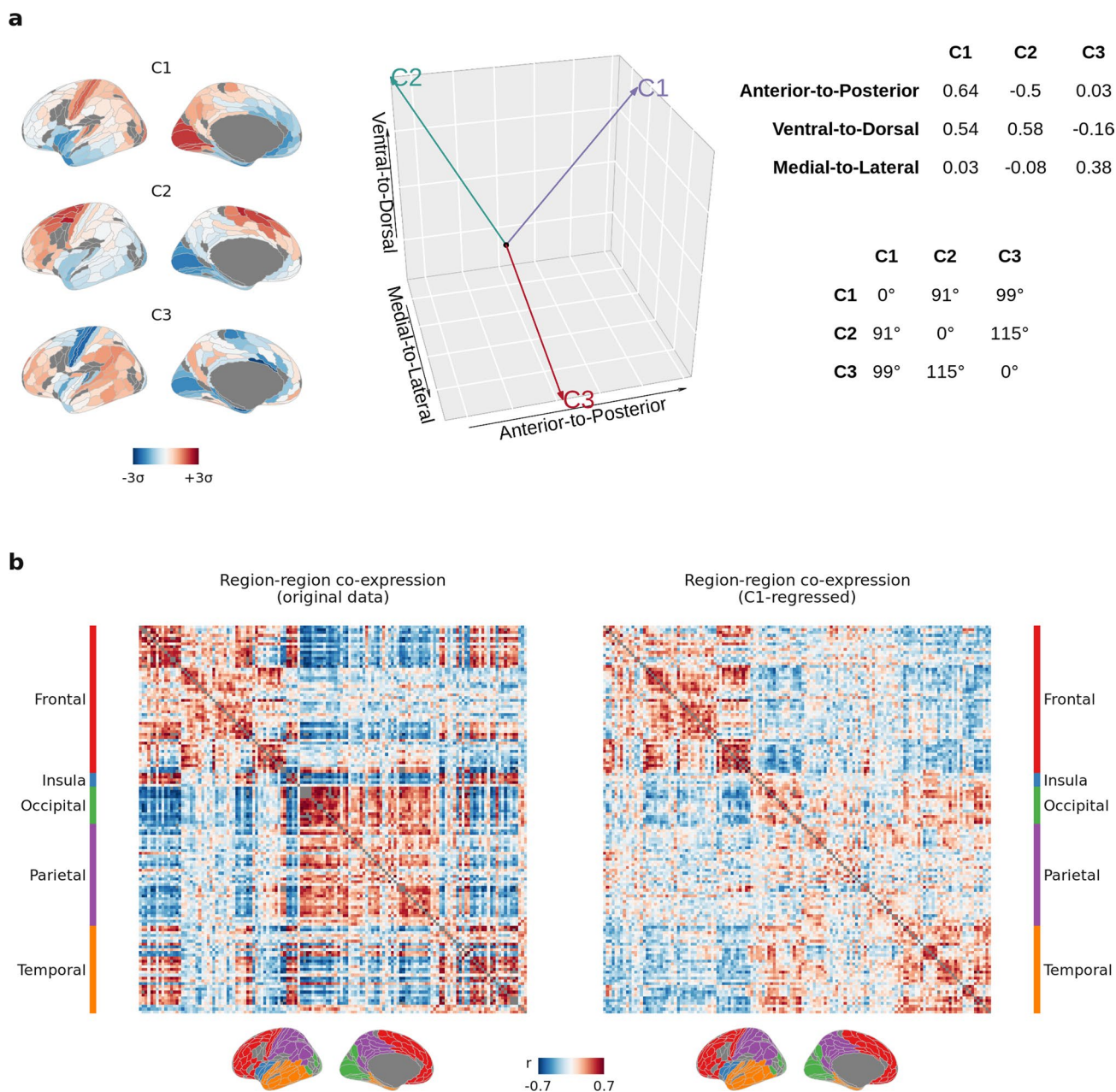
Killiany parcellation, C2 and C3 were less well aligned to the other parcellations. **b**, A wide range of parameters for processing the AHBA data were varied, and the resulting component region scores were correlated with the components obtained from the optimised parameters. For nearly all variations in parameters, highly consistent components were obtained, demonstrating the robustness of C1-C3.



Extended Data Fig. 3 | AHBA transcriptional components were reproducible in independent PsychENCODE control data, with differential spatial expression in autism.

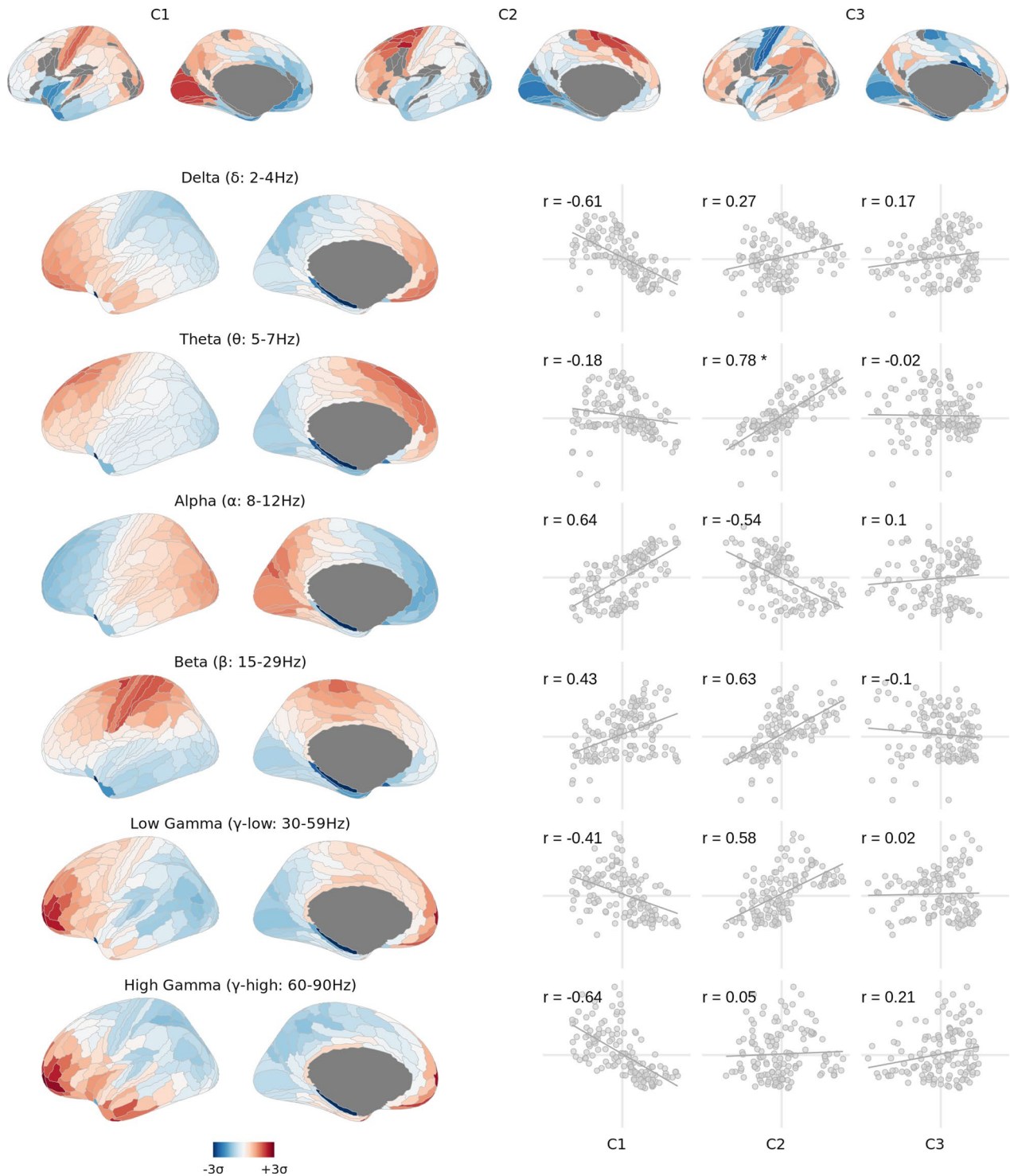
a, Gene weights from dimension reduction applied to group-averaged bulk RNA-seq measurements from 11 cortical regions in N = 54 healthy control brains from the PsychENCODE dataset³⁶ were correlated with gene weights from the components of the AHBA (derived by DME in the 180-region HCP-MMP parcellation), showing that the genetic profiles of AHBA C1, C2, and C3 were reproduced by PsychENCODE C1, C2, and C4, respectively (highlighted in green). **b**, Regional scores of PsychENCODE C1, C2 and C4 were also correlated with region scores of AHBA C1, C2 and C3, showing that the matching genetic profiles correspond to matching spatial expression patterns. **c**, Variance explained by the first five components of each dataset, showing that AHBA C3 and PsychENCODE C4 account for similar proportions of variance (6.5% and 7.1%, respectively). **d**, 1st row: Cortical maps of AHBA C1-C3 in the same 11 regions sampled in the PsychENCODE data. 2nd row: Cortical maps of PsychENCODE C1, C2, and C4 demonstrating their spatial similarity to AHBA

C1-C3. 3rd row: Gene weights from the PsychENCODE healthy control data were projected onto transcriptional data of cases with autism spectrum disorder (ASD; N = 58) from the same dataset, demonstrating lower regional expression at the positive (red) pole of each component in the ASD cases compared to healthy controls. **e**, Distributions of regional scores for C1, C2 and C4, computed on group-average healthy controls as in **a-d** and projected to individual donor brains in the PsychENCODE dataset, demonstrating significant case-control differential expression for regions at the positive poles of C1-C3. T-tests of case-control differences were corrected for multiple comparisons across all 33 tests; boxplots represent the median, first, and third quartiles with whiskers showing 1.5* inter-quartile range; *, **, *** indicate FDR-corrected two-sided p-value < 0.05, 0.01, 0.001 respectively. Region names refer to the sampled Brodmann Areas (BA)³⁶: Visual = BA17, Temporal Pole = BA38, Somatosensory = BA3-1-2-5, Motor = BA4-6, Anterior Cingulate = BA24, Prefrontal = BA9, Broca's Area = BA44-45, Fusiform Gyrus = BA20-37, Auditory = BA41-42-22, Lateral Parietal = BA39-40, Dorsal Parietal = BA7.



Extended Data Fig. 4 | Higher-order components of cortical gene expression reflect anatomically relevant co-expression structure. **a**, C1-C3 were orthogonally aligned in anatomical space, as computed by the Pearson's correlations of the regional scores with the XYZ coordinates of the region centroids: C1 and C2 were both aligned with the anterior-to-posterior (y) and ventral-to-dorsal (z) plane, but with opposite signs along the anterior-to-posterior axis, while only C3 was aligned to the medial-lateral (x) axis. The middle panel represents these alignments as vectors in 3D space. The right-hand upper

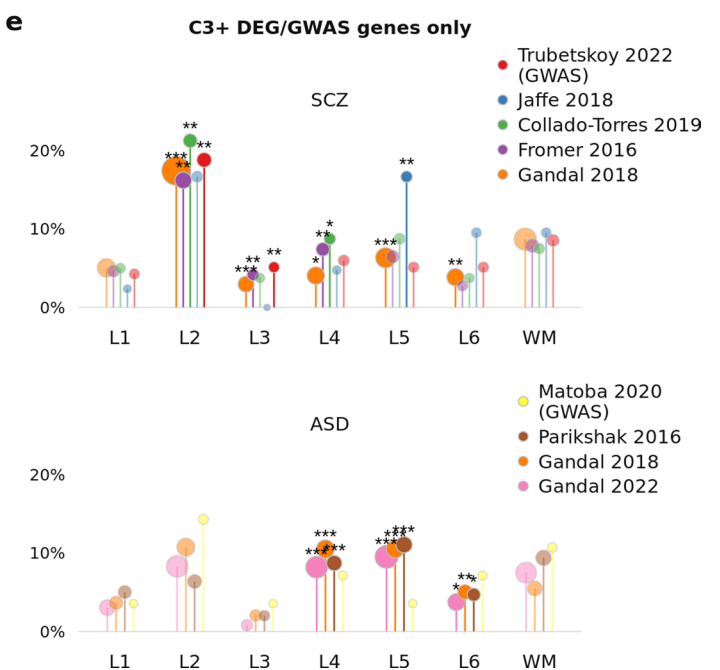
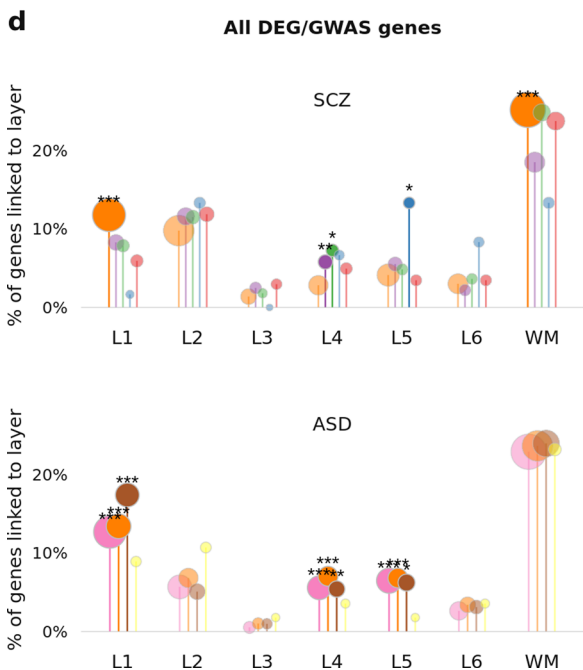
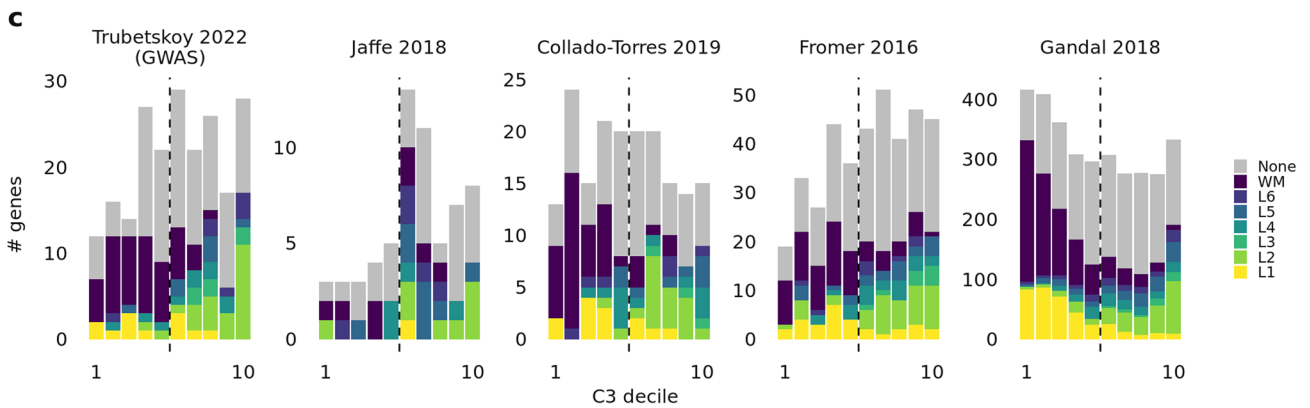
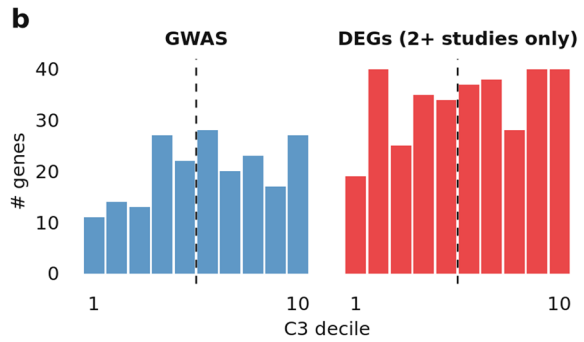
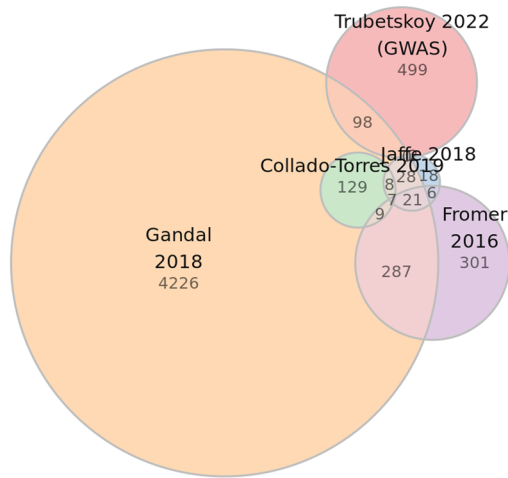
table shows the correlations of C1-C3 with each anatomical axis, and the lower table shows the angle in degrees between the vectors, showing that C1-C3 are orthogonal. **b**, Co-expression matrices computed by Pearson's correlations of gene expression between brain regions, computed with and without regressing out the first component C1, and annotated by the major cortical lobes as defined in the HCP-MMP parcellation³². This further demonstrates that the gene co-expression structure captured by C2 and C3 (that is, the residual variation beyond C1) is anatomically relevant.



Extended Data Fig. 5 | Transcriptional components were distinctively associated with the regional power of canonical brain oscillation frequencies. Several MEG power bands⁵⁵ were highly correlated ($|r| > 0.6$) with

C1 (delta, alpha, high-gamma) and C2 (beta, theta), although only the theta association to C2 survived FDR correction of the spin-test p-values ($r = 0.78$, $FDR_{spin} = 0.05$). No MEG band was aligned with C3.

a SCZ DEG/GWAS studies



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | C3 reveals shared biology across inconsistent postmortem brain RNA-seq studies of differentially expressed genes (DEGs) in schizophrenia. **a**, Euler diagram demonstrating the relative lack of overlap of genes linked to schizophrenia in four independent RNA-seq postmortem brain studies, as well as the latest GWAS study. **b**, Histogram of the schizophrenia GWAS and consensus DEG genes by C3 decile. The skew of the histograms towards higher C3 deciles reflects the significant enrichment of both non-overlapping gene sets, as in Fig. 4c,d. **c**, Histograms of the schizophrenia GWAS and DEG genes from each separate study by C3 decile, coloured by cortical layer where the gene was identified as a marker gene³⁸. L2 genes are distinctly clustered towards the C3+ pole, while L1 and WM genes are clustered towards C3-. **d**, For schizophrenia and ASD, enrichments of the GWAS/DEG genes from each

separate study for marker genes of cortical layers, showing that no consistent significant enrichments are found across the entire gene sets for studies of either disorder. **e**, Enrichments as in d, except for only genes positively weighted in C3 (corresponding to the right-hand five deciles of each histogram in panel c). For schizophrenia, significant enrichments for L2 and L3 are observed for three of the four DEG studies, as well as the GWAS study. No such enrichments were observed for ASD, demonstrating that C3 reveals convergent biology across otherwise inconsistent results specifically for schizophrenia. Significance was tested by one-sided Fisher's exact test and corrected for multiple comparisons across all tests in each panel. *, **, *** indicate FDR-corrected one-sided p-value < 0.05, 0.01, 0.001 respectively.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|--|
| Data collection | No software was used to collect data in this study. |
| Data analysis | Analyses were performed with Python v3.10.5 and R v.2.2. Key python packages include: abagen==0.1.3, brainspace==0.1.10, neuromaps==0.0.3. Full details of all packages, a Dockerfile and link to docker image, and all code used for these analyses are publicly available at https://github.com/richardajdear/AHBA_gradients . |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Gene expression datasets used are all publicly available:

- The Allen Human Brain Atlas is available at <http://human.brain-map.org>, and individual donor HCP-MMP parcellation images at <https://figshare.com/articles/>

dataset/AHBAdata/6852911.

- The BrainSpan Atlas is available at <https://www.brainspan.org/>.
- The Allen Human Cell Atlas is available at <https://portal.brain-map.org/atlas-and-data/rnaseq>.
- The PsychENCODE dataset is available at <https://github.com/dhglab/Broad-transcriptomic-dysregulation-across-the-cerebral-cortex-in-ASD>.

Neuroimaging maps of healthy brain features are available in the neuromaps package (<https://github.com/netneurolab/neuromaps>). For convenience all brain maps used are provided in Supplementary Table 3-4. Gene lists used for enrichment analyses were all obtained from prior publications as detailed in Methods.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	Sex-and gender-based analysis was not performed due to lack of sufficient sample size in the AHBA dataset (n=6, 1 female).
Population characteristics	The AHBA dataset brains have mean age of 42.5 (SD 13.4).
Recruitment	Recruitment of the AHBA data was as detailed on the Allen Institute website. Specifically, postmortem tissue from males and females between 18 – 68 years of age and no known history of neuropsychiatric or neurological conditions ('control' cases) were eligible for inclusion.
Ethics oversight	Ethics oversight was performed by the Allen Institute for Brain Science.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed as the AHBA dataset comprises only 6 brains; instead, our analysis proposed analysis by disjoint triplets as a proxy for generalisability. To ensure that this sample size was sufficient for the analysis performed, we reproduced and validated the results in three other transcriptomic datasets representing the three main dimensions of variability: developmental variability (BrainSpan), cell-type variability (Allen Cell Atlas), and inter-individual variability (PsychENCODE).
Data exclusions	In the BrainSpan dataset, the 7 donor brains from which fewer than 4 regions were sampled were excluded from analysis due to insufficient spatial representation to normalise by donor, leaving 35 donor brains. In the PsychENCODE dataset, samples with RIN < 3 were excluded from analysis.
Replication	While replication was not possible as the AHBA is the only transcriptomic data of the entire human cortex with high spatial resolution, we performed one replication of the three components in external RNAseq data from the PsychENCODE consortium.
Randomization	Randomization was not relevant to our analysis as we were identifying generalisable transcriptomic patterns across all human brains, for which we proposed the disjoint triplets analysis as a proxy for generalisability.
Blinding	Blinding was not relevant to our study as we did not perform randomized experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging