Three components of human brain gene expression reflect normative developmental programmes with specific links to neurodevelopmental disorders

Richard Dear, Konrad Wagstyl, Jakob Seidlitz, Ross D. Markello, Aurina Arnatkevičiūtė, Kevin M. Anderson, Richard A.I. Bethlehem, Lifespan Brain Chart Consortium, Armin Raznahan, Edward T. Bullmore, Petra E. Vértes

Abstract

Human brain organisation emerges from the coordinated transcription of thousands of genes, and the first principal component (C1) of spatial whole genome expression was shown to reflect cortical hierarchy. Here, optimised processing of the Allen Human Brain Atlas revealed two new components of brain transcription, C2 and C3, which were distinctively enriched for neuronal, metabolic and immune processes, cell-types and cytoarchitecture, and genetic variants associated with intelligence. Using additional datasets (PsychENCODE, Allen Cell Atlas, and BrainSpan), we found that C1-C3 represent generalisable transcriptional programmes that are coordinated within cells, and differentially phased during foetal and postnatal development. Autism spectrum disorder and schizophrenia were specifically associated with C1/C2 and C3, respectively, across neuroimaging, gene expression, and genome-wide association studies. Evidence converged especially in support of C3 as a normative transcriptional programme for adolescent brain development, which can lead to atypical supragranular brain connectivity in people at high genetic risk for schizophrenia.

Introduction

How does the complex anatomical and functional organisation of the human brain develop from the expression of over twenty thousand genes [1]? And how does this process go awry in neurodevelopmental disorders? In the last 10 years, whole-brain, whole-genome transcriptional atlases, such as the Allen Human Brain Atlas (AHBA) [2], have suggested that healthy brain organisation may depend on "transcriptional programmes" representing the coordinated expression of large numbers of genes over development [3–7].

In 2012, Hawrylycz *et al*. showed that principal components of the AHBA dataset capture distinct features of cortical anatomy [2]. 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- and T2-weighted MRI signals, a putative marker of cortical myelination [8]. These discoveries echoed prior findings from studies of embryonic development of chick, mouse and human brains where spatially patterned transcriptional gradients have been shown to organise neurodevelopmental processes such as areal differentiation, axonal projection, and cortical lamination [6,11–13]. Single-cell RNAseq data has also revealed an anterior-to-posterior gradient in the gene expression of inhibitory interneurons, which is conserved across multiple species including humans [14]. It is therefore likely that the principal component of gene expression in the adult human cortex represents a transcriptional programme key to its normative development.

However, it is not so clear that C1 is the only component of spatially patterned and neurodevelopmentally coordinated gene expression in the human brain. Hawrylycz *et al.* 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 [2] (**Supplementary Fig. S1**). Later, Goyal *et al.* 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 [15]. To our knowledge, more recent studies using all available AHBA data have reliably found only C1 [8,16]. This first component has been linked to a general "sensorimotor-association axis" of brain organisation [10] derived from several macro-scale brain phenotypes, including among others the principal gradient of functional connectivity [17], maps of brain metabolism and blood flow [15], and the map of human cortical expansion compared to other

primates [18]. Although it is parsimonious to assume that such diverse brain phenotypes could all be determined by a single transcriptional programme, it seems more realistic to expect that multiple transcriptional programmes are important for human brain development, as is generally the case for brain development in other species [19].

Here we present two higher-order components of human cortical gene expression, C2 and C3, that likely represent additional transcriptional programmes distinct from the C1 component already reliably described [8]. These higher-order components only emerged when optimised 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 macro-scale brain maps. Leveraging independent ribonucleic acid sequencing (RNA-seq) datasets on single-cell and developmental gene expression, we further demonstrated that all three components are generalisable to other datasets, representative of coordinated transcription within cells of the same class, and dynamically differentiated over the course of foetal, childhood and adolescent brain development. Finally, by triangulating evidence across case-control neuroimaging, differential gene expression, and genome-wide association studies (GWAS), we demonstrated that components C1 and C2 are specifically associated with autism spectrum disorder (ASD), and C3 with schizophrenia.

Results

Three generalisable components of human cortical gene expression were enriched for biological processes, cytoarchitecture, and cognitive capacity

We first applied PCA to the entire AHBA dataset of 6 adult brains (mean age at death = 42 years) [2]. Microarray measurements of relative messenger RNA levels were processed to represent mean expression of ~16,000 genes at each of the 180 regions of the left hemispheric cortex defined by the HCP-MMP parcellation [20,21] (**Methods**). We found that higher-order components (C2, C3) estimated by PCA of the resulting {180 × 16,000} data matrix were not statistically robust to sampling variation of the six donor brains, with low generalisability, 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 generalisability of higher order components. First, we optimised the trade-off involved in excluding noisy data – by filtering spatially inconsistent genes (with low differential stability [22]) and under-sampled brain regions – while seeking to maximise the anatomic and genomic scope of the data matrix (**Extended Data Fig. 1**). Second, we used the non-linear 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 while 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 generalisability 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 generalisability of the first three components was substantially increased, i.e., $g_{C1} = 0.97$, $g_{C2} = 0.72$, $g_{C3} = 0.65$, while the generalisability of even higher-order components remained low, e.g., $g_{C4} = 0.28$ (**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 higher generalisability indicating 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 controls [23] (regional correlation $r_{C1} = 0.85$, $r_{C2} = 0.75$, $r_{C3} = 0.73$; see **Extended Data Fig. 3** and **Supplementary Table S5**).

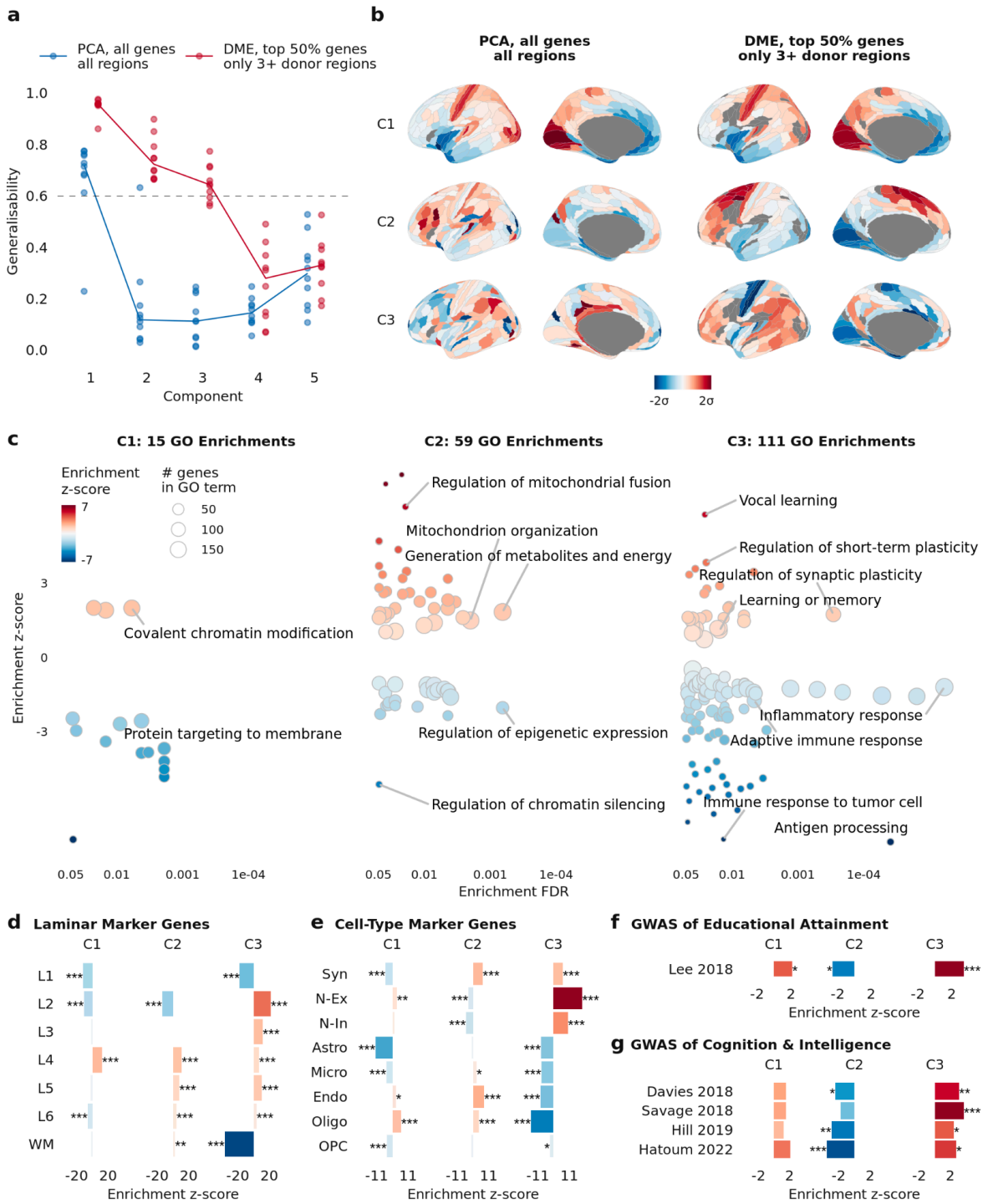


***Figure 1**: Three generalisable components of human cortical gene expression were enriched for biological processes, cytoarchitecture, and cognitive capacity. **a**, To test whether the higher-order components of gene expression were statistically robust and likely generalisable to the population, or merely partitions of noise in*

the available data, we estimated generalisability by splitting the AHBA dataset into two disjoint triplets of three brains, applying PCA to each triplet and correlating the resulting matched components (C1, C2, C3…) between all possible (10) pairs of triplets (Methods). For each component, the median absolute correlation over all triplet pairs was a proxy for its generalisability, g*. Using PCA and previously published best practices for processing the AHBA dataset [20,21], generalisability decreased markedly beyond the first component: $g_{C1} = 0.78$, $g_{C2} = 0.09$, $g_{C3} = 0.14$ (**Fig. 1a;** blue points represent median absolute correlation for each of 10 triplet pairs). Using diffusion map embedding (DME) with the top 50% most stable genes and the 137 regions with data available from at least three brains, the generalisability of the first three components was substantially increased: $g_{C1} = 0.97$, $g_{C2} = 0.72$, $g_{C3} = 0.65$ (**Fig 1a;** red points represent g for each of 10 triplet pairs), although the generalisability of even higher-order components remained low, e.g. $g_{C4} = 0.28$. **b**, Cortical maps of brain regional scores of components C1-C3 estimated by PCA on the whole AHBA dataset and DME on filtered AHBA dataset. Quantifying smoothness with Moran's I metric for spatial autocorrelation [24], the first three components estimated by PCA had I = 0.50, 0.09 and 0.07, respectively, whereas the C1, C2, and C3 components estimated by DME were more spatially smooth: $I_{C1} = 0.48$, $I_{C2} = 0.58$, and $I_{C3} = 0.21$. **c**, Gene Ontology Biological Process enrichments for C1-C3 showing that the number of enrichments was greater for higher-order components, illustrating that they are more biologically specific. C2-positive genes were enriched for metabolism while C2-negative genes were enriched for regulatory processes; C3-positive genes were enriched for synaptic plasticity and learning, while 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) [25]. Significance computed as percentile of mean weight of marker gene set relative to randomly permuted gene weights, and FDR-corrected across all 24 tests. *, **, *** respectively indicate FDR-corrected p-value < 0.05, 0.01, 0.001. **e**, C1-C3 were also distinctively enriched for marker genes of cell types and synapses [26,27]. Significance tested as in c. **f**, C2 and C3 (but not C1) were significantly enriched for genes mapped to common variation in intelligence and cognition across four independent genome-wide association studies (GWAS) [28–31]. Significance tested as in d, corrected for multiple comparisons across all four studies. **g**, All three components, C1-C3, were significantly enriched for genes mapped to common variants associated with educational attainment in prior GWAS data [32]. Significance tested as in d. Code and data for these analyses are openly provided (**Supplementary Table 1, Methods**).*

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. S2**). 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 [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*, *CCK*) and glutamatergic neuronal genes (*GRIN* and *GABRA*) were strongly weighted with opposite signs (positive or negative) on C1 (**Supplementary Fig. S3**).

For genes positively weighted on C2, 23 of the 36 FDR-significant enrichments were for metabolic processes, and for negatively-weighted genes, 19 of the 23 enrichments were for epigenetic processes (**Fig. 1c, Supplementary Table 2**). Whereas, for genes positively-weighted on C3, 19 of 27 FDR-significant enrichments were related to synaptic plasticity or learning, and for negatively-weighted genes, 33 of 84 enrichments were related to the immune system. We further analysed enrichment for genes identified as markers of specific cortical layers [25] (**Fig. 1e**) and cell types [33] (**Fig. 1f**), and in each case observed distinct enrichment profiles for C1-C3. For example, genes positively weighted on C3 were enriched for neuronal and synaptic marker genes and at cortical layers 2 and 3 (L2, L3), whereas negatively weighted genes 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 attainment (**Fig. 1f**). Across four independent GWAS studies of intelligence and cognition [28–31], genes strongly weighted on C1 were not significantly enriched, but genes strongly negatively-weighted on C2 were enriched for genetic variants associated with intelligence in three of the four studies, and genes strongly positively-weighted on C3 were enriched for genes identified by all four prior GWAS studies of intelligence (**Fig. 1g**).

Neuroimaging and macro-scale maps of brain structure, function, and development were distinctively co-located with three components of cortical gene expression

Prior work has linked gene transcription to a multimodal "sensorimotor-association axis" (S-A axis) [10] of brain organisation, defined as the composite of 10 brain maps, comprising the first principal component of gene expression (C1) and 9 other MRI or PET neuroimaging maps that were selected to differentiate sensorimotor and association cortices. We first aimed to build on this work by analysing the correlation matrix of the same set of 9 brain maps together with the three transcriptional components derived from DME of the filtered AHBA dataset. Data-driven cluster analysis of this {12 x 12} correlation matrix identified three phenotypic clusters, each nucleated by one of the DME-derived components (**Fig. 2a, Methods**). C1 was clustered together with 2 MRI maps, representing the T1w/T2w myelination marker [34] and cortical thickness [35]; C2 was clustered with 5 maps representing aerobic glycolysis [36], cerebral blood flow [37], cortical expansion in humans relative to non-human primates [18], inter-areal allometric scaling [38] and external pyramidal cell density [39]; and C3 was clustered with 2 maps, representing the principal gradient of fMRI connectivity [17] and first principal component of cognitive terms meta-analysed by Neurosynth [40]. While some maps were specifically correlated to one component, e.g. aerobic glycolysis $r_{C2} = 0.66$ ($p_{spin} = 0.004$, FDR < 5%), others were moderately correlated with multiple transcriptional components, e.g., for cerebral blood flow: $r_{C1} = 0.25$, $r_{C2} = 0.28$ and $r_{C3} = 0.33$. This clustering analysis suggests that it is overly parsimonious to align all 9 neuroimaging phenotypes with just one transcriptional component (C1) as part of a singular sensorimotor-association cortical axis.

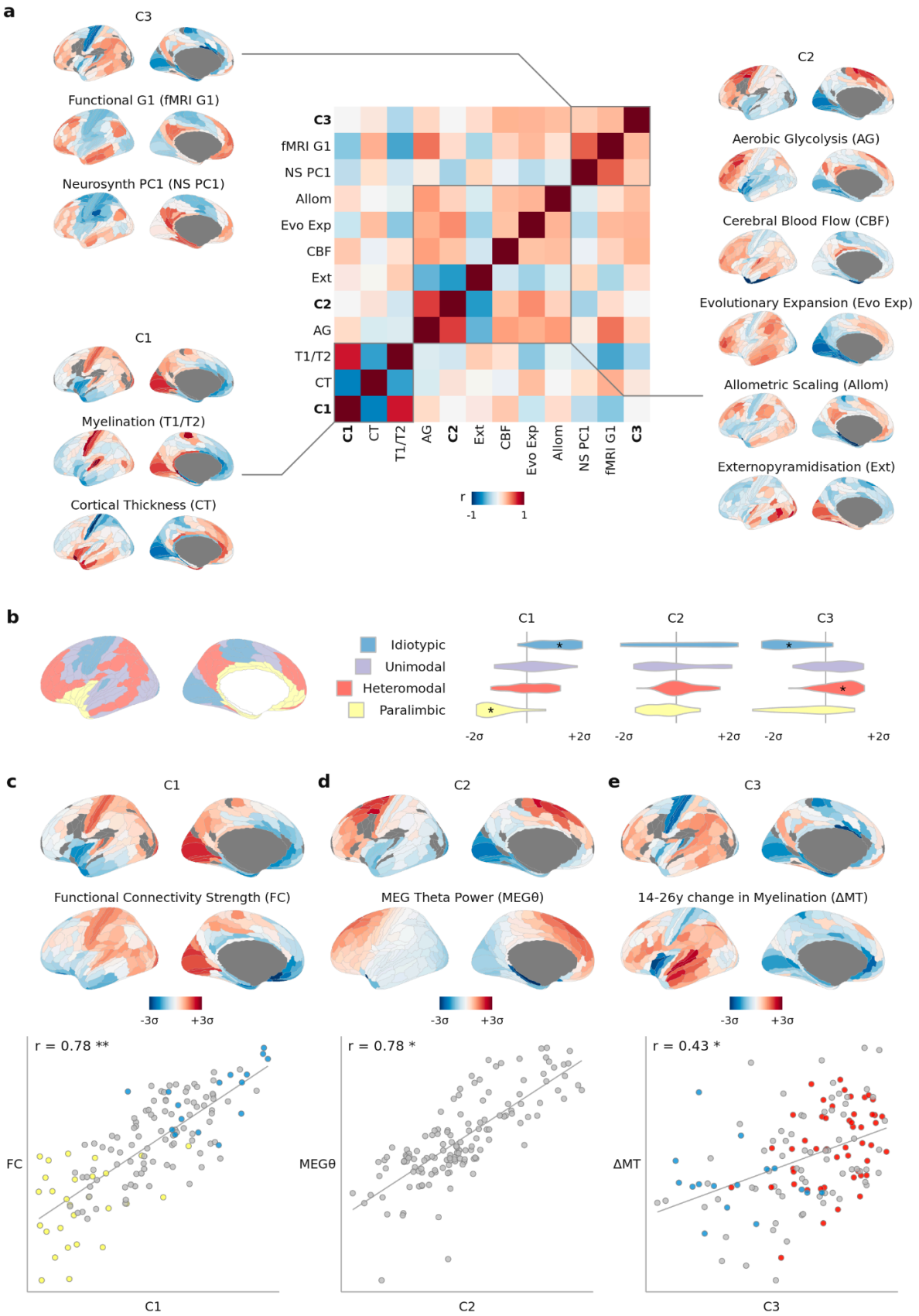

a
C3
Functional G1 (fMRI G1)
Neurosynth PC1 (NS PC1)
C1
Myelination (T1/T2)
Cortical Thickness (CT)
C3
fMRI G1
NS PC1
Allom
Evo Exp
CBF
Ext
C2
AG
T1/T2
CT
C1
r
-1
1
C2
Aerobic Glycolysis (AG)
Cerebral Blood Flow (CBF)
Evolutionary Expansion (Evo Exp)
Allometric Scaling (Allom)
Externopyramidisation (Ext)
b
Idiotypic
Unimodal
Heteromodal
Paralimbic
C1
C2
C3
-2σ
+2σ
c
C1
Functional Connectivity Strength (FC)
-3σ
+3σ
r = 0.78 **
FC
C1
d
C2
MEG Theta Power (MEGθ)
r = 0.78 *
MEGθ
C2
e
C3
14-26y change in Myelination (ΔMT)
r = 0.43 *
ΔMT
C3

***Figure 2**: **Neuroimaging and macro-scale 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- and physiologically-derived maps that Sydnor et al. combined with C1 to define a 'sensorimotor-association axis' of brain organisation [10]. 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 organisation. **b**, Distributions of regional scores of C1-C3 in histologically-defined regions of laminar cytoarchitecture [41]. C1 distinguished idiotypic and paralimbic regions, while C3 distinguished idiotypic and heteromodal regions. * indicates p_{spin} < 0.05, where p-value was computed as the percentile of the mean z-score relative to null spin permutations and has been adjusted for multiple comparisons. **c**, Degree of fMRI functional connectivity [42,43] was closely aligned to C1. *, **, *** indicate FDR_{spin} < 0.05, 0.01, 0.001 when corrected for all maps in panels c-d being compared with all of C1-C3. Blue/yellow highlighted points correspond to idiotypic/paralimbic cytoarchitectural regions as in b. **d**, MEG-derived theta power [44] was significantly aligned to C2. Significance assessed as in b. **e**, Regional change in myelination over adolescence [45,46] was significantly aligned to C3. Significance assessed as in b. Blue/red highlighted points correspond to idiotypic/heteromodal cytoarchitectural regions as in b.*

We also found that each of the three transcriptional components was associated with a wider range of cellular, functional and developmental phenotypes than the 9 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 differentiation [41] were co-located with C1 and C3, but not C2 (ANOVA, p < 0.001; **Fig. 2b**). In functional MRI and magnetoencephalography (MEG) data, we found that weighted nodal degree of cortical regions in an fMRI network [42,43] was strongly correlated with C1 (r_{C1} = 0.78, p_{spin} < 0.001, FDR = 5%, **Fig. 2c**) but not C2 or C3 (r_{C2} = -0.01, r_{C3} = 0.00); across all canonical frequency intervals of MEG data [44], an FDR-significant association was observed between theta band (4-7 Hz) oscillations and C2 (r_{C2} = 0.78, p_{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. And in support of the hypothetical prediction that adult brain transcriptional programmes are neurodevelopmentally relevant, we found that a prior map of adolescent cortical myelination, as measured by change in magnetisation transfer between 14-24 years (ΔMT) [45,46], was significantly co-located with C3 (r_{C3} = 0.43, p_{spin} = 0.009, **Fig. 2e**) but not C1 or C2 (r_{C2} = 0.17, r_{C3} = 0.15).

Transcriptional components represent intracellular coordination of gene expression programmes with distinct developmental trajectories

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

First, for single-cell RNA-seq data comprising 50,000 nuclei sampled from five cortical regions of three donor brains [47], 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, e.g. 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. S5**). 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 even stronger when the single-cell data were stratified by sub-classes of cells in specific cortical layers, e.g., L2 VIP-expressing interneurons (**Fig. 3a inset**). In 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.

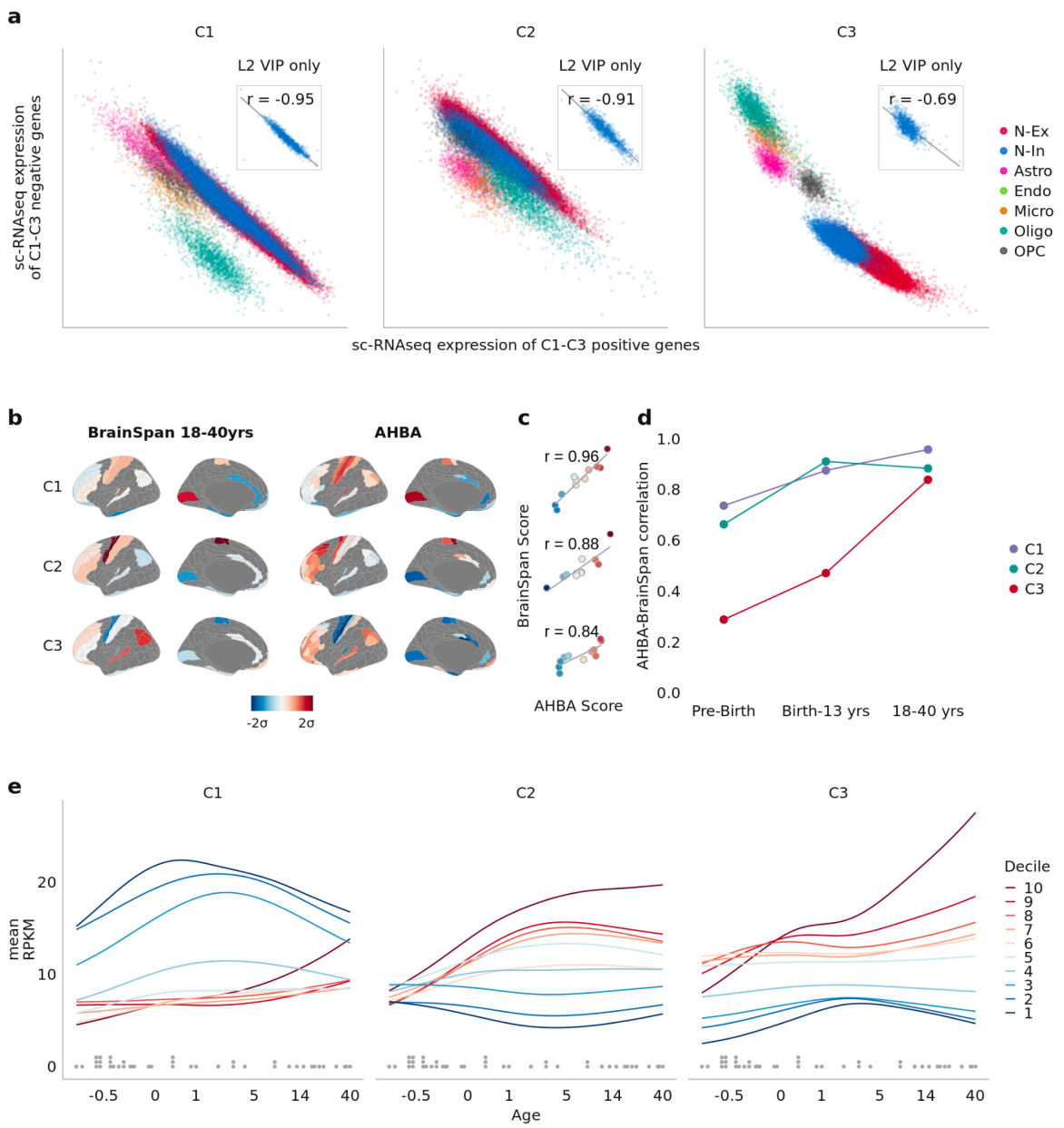


***Figure 3**: **Transcriptional components represent intracellular coordination of gene expression programmes with distinct developmental trajectories. a**, For each of ~50,000 single-cell RNAseq samples, weighted average expression of the positive and negative genes of each AHBA transcriptional component C1-C3 (Methods). Samples are coloured by cell-type, demonstrating that genes positively and negatively-weighted on C1-C3 have correlated expression within each major class of brain cells: N-Ex, excitatory neurons; N-In, inhibitory neurons; Astro, astrocytes; Endo, endothelial cells; Micro, microglia; Oligo, oligodendrocytes; and OPC, oligodendrocyte precursor cells. Inset, a subset of samples from Layer 2 VIP interneurons (one of several cells included in the N-In class), 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 on components C1-C3 for each of 11 regions with transcriptional data available in the BrainSpan cohort of adult brains (left column) and C1-C3 component scores for the matching subset of regions with transcriptional data available in the AHBA (right*

column). ***c****, Scatter plots of matched regional C1-C3 component scores (from b) demonstrating that the three transcriptional components first defined by analysis of AHBA data are consistently expressed in the BrainSpan dataset.* ***d****, Plots of the correlation between AHBA C1-C3 scores and BrainSpan C1-C3 scores (y-axis) calculated for each of 3 age-defined subsets of the BrainSpan dataset (prebirth, birth-13 years, and 18-40 years). C1 and C2 component scores were strongly correlated between datasets for all epochs of the BrainSpan dataset, whereas C3 component scores were strongly correlated between datasets only for the 18-40y epoch of BrainSpan. This indicates that C1 and C2 components were expressed in nearly adult form 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 to 40 years; x-axis, log scale) were estimated by generalised additive models for each gene and then averaged within each decile of gene weights for each of the three components; fitted lines are colour-coded by decile, dark blue representing the first (most negatively weighted) decile of genes, and dark red representing the 10th (most positively weighted) 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 post-mortem ages of the donor brains used to compute the curves. RPKM: reads per kilobase million.*

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 between 4 and 14 cortical regions for each of 35 donor brains ranging in age from -0.5 years (mid-gestation) to 40 postnatal years [6]. We first asked if the gene weights for each of the components derived from the filtered AHBA dataset would define similar cortical 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**, **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 generalisable 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-13 years, N = 14). We observed that for C1 and C2, AHBA component scores were almost as highly correlated with BrainSpan component scores in foetal (prebirth) and childhood (birth-13 years) brains as in the adult (18-40 years) brains (birth-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 foetal and childhood subsets of the BrainSpan dataset (prebirth, r_{C3} = 0.29; birth-13 years, r_{C3} = 0.47). These results suggested 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, modelling the non-linear developmental trajectories of each gene over the age range -0.5 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 adult life (**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.

Genetics, transcriptomics, and neuroimaging of autism and schizophrenia were consistently and specifically linked to normative transcriptional programmes

We explored the clinical relevance of C1-C3 by analysis of additional independent GWAS and case-control studies of neuroimaging and gene expression in autism spectrum disorder (ASD), major depressive disorder (MDD), and schizophrenia.

First, we leveraged the BrainChart neuroimaging dataset of >125,000 MRI scans [48], 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 life-cycle (**Fig. 4a**). Using the relatively coarse-grained Desikan-Killiany parcellation of 34 cortical regions necessitated by alignment with this dataset (**Fig. 4b**, **Methods**), we found that cortical shrinkage in ASD was significantly associated with both C1 and C2 (r_{C1} = 0.49, P_{spin} = 0.0002, FDR < 5%; r_{C2} = -0.28, P_{spin} = 0.0006, FDR < 5%), while shrinkage in schizophrenia was specifically associated with C3 (r_{C3} = 0.43, p_{spin} = 0.008, FDR < 5%) (**Fig. 4b**).

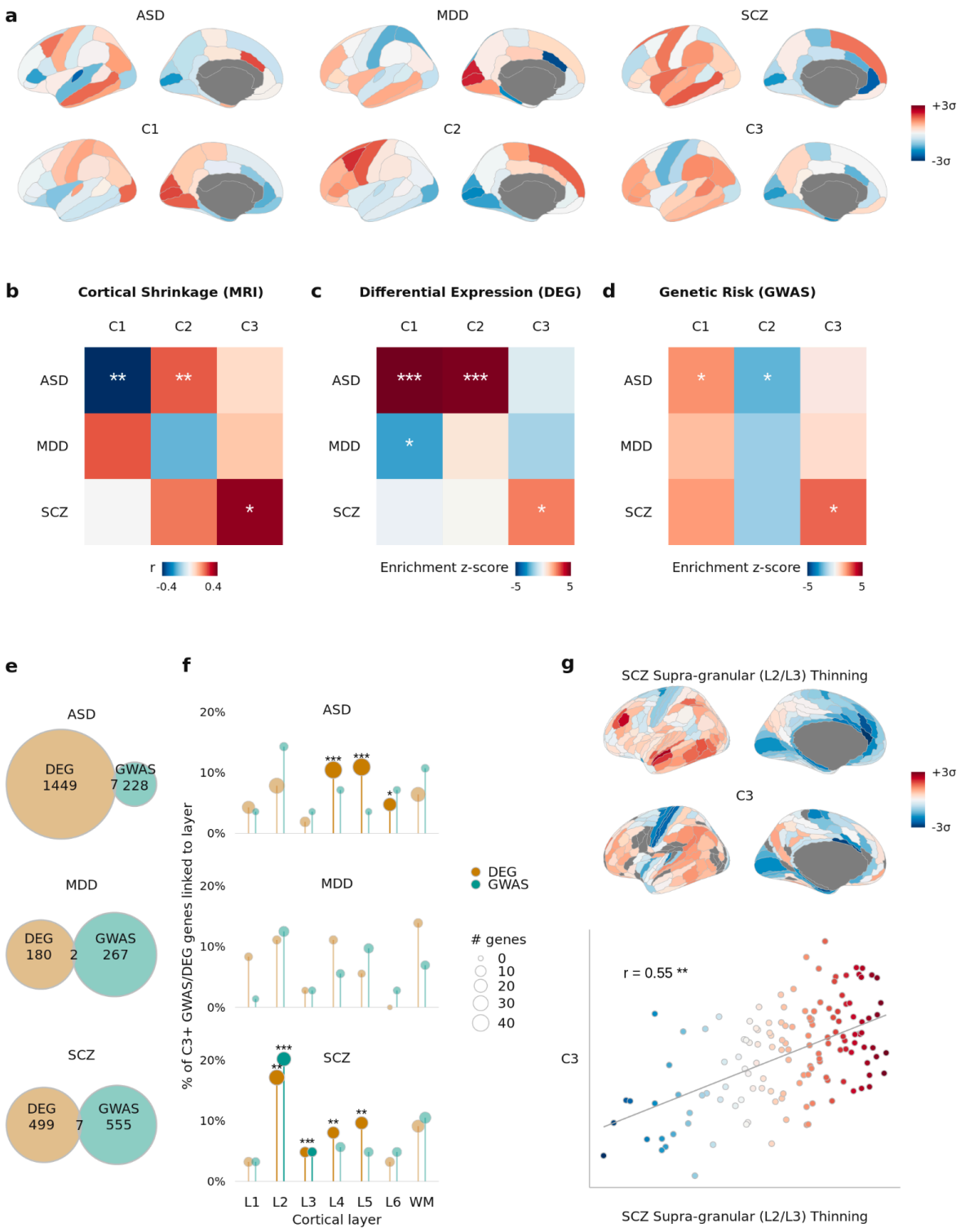


***Figure 4**: Genetics, transcriptomics, and neuroimaging of autism and schizophrenia were consistently and specifically linked to normative transcriptional programmes. **a**, First row: cortical volume shrinkage in autism spectrum disorder (ASD), major depressive disorder (MDD), and schizophrenia (SCZ) cases. Red indicates greater shrinkage, computed as z-scores of centiles from normative modelling of >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 tested by spatially autocorrelated spin permutations, and corrected for multiple comparisons: *, **, *** indicate FDR_{spin} < 0.05, 0.01, 0.001 respectively.* ***c****, Enrichments in C1-C3 for differentially expressed genes (DEGs) in postmortem brain tissue of donors with ASD, MDD, and SCZ compared to healthy controls. DEGs were consensus lists of genes identified in two or more studies (****Methods****). Significance assessed as percentile of mean weight of DEGs in each component relative to randomly permuted gene weights, and corrected for multiple comparisons: *, **, *** indicate FDR < 0.05, 0.01, 0.001 respectively.* ***d****, Enrichment in C1-C3 for GWAS risk genes for ASD[49], MDD[50], and SCZ[51] aligned with both spatial associations of volume changes and enrichments for DEGs. Significance tested as in c.* ***e,*** *Venn diagrams showing the lack of overlap of DEGs and GWAS risk genes reported by the primary studies summarised in panels c and d.* ***f****, DEGs and GWAS risk genes for each disorder were filtered for only C3-positive genes, then tested for enrichment with marker genes for each cortical layer [25]. Significance was tested by 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 (panel e, see* ***Extended Data Fig. 6*** *for more detail).* ***g,*** *Convergent with the focus on L2/L3 from enrichment analysis of gene weights, a whole cortex map of supragranular-specific, L2-predominant cortical thinning in schizophrenia [52] was specifically co-located with C3. In the brain maps, red represents cortical areas with greatest thinning of supragranular layers of cortex in a prior case-control MRI study of schizophrenia. The scatterplot shows the strong co-location between regional component C3 scores (y-axis) and case-control differences in supragranular cortical thickness (x-axis); each point represents a region in the HCP-MMP parcellation; color represents C3 score.*

Second, we compiled consensus lists of differentially expressed genes (DEGs) from RNA-seq measurements of dorsolateral prefrontal cortex in patients with ASD [23,53,54], MDD [55], or schizophrenia [54,56–59] (**Methods**). We found that genes differentially expressed in ASD were specifically enriched in both C1 and C2 (but not C3); whereas genes differentially expressed in schizophrenia were enriched in C3 (but not C1 or C2); and genes differentially enriched in MDD were enriched only in C1 (**Fig. 4b**). Analysis of case-control differences in spatial expression of C1-C3 at 11 cortical regions for ASD cases compared to healthy controls demonstrated, as expected, 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 GWAS studies of ASD [49], MDD [50], and schizophrenia [51], we found that genetic variants significantly associated with ASD were enriched in both C1 and C2 (but not C3); whereas genes associated with schizophrenia were enriched in C3 (but not C1 or C2) (**Fig. 4d**). Genes associated with MDD were not significantly enriched in any transcriptional component. These associations were largely replicated using alternative methods (MAGMA [60] and H-MAGMA [61]) to test the association between GWAS-derived P-values for the association of each gene with ASD,

MDD or SCZ and the weights of each gene on the transcriptional components C1-C3 without requiring an explicit prioritisation of GWAS-associated genes (**Methods**, **Supplementary Fig. S6**). This pattern of results for autism and schizophrenia GWAS data evidently mirrored the pattern of prior 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 disorder [62] (**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 layers L2 and L3 (**Fig. 4g; Extended Data Fig. 6; Methods**). These shared laminar associations between the non-overlapping DEG and GWAS gene sets were only present when subsetting to C3-positive genes and were specific to schizophrenia (i.e. 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, we found that the cortical map of C3 was significantly co-located with an MRI-derived map of supragranular, L2/L3 predominant thinning in schizophrenia [52] (r_{C3} = 0.55, p_{spin} = 0.002, FDR < 1%, **Fig. 4g**).

Discussion

We have shown that the transcriptional architecture of the human brain includes at least three components of coordinated gene expression that are relevant to its normative organisation and development and to the emergence of neurodevelopmental disorders.

We used optimised data processing methods and leveraged multiple gene expression datasets to identify and validate two higher-order transcriptional components, C2 and C3. These components have not previously been robustly demonstrated, although in their initial study Hawrylycz *et al.* identified similar components to C1 and C2 by applying PCA to one of the six AHBA brains and filtering for only 1000 genes [2] (**Supplementary Fig. S1**). Each component was defined by signed weights of individual gene expression patterns, and by the spatial pattern shared by those genes

across the cortex. We therefore used two approaches to assess the biological relevance of these components: functional enrichment analysis of component gene weights, and co-location analysis of component brain maps with neuroimaging and other macro-scale brain phenotypes.

We found that specific, significant functional enrichments were much clearer for the higher-order components (C2, C3) than for the first component (C1). This reflects the fact that C2 and C3 represent coordinated expression across hundreds of genes, rather than the several thousand genes aligned to C1, and suggests that C2 and C3 play more specific biological roles than the C1 neuronal hierarchy [8]. Broadly, the positive pole of C2 gene weights was enriched for "metabolic" processes and its negative pole for "epigenetic" processes; whereas the positive pole of C3 gene weights was enriched for "synaptic" processes and its negative pole for "immune" processes (**Fig. 1c**). Notwithstanding their relatively specialised functional enrichments, 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 behaviour.

We also found that each of the three component brain maps was distinctively co-located with multiple independent neuroimaging or other macroscale brain phenotypes (**Fig. 2**). It was notable that these co-locations were often convergent with the gene enrichment results, triangulating the evidence for C2 as a metabolically specialised component and for C3 as a component specialised for synaptic and immune processes underpinning adolescent plasticity; see **Table 1**. For example, based on prior knowledge that theta oscillations are linked to intelligence and cognition [63–65] as well as to glucose metabolism [66], 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, prior knowledge implicates microglia and oligodendrocytes in the immune-mediated synaptic pruning and myelination that over adolescence gives rise to adult cognitive capacity [67–69], 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 study explicitly linked intelligence to myelination [29]. Together, these results expand on the proposal of a single "sensorimotor-association axis" [10,70] by demonstrating that macro-scale brain organisation emerges from multiple biologically-relevant transcriptional components.

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

Table 1: Summary of convergent results on the biological and clinical relevance of three human brain transcriptional programmes. *Each of three components of normative human brain gene expression (C1, C2, 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 macro-scale brain phenotypes, in healthy brain samples (normative) and in studies of neurodevelopmental disorders (atypical). Each row summarises results for a distinct gene enrichment analysis (italicised, grey) or spatial co-location analysis (plain font, white). Red text indicates enrichment in positively weighted genes on each component or positive spatial correlation with regional component scores; blue text indicates enrichment in negatively weighted genes or negative spatial correlation with regional component scores.*

However, the discovery of these biologically-relevant, higher-order transcriptional components in the AHBA dataset raised further questions: i) can the components be generalised to other human brain gene expression datasets; ii) do they reflect intracellular coordination of gene expression or regional differences in the proportion of different cell types; iii) when do the components emerge during normative brain development; and (iv) how do they intersect with neurodevelopmental disorders? We addressed these questions directly using additional RNA-seq datasets.

First, we found that the components derived from the adult AHBA brains (N = 6) could be replicated in an independent dataset (PsychENCODE) comprising bulk RNA-seq measurements on a smaller number of brain regions (11) sampled in a larger number of healthy controls (N = 54; **Extended Data Fig. 3** and **Supplementary Table S5**). Second, we analysed measurements of gene expression in single cells, e.g., individual neurons and glial cells, sampled from an independent cohort of post-mortem brains. We found that genes positively or negatively weighted on components derived from the AHBA bulk tissue samples had consistently coupled co-expression across single cells. This indicated that components C1-C3 represent transcriptional programmes coordinated at the intracellular level, and not merely inter-areal variation in the proportion of different cell types. Third, we found that C1-C3 have differentially phased developmental trajectories of expression, e.g. 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. Finally, we established that these transcriptional programmes are not only critical for healthy brain development but, as might be expected, are also implicated in the pathogenesis of neurodevelopmental disorders.

The pattern of results was strikingly convergent across multiple datasets and disorders: C1 and C2 were both enriched for genes implicated by both GWAS and DEG data on autism spectrum disorder, whereas C3 was specifically enriched for genes implicated by both GWAS and DEG data on schizophrenia (**Table 1**). We observed a similar pattern of co-location between C1-C3 maps and MRI phenotypes: developmentally normalised scores on reduced cortical volume in ASD were spatially correlated with maps of C1 and C2 (**Fig. 4a-b**), whereas the map of C3 was spatially correlated with both full-thickness cortical shrinkage (**Fig. 4a-b**) and supra-granular layer-specific cortical thinning (**Fig. 4g**) in schizophrenia. In contrast, there was no evidence for enrichment of C1-C3 by genes associated with increased risk of Alzheimer's disease (derived from SNP-level data reported in prior GWAS studies [71], **Supplementary Fig. S6**). An intuitive generalisation of these results is that the developmental processes which 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.

Overall, our results were strongly supportive of the motivating hypothesis that the transcriptional architecture of the human brain represents developmental programmes crucial both to the brain's healthy organisation for adaptive function, and to the emergence of neurodevelopmental disorders. For example, when interpreting C3 as a transcriptional programme mediating adolescent plasticity (**Table 1**), our finding that C3 represents coupled (anti-correlated) transcription of synapse- and immune-related genes within cells (as well as between different cell types; **Fig. 3a**) is consistent with prior work indicating that the neuronal expression of immune-related, typically glial genes can play a mechanistic role in synaptic pruning [72]; and, vice-versa, that neuronal genes associated with synapse and circuit development can also be expressed in glial cells [73]. While atypical synaptic pruning has long been hypothesised to be a mechanistic cause of schizophrenia [74–76], prior results on the biology of schizophrenia have shown limited consistency, both between the primary data modalities of GWAS, postmortem expression, and neuroimaging [77,78], and even between DEG studies [62]. Here, we demonstrate that the C3 transcriptional programme offers a unifying link between these disparate prior results. In addition to the non-overlapping schizophrenia GWAS and DEG gene sets both being enriched with C3 (**Fig. 4c-e**), we found that focusing on only C3-positive genes revealed a common enrichment for L2 and L3 marker genes in the schizophrenia-associated genes identified by both prior GWAS and DEG studies (**Fig. 4f**). L2 and L3 (supragranular) layers have dense cortico-cortical connections and are expanded in humans relative to other species [79–81], mature latest in development [82,83], and have been linked to intelligence [84], again convergent with the emergence of C3 in adolescence (**Fig. 3b-c**) and its links to intelligence (**Fig. 1f-g**). Supragranular layers have also been previously implicated in schizophrenia [85–89]; accordingly, we found a strong correlation between C3 and a map of supragranular-specific cortical thinning in schizophrenia [52] (**Fig. 4g**). This triangulation of evidence strongly suggests that the third major component of the brain's gene expression architecture represents the transcriptional programme coordinating the normative, neuro-immune processes of synaptic pruning and myelination in adolescence, 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 [45].

Note that these associations between C3 and schizophrenia (or between C1/2 and ASD) are conceptually distinct from prior studies that have used the AHBA to derive gene sets correlated with disorder-related MRI phenotypes [26,90–94]. A disorder-first, "imaging transcriptomics" [95–97] approach is

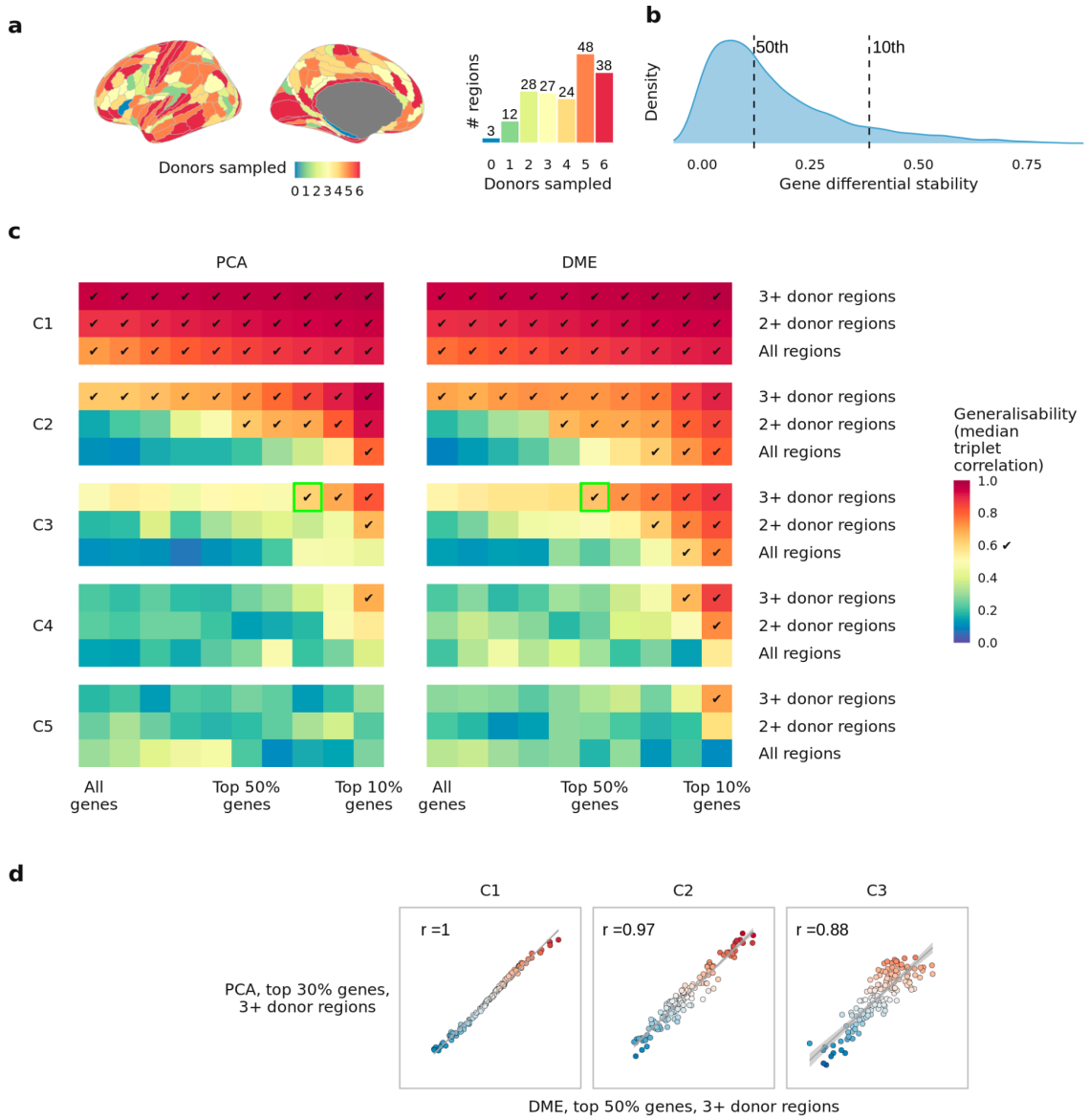
susceptible to identifying genes whose co-location with MRI phenotypes reflects secondary associations or consequences of a disorder, such as behavioural changes (e.g. smoking, alcohol use), physical health disorders (e.g. obesity, diabetes), or pharmacological treatment [98–100]. However, what is of 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 to disorders distinctively addresses the question of what "develops differently" based on an understanding of "normal development", by linking genetic risks and atypical phenotypes to a prior transcriptional architecture of healthy brain development.

Clearly there are limits to what can be learnt from RNA measurements of bulk tissue samples from 6 healthy adult brains. Here, we used consistency within the AHBA dataset to define generalisability, *g*, an unbiased metric for optimising the trade-off involved in excluding noisy data while retaining as many genes and brain regions as possible in the HCP-MMP parcellation, which boosted signal-to-noise and yielded three generalisable components. While it is likely that the intrinsic architecture of human brain gene expression involves more than three components, our analysis suggests that their discovery will rely on additional high-granularity transcriptional data. Meanwhile the publicly available processing code and data that supplements our results should benefit future studies seeking to explicitly navigate the limits of the unique AHBA resource. For example, many imaging-transcriptomic studies are explicitly aiming to uncover components of gene expression which account for less variance in the AHBA than C1-C3, such that optimising signal-to-noise whilst retaining a large fraction of genes and brain regions is of key importance.

One particular constraint of the AHBA is that the small number of brains focuses attention on genes with minimal inter-individual variation. Our results were robust to some relaxation of this condition (**Extended Data Fig. 2)**. However, gene expression is known to vary markedly due to sex, age, genetic and environmental factors [101], so future data may reveal that the brain's transcriptional architecture includes additional components that are more individually variable and/or demographically diverse than the three generalisable components we have characterised here.

Extended Data Figures

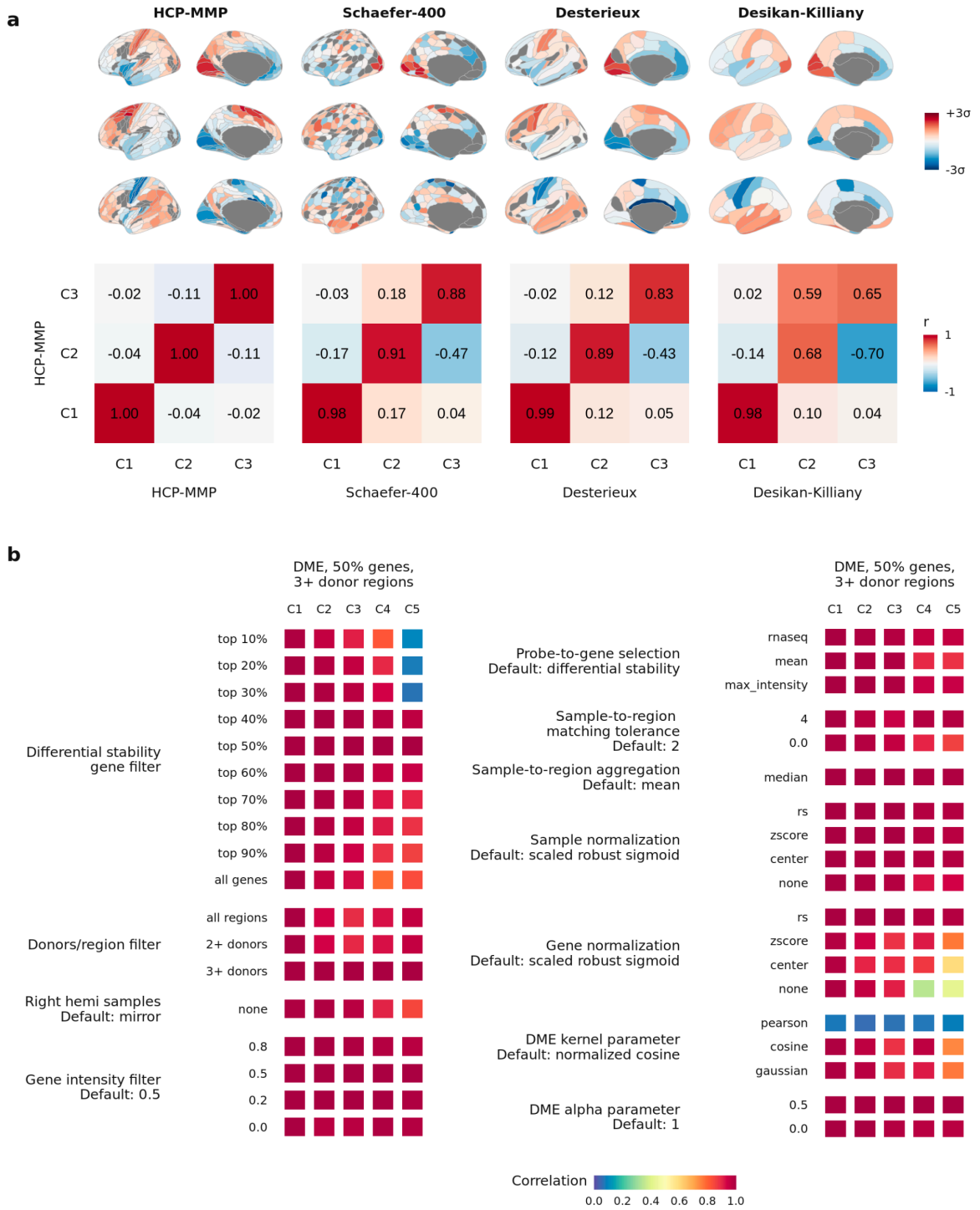
Optimised processing of the AHBA identified three generalisable components



Extended Data Figure 1: Optimised processing of the AHBA identified three generalisable components. a*, In the HCP-MMP parcellation template, 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 when processed in the HCP-MMP parcellation template.* ***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, estimated as median absolute correlation between matched components computed on disjoint triplet pairs; 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 result in spatially equivalent components.

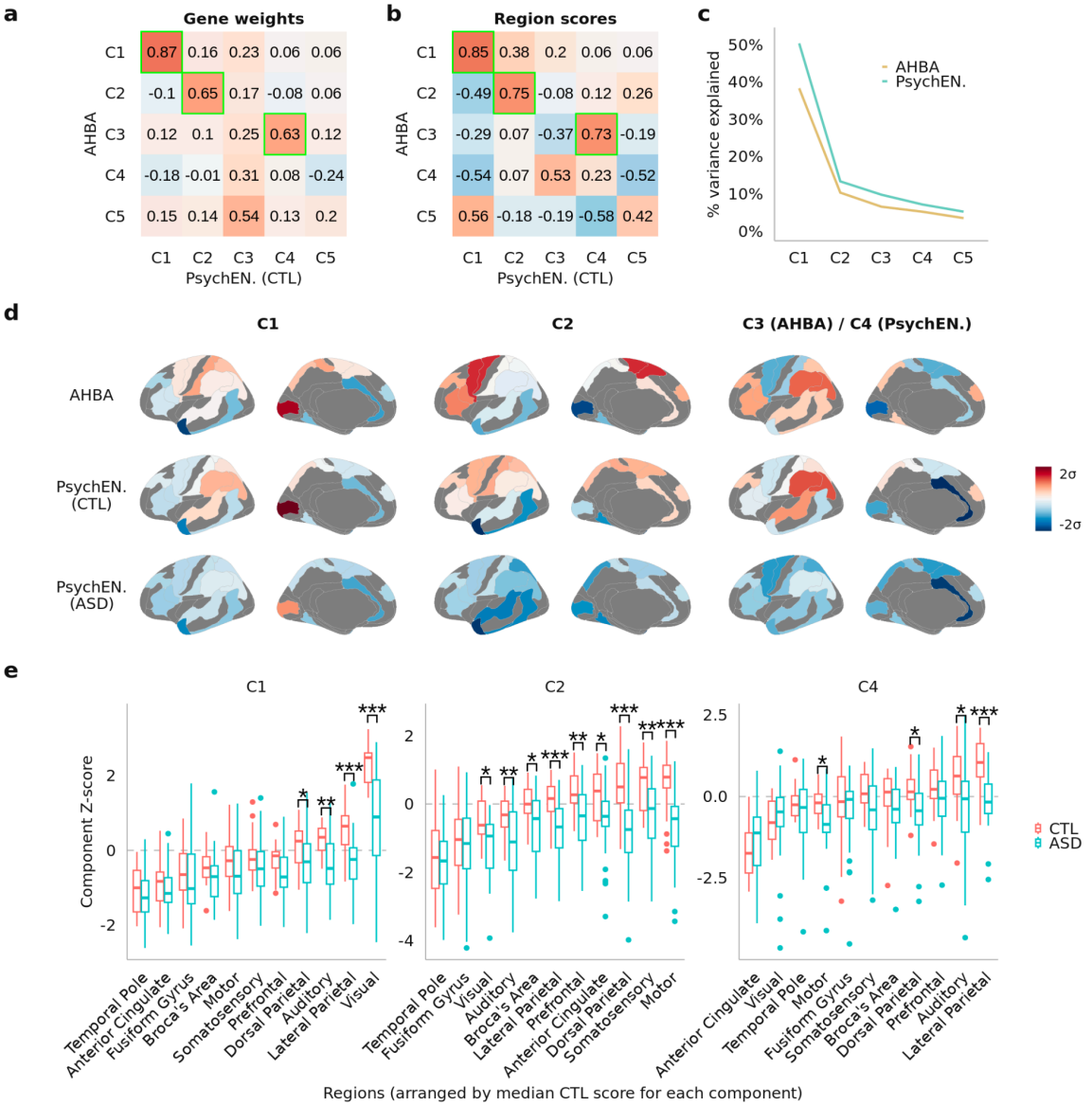
Transcriptional components were robust to parcellation and processing



Extended Data Figure 2: Transcriptional components were robust to parcellation and processing. a, *Transcriptional components were computed in four different parcellation templates: HCP-MMP1.0 [102], Schaefer-400 [103], Desterieux [104], and Desikan-Killiany [105]. 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.*

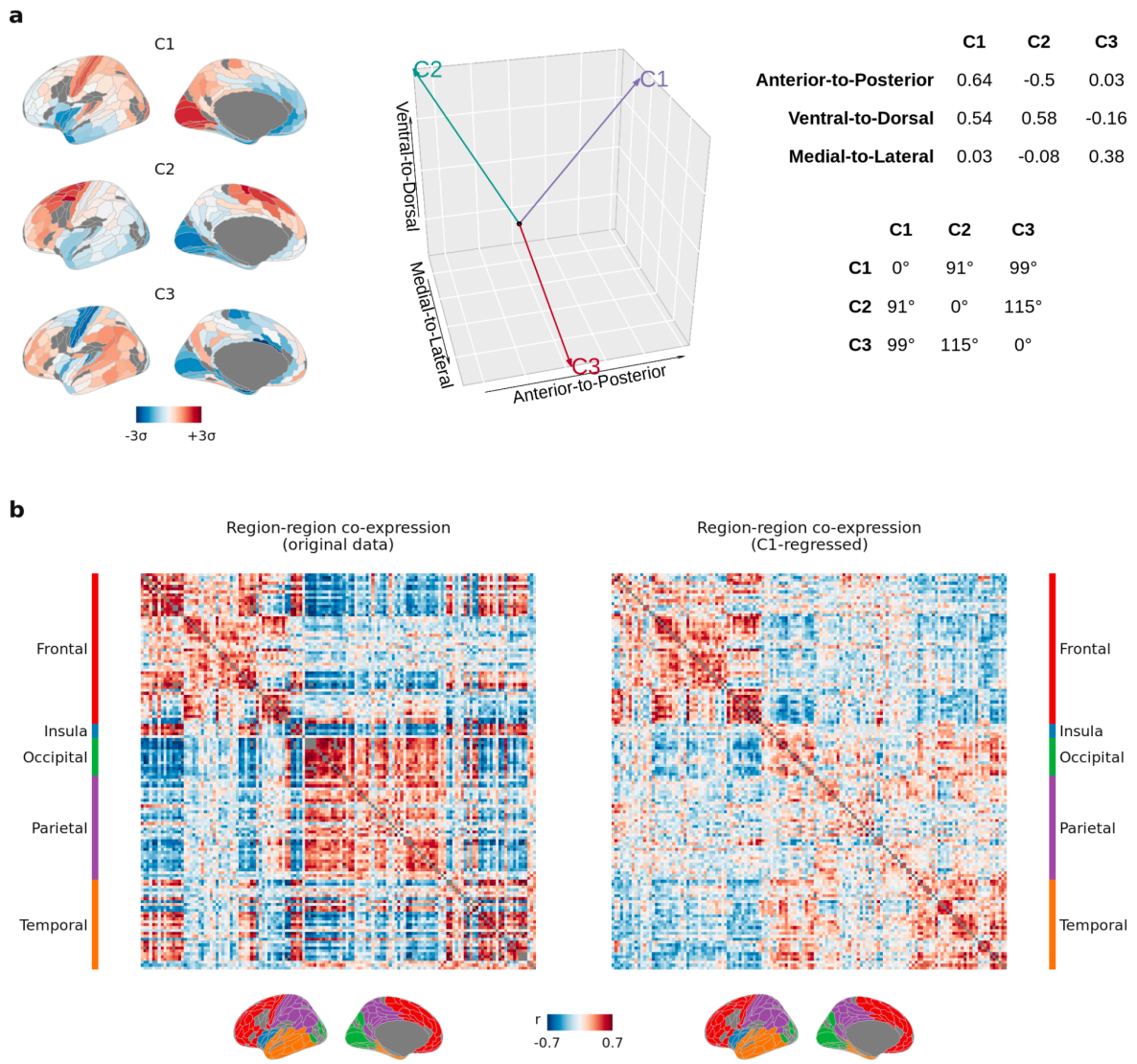
AHBA transcriptional components were reproducible in independent PsychENCODE control data, with differential spatial expression in autism



Extended Data Figure 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 data [23] were correlated with gene weights from the components of the AHBA (defined by DME in the 180-region HCP-MMP parcellation, as elsewhere), 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 components C1, C2 and C4 of the PsychENCODE PCA were correlated with scores derived from C1, C2 and C3 of the AHBA DME for the same 11 regions defined by PsychENCODE, showing that the matching genetic profiles*

*(in **a**) correspond to matching spatial expression patterns between these independent datasets. **c**, The proportions of variance explained by the first five components of the AHBA and PsychENCODE datasets, 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 the AHBA C1-C3 regional scores in the same Brodmann regions sampled in the PsychENCODE data. 2nd row: Cortical maps of C1, C2, and C4 from PCA of the PsychENCODE healthy control (CTL) data demonstrating their spatial similarity to cortical maps of AHBA C1-C3 (as in **b**). 3rd row: Gene weights from the PsychENCODE healthy control data were projected onto transcriptional data from the same 11 regions in a group of cases with autism spectrum disorder (ASD; N = 58) from the same dataset, visually demonstrating lower regional expression at the positive (red) pole of each component in the ASD cases compared to healthy controls. **e**, Boxplots showing the distribution over individual donor brains in the PsychENCODE dataset of regional scores for principal components C1, C2 and C4, for each of the 11 cortical regions, split by diagnosis. Components were computed on the group-average healthy control PsychENCODE data (as in **a-d**) then projected onto each individual donor's transcriptional data. Cortical regions are arranged on the x-axis by their median component score in healthy controls. Case-control differences in expression were tested for each component in each region and corrected for multiple comparisons across all 33 tests, demonstrating the following significant differences in transcriptional component expression in ASD cases compared to healthy controls: i) dorsal and lateral parietal cortices have reduced expression in all three components; ii) visual and auditory cortices have reduced expression specifically of C1; and iii) almost all regions tested have reduced expression of C2. Region names refer to the sampled Brodmann areas (BA) [23]: 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. *, **, *** indicate FDR-corrected P-value < 0.05, 0.01, 0.001 respectively.*

Higher-order components of cortical gene expression reflect anatomically relevant co-expression structure



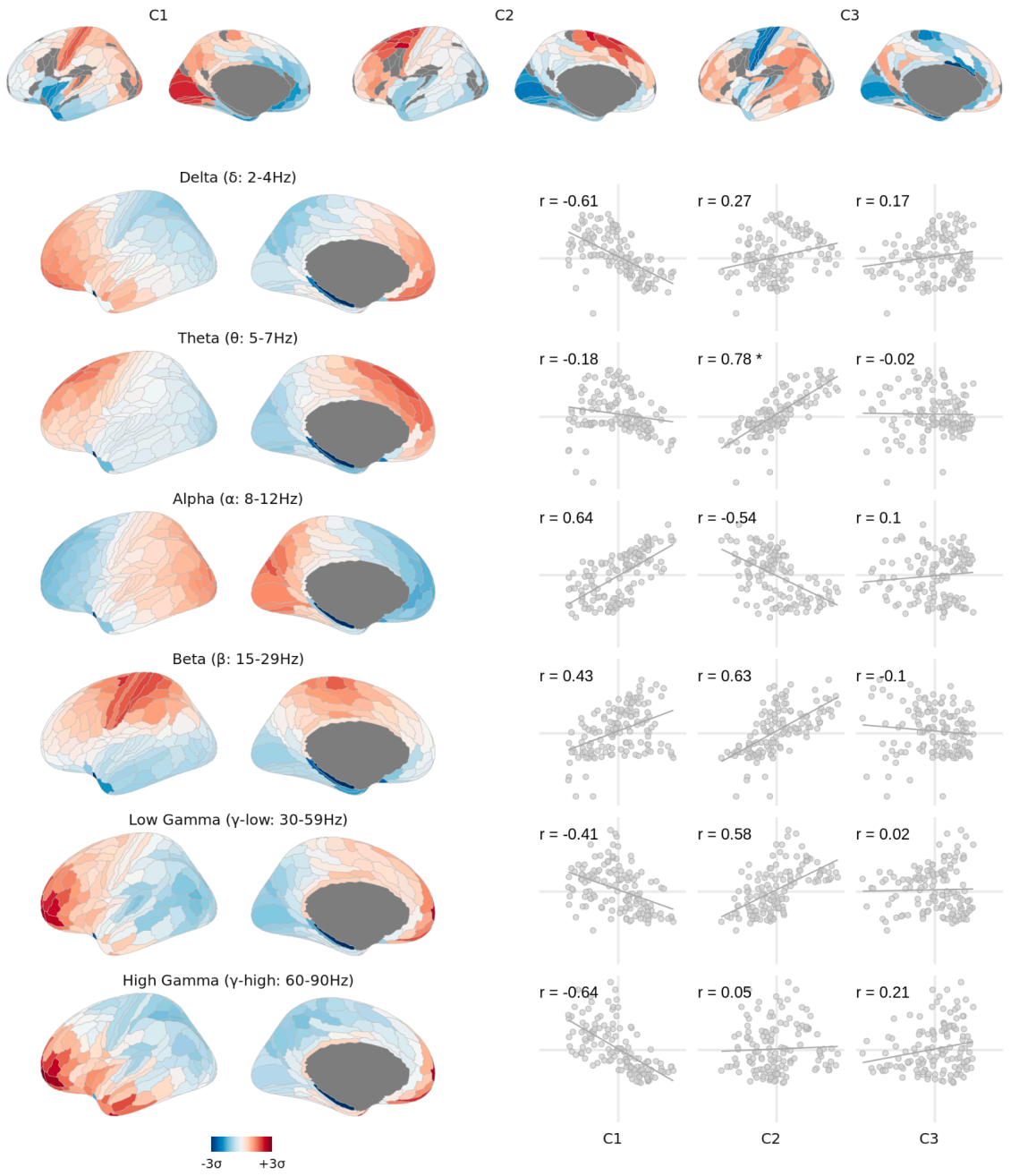
	C1	C2	C3
Anterior-to-Posterior	0.64	-0.5	0.03
Ventral-to-Dorsal	0.54	0.58	-0.16
Medial-to-Lateral	0.03	-0.08	0.38

	C1	C2	C3
C1	0°	91°	99°
C2	91°	0°	115°
C3	99°	115°	0°

Extended Data Figure 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*

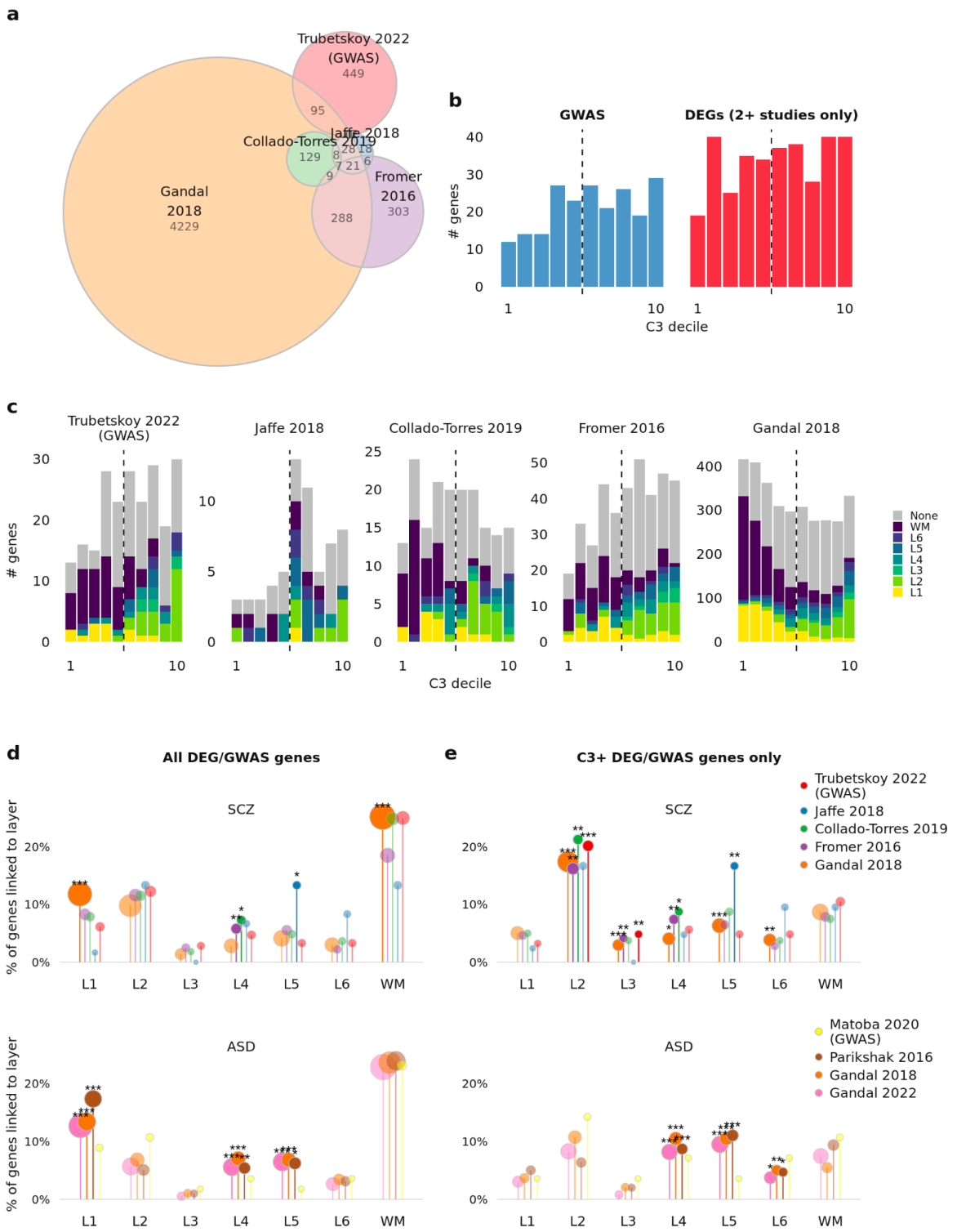
lobes as defined in the HCP-MMP parcellation [102]. This further demonstrates that the gene co-expression structure captured by C2 and C3 (i.e., the residual variation beyond C1) is anatomically relevant.

Transcriptional components were distinctively associated with the regional power of canonical brain oscillation frequencies



Extended Data Figure 5: Transcriptional components were distinctively associated with the regional power of canonical brain oscillation frequencies. *Several MEG power bands 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.*

C3 reveals shared biology across inconsistent postmortem brain RNA-seq studies of differentially expressed genes (DEGs) in schizophrenia



Extended Data Figure 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 gene [25]. 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.*

Methods

AHBA data and donor-level parcellation images

Probe-level gene expression data with associated spatial coordinates wer 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) [20]. The use of native donor parcellation images (rather than a standard parcellation image with sample coordinates mapped to MNI space) was chosen as it optimised the triplet generalisability metric (see 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 (i) *abagen* version 0.1.3 or greater (for Python) [21], or (ii) the version of the *AHBAprocessing* pipeline updated in June 2021 or later (for Matlab) [20].

Here we processed the AHBA with the *abagen* package, with one modification: we filtered the AHBA samples for only those annotated as cortical samples prior to 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 optimised the triplet generalisability metric (see 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, per published recommendations [20], 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 samples [20].

- *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 2mm, using the voxel-mass algorithm in the *abagen* package.
- *Sample normalisation*: Prior to aggregating over donors, samples were normalised across all genes, using the scaled robust sigmoid method (a sigmoid transformation that is robust to outliers [20]).
- *Gene normalisation*: Prior to aggregating over donors, genes were normalised 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, e.g. not mirroring right hemisphere samples to the left hemisphere, different or no intensity filter for genes, different methods for aggregating and normalising probes. Sensitivity analyses also included running the pipeline with alternative parcellation templates (**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 (DS) as previously described by Hawrylycz *et al.* [22]. For each gene, DS 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 DS 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 optimise for generalisability) where the threshold for DS 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 137/180 regions. Note that in the triplets analysis (see below), this means 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 chosen on the basis of a grid-search in combination with the differential stability gene filter to optimise for generalisability (**Extended Data Fig. 1**).

Triplets analysis: disjoint triplet correlation as a proxy for generalisability

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 e.g. probe aggregation, normalisation, and filters) independently to each of the twenty possible combinations of triplets, and correlated the regional scores for each DME or PCA component between each of the ten 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 employed a matching algorithm in which the full correlation matrix was computed between the top five principal components of both triplets (e.g. 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 ten disjoint triplet pairs represented the generalisability, *g*, of the AHBA components processed using the given set of parameters. Processing parameters, in particular the filters for regions and donors, were optimised so as to maximise *g* while retaining as many genes and regions as possible; see **Extended Data Fig. 1**.

Dimension reduction methods

Dimension reduction was performed using both principal component analysis (PCA) and diffusion map embedding (DME), the latter having been described for use in spatial gradient analysis of brain imaging data by Margulies *et al.* [17]. For DME, the normalised 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 optimised the inter-triplet correlation metric for generalisability. Both PCA and DME methods were implemented using the BrainSpace package [106].

Diffusion map embedding (DME) is a graph Laplacian dimension reduction technique that can be understood as a form of non-linear PCA [107,108]. When applied to the region-by-gene expression matrix of the AHBA, PCA finds principal components by directly computing the eigenvectors of the matrix. In contrast, DME first projects the data into a region-by-region affinity matrix using a kernel (e.g.

normalised cosine, Gaussian, correlation, etc), then finds successive components such that, in the low-dimensional space defined by the components, regions which are highly connected in the affinity matrix are close together and regions that are weakly connected are far apart.

In both PCA and DME, the regional scores of the components have zero correlation by construction. However, in PCA, components are also forced to have uncorrelated gene weights, i.e. the components have zero correlation between their eigenvector coefficients. DME relaxes this second constraint, resulting in components that are allowed to have some correlation in their genetic loadings. In practice this cross-correlation of gene weights is minimal ($r < 0.05$) but nevertheless allows for some flexibility that results in more robust components. There is also no prior evidence that transcriptional components should represent mutually uncorrelated sets of co-expressed genes, as mandated by PCA eigenvectors, so the less constrained weighting of genes in DME components may more accurately reflect the underlying transcriptional architecture. Importantly, PCA and DME components were in practice very highly correlated ($r > 0.9$ for each of PC1/C1, PC2/C2, PC3/C3; **Extended Data Fig. 1**), demonstrating that DME identifies broadly the same transcriptional components as PCA in a more robust and biologically plausible manner.

Independent component analysis (ICA) was also tested as an alternative to DME and PCA, but we found that ICA did not reproduce the well-known PC1 of the AHBA that has been extensively characterised in the prior literature [2,8,10]. ICA components were very different ($|r| < 0.4$) to the PCA and DME components, and also introduced additional uncertainty because: i) lower-order ICA components vary depending on the number of components to be estimated ; ii) ICA components do not have a natural ordering; and iii) the ICA algorithm uses a quadratic optimisation that depends on initial conditions [109]. In transcriptomics, ICA is most often used to isolate systematic sources of noise such as technical confounds, which is less applicable for our analysis with the AHBA where rows are not independent and identically-distributed samples but rather represent unique spatial locations on the cortex.

Component gene weights

For each component, gene weights were computed as the Pearson 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 $e_{g,i}$, the total variance V_i of the residualised region-by-gene expression matrix is

$$V_i = \sum_g Var\left(e_{g,i}\right)$$

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

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

Gene Ontology 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 online software *STRING* [110], 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 [110,111]), and includes a Benjamini-Hochberg adjustment of the False Discovery Rate (FDR).

The aggregate fold change method was chosen as it does not require thresholding the gene weights of the components to define 'target' vs 'background' gene lists (as in e.g. 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 Gene Ontology 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 dorsolateral prefrontal cortex [25] (columns Q-W of Table S4B in Maynard *et al.* [25]).

Cell-type gene lists were obtained from Seidlitz *et al.* [26], who compiled lists of significantly differentially expressed genes from five independent single-cell RNA-seq studies [33,112–115]. The gene list for synaptic marker genes was the unfiltered gene list from SynaptomeDB [27].

All enrichments for layers and cell-types were computed by the same aggregate fold change method [111] as in the *STRING* software [110], 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.* 2018, Supplementary Table 7 [32] (educational attainment).
- Davies *et al.* 2018, Supplementary Table 6 [28]
- Savage *et al.* 2018, Supplementary Table 15 [30]
- Hill *et al.* 2019, Supplementary Table 5 [29]
- Hatoum *et al.* 2023, Supplementary Table 16 [31]

Enrichment tests were performed by the aggregate fold change method [111], as above.

Neuroimaging and other macro-scale brain maps (Fig. 2)

Neuroimaging and other macro-scale maps were obtained as follows:

- The 9 neuroimaging and macro-scale maps in the clustering analysis (**Fig. 2a**) were obtained from the *Neuromaps* package [116], and are also available in Sydnor *et al.* [10].
- The regions of cytoarchitectural differentiation (**Fig. 2b**) were obtained from Paquola *et al.* 2019 [117] and averaged into the HCP-MMP parcellation using the *Neuromaps* package [116].
- The map of fMRI degree (**Fig. 2c**) was obtained from Paquola *et al.* 2020 [39], and was originally computed from the HCP S900 release [118].

- The maps of MEG power bands (**Fig. 2d, Extended Data Fig. 5**) were obtained from the *Neuromaps* package [116].
- The map of adolescent change in cortical myelination was obtained from Váša *et al.* 2020 [46].

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

Spatial associations between maps and the transcriptional components were computed by Pearson correlations and tested for significance using spin permutation tests (5,000 spins) by the Cornblath method [119], leveraging tools from *Neuromaps* [116], 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 analysis of variance (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 normalised 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/atlases-and-data/rnaseq) [47].

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 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+, 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 website [6] (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 online 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 4 regions were sampled were dropped.
4. Within each donor, expression of each gene was Z-normalised over regions.
5. Donors were aggregated into three age ranges (pre-birth, birth-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:

1. Processed BrainSpan data were filtered for only the 7,973 genes retained in the filtered AHBA dataset (top 50% by differential stability; see above).
2. 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-13 years, and 18-40 years).
3. 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.
4. 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 $\bar{y}$ of BrainSpan donors in each age range.

BrainSpan developmental trajectory modelling (Fig. 3e)

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

1. The ages in the BrainSpan data were converted to post-conception days on a log10 scale.
2. For each gene, a Generalised 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.
3. Developmental curves were plotted from the fitted models for each gene, sex, and region, 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, Seidlitz, White *et al.* [48], in which normative models were computed for multiple brain phenotypes across the human lifespan from a harmonised dataset of >125,000 total MRI scans ($N_{controls}$ = 38,839, N_{ASD} = 381, N_{MDD} = 3,861, N_{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) then re-averaged into the Desikan-Killiany parcellation. Pearson correlations with cortical maps of C1-C3 scores were computed, 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)

Differentially expressed genes (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.* 2022, Supplementary Table S3 [23], WholeCortex_ASD_FDR < 0.05
    - Gandal *et al.* 2018, Supplementary Table S1 [54], ASD.fdr < 0.05
    - Parikshak *et al.* 2016, Supplementary Table S2 [53], FDR-adjusted P value, ASD vs CTL < 0.05
- MDD
    - Jaffe *et al.* 2022, Supplementary Table S2 [55], Cortex_adjPVal_MDD < 0.05
- Schizophrenia
    - Fromer *et al.* 2016, Supplementary Table S16 [56], FDR estimate < 0.05

- Gandal *et al.* 2018, Supplementary Table S1 [54], SCZ.fdr < 0.05
- Jaffe *et al.* 2018, Supplementary Table S9 [59], fdr_qsva < 0.05
- Collado-Torres *et al.* 2019, Supplementary Table S11 [58], 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 2 studies.

Enrichments for these gene sets in each disorder were computed by the aggregate fold change method [111], i.e. 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.* 2020, Supplementary Table S7 [49]
- MDD: Howard *et al.* 2019, Supplementary Table S9 [50]
- Schizophrenia: Trubetskoy *et al.* 2022, Extended GWAS [51]: https://figshare.com/articles/dataset/scz2022/19426775?file=35775617

Associations with GWAS were calculated using three methods (**Supplementary Figure S6**):

- Enrichment of the prioritised genes identified in each of the specific studies, using the aggregate fold change method [111] as described above.
- MAGMA [60], a regression technique which tests for association between each of the components C1-C3 and the *P*-values for each gene's association with ASD, MDD or SCZ (from corresponding primary GWAS studies) without requiring a threshold to be applied to the GWAS-derived *P*-values to define a prioritised 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 SNP-to-gene mapping window of +35kb/-10kb was used.
- H-MAGMA [61], 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 layer [25] were computed for the disorder-associated gene lists from DEGs and GWAS 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.* [52] (N=90 subjects, 46 cases), and parcellated using the HCP-MMP1.0 parcellation. Pearson's correlations were computed with C1-C3 and significance assessed by spin permutation tests, corrected for FDR.

Author information

Department of Psychiatry, University of Cambridge, Cambridge, UK

Richard Dear, Richard A.I. Bethlehem, Edward T. Bullmore, Petra E. Vértes

Lifespan Brain Institute, The Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, USA

Jakob Seidlitz

Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Jakob Seidlitz

Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

Jakob Seidlitz

McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

Ross D. Markello

Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia

Aurina Arnatkevičiūtė

Department of Psychology, Yale University, New Haven, CT, USA

Kevin M. Anderson

Wellcome Centre for Human Neuroimaging, London, UK

Konrad Wagstyl

Section On Developmental Neurogenomics, National Institute of Mental Health, Bethesda, MD, USA

Armin Raznahan

Lifespan Brain Chart Consortium

List of members https://github.com/brainchart/Lifespan

Author contributions

P.E.V., R.D., J.S., K.W., and A.R. designed research. J.S., A.A., K.W., and R.B. contributed data. R.D. and P.E.V. performed research, and 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.

Data availability

Regional scores and gene weights for the transcriptional components C1-C3 are provided in **Supplementary Table 1**.

Spatial gene expression data are available from http://human.brain-map.org and http://brainspan.org, with parcellation images from https://figshare.com/articles/dataset/AHBAdata/6852911. 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**.

Code availability

All the code used for these analyses is publicly available at https://github.com/richardajdear/AHBA_gradients.

Disclosures

KMA is an employee of Neumora Therapeutics. RDM is an employee of Octave Biosciences. ETB has consulted for Boehringer Ingelheim, SR One, GlaxoSmithKline, Sosei Heptares, Monument Therapeutics. All other authors have no disclosures to make.

Acknowledgements

The authors acknowledge and thank Sofie Valk and Varun Warrier for their helpful comments on the manuscript.

R.D. was supported by the Gates Cambridge Trust. J.S. was supported by NIMH T32MH019112-29 and K08MH120564. A.A. was funded by a grant from the Australian Research Council (ARC) 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 an NIHR Senior Investigator award and the Wellcome Trust collaborative award for the Neuroscience in Psychiatry Network (NSPN). A.R. was supported by the National Institute of Mental Health Intramural Research Program (NIH Annual Report Number, 1ZIAMH002949-04). P.E.V. is a Fellow of MQ: Transforming Mental Health (MQF17_24).

Data were curated and analysed using a computational facility funded by an MRC research infrastructure award (MR/M009041/1) to the School of Clinical Medicine, University of Cambridge. All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and NIHR Applied Research Centre. The views expressed are those of the authors and not necessarily those of the NIH, NHS, 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.

References

1. van den Heuvel, M. P. & Yeo, B. T. T. A spotlight on bridging microscale and macroscale human brain architecture. *Neuron* **93**, 1248–1251 (2017).
2. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
3. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).
4. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
5. 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. U. S. A.* **108**, 14950–14955 (2011).
6. 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. U. S. A.* **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**, (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 neotenous 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. U. S. A.* **113**, 12574–12579 (2016).
18. Hill, J. *et al.* Similar patterns of cortical expansion during human development and evolution. *Proc. Natl. Acad. Sci. U. S. A.* **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. Arnatkevičiūtė, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).
21. Markello, R. D. *et al.* Standardizing workflows in imaging transcriptomics with the abagen toolbox. *bioRxiv* 2021.07.08.451635 (2021) doi:10.1101/2021.07.08.451635.
22. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).

23. Gandal, M.J., Haney, J.R., Wamsley, B. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
24. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
25. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
26. Seidlitz, J. *et al.* Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).
27. Pirooznia, M. *et al.* SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* **28**, 897–899 (2012).
28. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
29. 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).
30. 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).
31. 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).
32. 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).
33. 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).
34. 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).
35. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**,

62–79 (2013).

36. Vaishnavi, S. N. *et al.* Regional aerobic glycolysis in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17757–17762 (2010).
37. Satterthwaite, T. D. *et al.* Impact of puberty on the evolution of cerebral perfusion during adolescence. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8643–8648 (2014).
38. Reardon, P. K. *et al.* Normative brain size variation and brain shape diversity in humans. *Science* **360**, 1222–1227 (2018).
39. 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).
40. 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).
41. Mesulam, M.-M. *Principles of behavioral neurology*. (Oxford University Press, USA, 1985).
42. Glasser, M. F. *et al.* The Human Connectome Project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187 (2016).
43. Paquola, C. *et al.* Shifts in myeloarchitecture characterise adolescent development of cortical gradients. *Elife* **8**, (2019).
44. Larson-Prior, L. J. *et al.* Adding dynamics to the Human Connectome Project with MEG. *Neuroimage* **80**, 190–201 (2013).
45. Whitaker, K. J. & Vértes, P. E. *et al.* Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9105–9110 (2016).
46. Váša, F. *et al.* Conservative and disruptive modes of adolescent change in human brain functional connectivity. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3248–3253 (2020).
47. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).

48. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
49. 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).
50. 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).
51. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
52. Wagstyl, K. *et al.* Multiple markers of cortical morphology reveal evidence of supragranular thinning in schizophrenia. *Transl. Psychiatry* **6**, e780 (2016).
53. Parikshak, N. N. *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
54. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
55. 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).
56. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
57. Birnbaum, R. *et al.* Investigating the neuroimmunogenic architecture of schizophrenia. *Mol. Psychiatry* **23**, 1251–1260 (2018).
58. 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.e8 (2019).
59. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome

illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).

60. 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).
61. 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).
62. Merikangas, A. K. *et al.* What genes are differentially expressed in individuals with schizophrenia? A systematic review. *Mol. Psychiatry* **27**, 1373–1383 (2022).
63. Cavanagh, J. F. & Frank, M. J. Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* **18**, 414–421 (2014).
64. Buzzell, G. A. *et al.* Adolescent cognitive control, theta oscillations, and social observation. *Neuroimage* **198**, 13–30 (2019).
65. Senoussi, M. *et al.* Theta oscillations shift towards optimal frequency for cognitive control. *Nat Hum Behav* **6**, 1000–1013 (2022).
66. 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).
67. Choudhury, S., Charman, T. & Blakemore, S.-J. Development of the teenage brain. *Mind Brain Educ.* **2**, 142–147 (2008).
68. Brenhouse, H. C. & Schwarz, J. M. Immunoadolescence: Neuroimmune development and adolescent behavior. *Neurosci. Biobehav. Rev.* **70**, 288–299 (2016).
69. Santos, E. N. & Fields, R. D. Regulation of myelination by microglia. *Sci Adv* **7**, eabk1131 (2021).
70. 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).
71. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related

dementias. *Nat. Genet.* **54**, 412–436 (2022).

72. 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).
73. 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. U. S. A.* **115**, E1051–E1060 (2018).
74. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17**, 319–334 (1982).
75. Johnson, M. B. & Hyman, S. E. A critical perspective on the synaptic pruning hypothesis of schizophrenia pathogenesis. *Biol. Psychiatry* **92**, 440–442 (2022).
76. Howes, O. D. & Onwordi, E. C. The synaptic hypothesis of schizophrenia version III: a master mechanism. *Mol. Psychiatry* (2023) doi:10.1038/s41380-023-02043-w.
77. Harrison, P. J. & Weinberger, D. R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* **10**, 40–68; image 5 (2005).
78. 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).
79. 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).
80. 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).
81. Galakhova, A. A. *et al.* Evolution of cortical neurons supporting human cognition. *Trends Cogn. Sci.* **26**, 909–922 (2022).
82. Petanjek, Z., Judas, M., Kostović, I. & Uylings, H. B. M. Lifespan alterations of basal dendritic trees of pyramidal neurons in the human prefrontal cortex: a layer-specific pattern. *Cereb. Cortex* **18**, 915–929 (2008).

83. Petanjek, Z. *et al.* Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13281–13286 (2011).
84. 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).
85. Harrison, P. J. The neuropathology of schizophrenia: A critical review of the data and their interpretation. *Brain* **122**, 593–624 (1999).
86. 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).
87. 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).
88. Batiuk, M. Y. *et al.* Upper cortical layer-driven network impairment in schizophrenia. *Sci Adv* **8**, eabn8367 (2022).
89. Smucny, J., Dienel, S. J., Lewis, D. A. & Carter, C. S. Mechanisms underlying dorsolateral prefrontal cortex contributions to cognitive dysfunction in schizophrenia. *Neuropsychopharmacology* **47**, 292–308 (2022).
90. 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. U. S. A.* **116**, 9604–9609 (2019).
91. 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).
92. Romero-Garcia, R. *et al.* Schizotypy-related magnetization of cortex in healthy adolescence Is colocated with expression of schizophrenia-related genes. *Biol. Psychiatry* **88**, 248–259 (2020).
93. Anderson, K. M. *et al.* Transcriptional and imaging-genetic association of cortical interneurons, brain function, and schizophrenia risk. *Nat. Commun.* **11**, 2889 (2020).

94. Anderson, K. M. *et al.* Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25138–25149 (2020).
95. Fornito, A., Arnatkevičiūtė, A. & Fulcher, B. D. Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).
96. Arnatkeviciute, A., Fulcher, B., Bellgrove, M. & Fornito, A. Imaging transcriptomics of brain disorders. *PsyArXiv* (2021) doi:10.31234/osf.io/4exug.
97. Martins, D. *et al.* Imaging transcriptomics: Convergent cellular, transcriptomic, and molecular neuroimaging signatures in the healthy adult human brain. *Cell Rep.* **37**, 110173 (2021).
98. 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).
99. Perzel Mandell, K. A. *et al.* Molecular phenotypes associated with antipsychotic drugs in the human caudate nucleus. *Mol. Psychiatry* **27**, 2061–2067 (2022).
100. Schulmann, A. *et al.* Antipsychotic drug use complicates assessment of gene expression changes associated with schizophrenia. *Transl. Psychiatry* **13**, 93 (2023).
101. 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).
102. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
103. Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
104. 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).
105. 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).

106. 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).
107. Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426–7431 (2005).
108. Lafon, S. & Lee, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1393–1403 (2006).
109. Kairov, U. *et al.* Determining the optimal number of independent components for reproducible transcriptomic data analysis. *BMC Genomics* **18**, 712 (2017).
110. 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).
111. Yu, C. *et al.* A strategy for evaluating pathway analysis methods. *BMC Bioinformatics* **18**, 453 (2017).
112. 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).
113. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
114. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7285–7290 (2015).
115. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, (2018).
116. Markello, R. D. *et al.* neuromaps: structural and functional interpretation of brain maps. *bioRxiv* 2022.01.06.475081 (2022) doi:10.1101/2022.01.06.475081.
117. Paquola, C. *et al.* Microstructural and functional gradients are increasingly dissociated in

transmodal cortices. *PLoS Biol.* **17**, e3000284 (2019).

118. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).

119. Markello, R. D. & Misic, B. Comparing spatial null models for brain maps. Preprint at https://doi.org/10.1101/2020.08.13.249797.