

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

---

In the format provided by the  
authors and unedited

## Supplementary Information

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

Richard Dear<sup>1</sup>, Konrad Wagstyl<sup>2</sup>, Jakob Seidlitz<sup>3,4,5</sup>, Ross D. Markello<sup>6</sup>, Aurina Arnatkevičiūtė<sup>7</sup>, Kevin M. Anderson<sup>8</sup>, Richard A.I. Bethlehem<sup>1</sup>, Lifespan Brain Chart Consortium<sup>10</sup>, Armin Raznahan<sup>9</sup>, Edward T. Bullmore<sup>1</sup>, Petra E. Vértes<sup>1</sup>

<sup>1</sup> Department of Psychiatry, University of Cambridge, Cambridge, UK; <sup>2</sup> Wellcome Centre for Human Neuroimaging, London, UK; <sup>3</sup> Lifespan Brain Institute, The Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, USA; <sup>4</sup> Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>5</sup> Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA; <sup>6</sup> McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; <sup>7</sup> Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia; <sup>8</sup> Department of Psychology, Yale University, New Haven, CT, USA; <sup>9</sup> Section On Developmental Neurogenetics, National Institute of Mental Health, Bethesda, MD, USA; <sup>10</sup> Lifespan Brain Chart Consortium (list of members <https://github.com/brainchart/Lifespan>)

### Supplementary Methods

1. Further details on dimension reduction

### Supplementary Tables

1. C1-C3 region scores and gene weights \*
2. GO terms table \*
3. Normative brain maps \*
4. Disorder brain maps \*
5. Summary table of transcriptional datasets used for validation

\* Provided in separate excel file

### Supplementary Figures

1. Optimised AHBA components C1-C3 aligned with previously published PCs
2. Proportion of variance explained for C1-C3 was proportional to the number of strongly weighted genes
3. Enrichments for C1 aligned with prior results
4. Enrichments were not distinct for the less stable 50% of genes
5. Covariation in single-cells between positive and negative genes was only consistent for the same component, not between different components
6. GWAS associations were consistent when not thresholded for prioritised genes

## Supplementary Methods

### Further details on dimension reduction

Diffusion map embedding (DME) is a graph Laplacian dimension reduction technique that can be understood as a form of non-linear PCA<sup>100,101</sup>. 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<sup>2,8,10</sup>. 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<sup>102</sup>. 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.

## Supplementary Tables

Table S5: Summary of transcriptional datasets used for validation

| Dataset                                                | N donors | # cortical locations                 | Analysis method                                                  | Validation result for C1, C2, C3 ( <i>g</i> or <i>r</i> ) <sup>a</sup> |
|--------------------------------------------------------|----------|--------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------------|
| AHBA <sup>1</sup><br>(bulk microarray)                 | 6        | 2000+<br>(summarised to 180 regions) | DME, 50% most stable genes, regions with samples from 3+ donors  | <i>g</i> = 0.97, 0.72, 0.65                                            |
| PsychENCODE <sup>2</sup><br>(bulk RNA-seq)             | 54       | 11                                   | PCA, 50% most stable genes <sup>b</sup>                          | <i>r</i> = 0.85, 0.75, 0.73                                            |
|                                                        |          |                                      | PCA, all genes <sup>c</sup>                                      | <i>r</i> = 0.83, 0.60, 0.56                                            |
|                                                        |          |                                      | DME, 50% most stable genes <sup>d</sup>                          | <i>r</i> = 0.82, 0.65, 0.64                                            |
| BrainSpan <sup>3</sup><br>(bulk RNA-seq)               | 8*       | 11                                   | Projection of AHBA C1-C3 gene weights <sup>e</sup>               | <i>r</i> = 0.96, 0.88, 0.84                                            |
| Allen Cell Atlas <sup>4</sup><br>(single-cell RNA-seq) | 3        | 8                                    | Projection of AHBA C1-C3 gene weights: pos. vs neg. <sup>f</sup> | <i>r</i> = -0.90, -0.85, -0.95                                         |

<sup>a</sup> For the AHBA, the validation was performed using the intra-dataset generalisability measure *g*, defined as the median regional correlation of components from disjoint triplets of donor brains.

<sup>b</sup> The 50% most stable genes were calculated by taking the healthy control subjects who had samples from at least 10 of the 11 cortical regions, then using the same differential stability method as for the AHBA, i.e. ranking genes by the mean correlation of spatial expression patterns across all pairs of subjects (12,418 genes for the PsychENCODE dataset). All RIN > 3 samples from N = 54 control subjects were then averaged into a {11 regions x 12,418 genes} matrix, to which PCA was applied.

<sup>c</sup> As in <sup>b</sup>, but without filtering genes, i.e. PCA applied to the full matrix {11 regions x 24,836 genes}.

<sup>d</sup> As in <sup>b</sup>, but using DME instead of PCA. DME first converts the {11 regions x 12,418 genes} matrix into a {11 regions x 11 regions} affinity matrix, so PCA is preferable for these low-granularity data.

<sup>e</sup> The Brainspan analysis examined whether C1-C3 were consistent at different donor ages. Therefore the gene weights from the AHBA-derived C1-C3 were projected onto the group-average Brainspan data, to examine if the AHBA gene profiles had consistent spatial expression patterns in Brainspan.

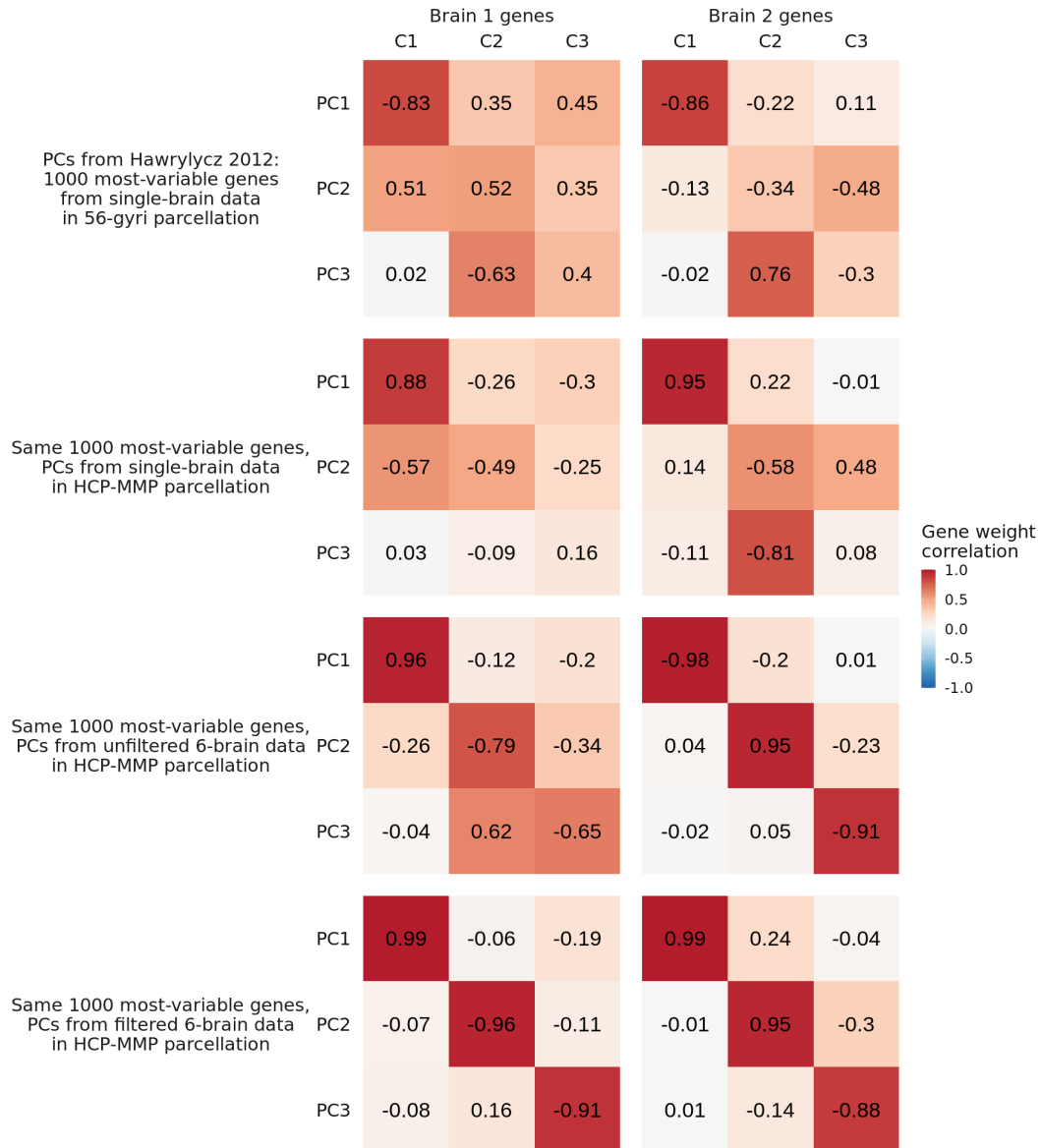
<sup>f</sup> The Allen Cell Atlas analysis examined whether C1-C3 were consistent within single cell nuclei.

Therefore the gene weights from the AHBA-derived C1-C3 were projected onto each single-cell nuclei, separately for positively- and negatively-weighted genes. Here the validation results are (anti-)correlations between the positive and negative genes for each component across single cells (i.e. these correlations are over 50,000 cells, not correlations over the 8 sampled brain regions).

\* 18-40 years subjects only; the same analysis was also performed for Pre-Birth and Birth-13y

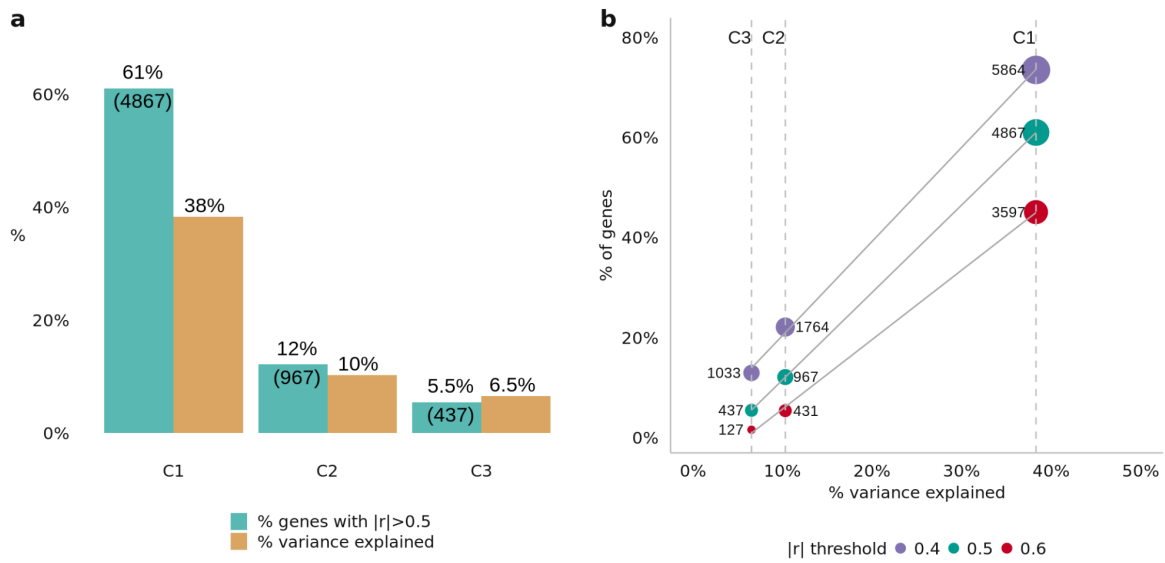
## Supplementary Figures

### S1: Optimised AHBA components C1-C3 aligned with previously published PCs



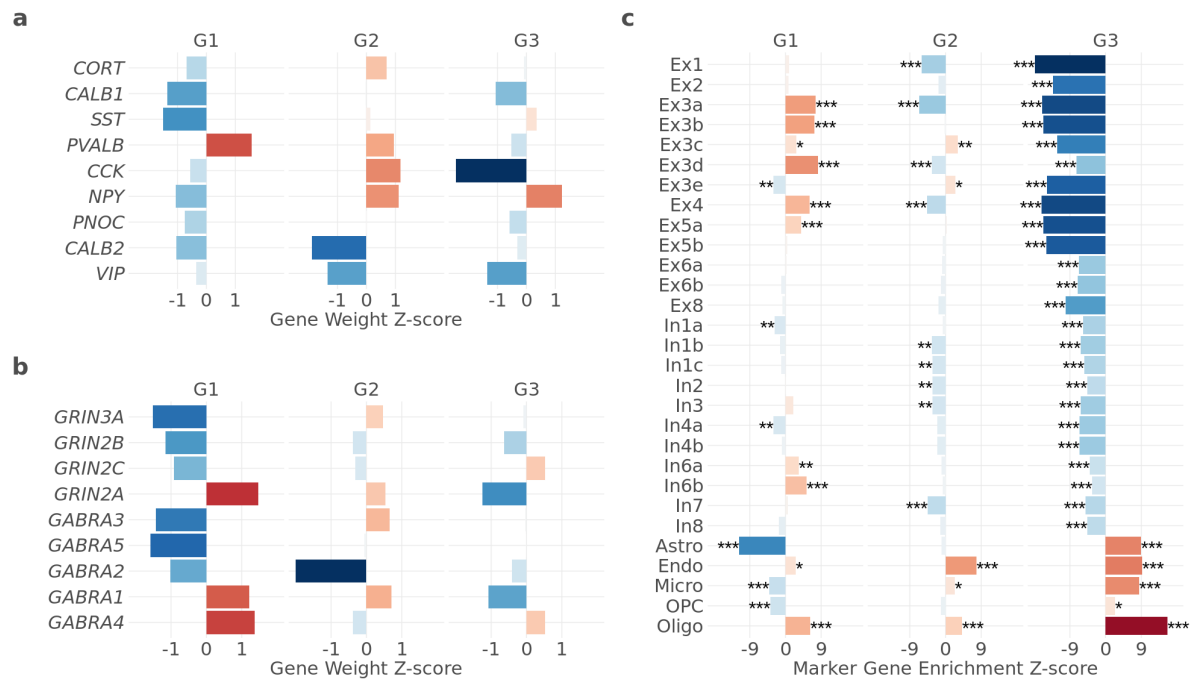
**Supplementary Figure S1: Optimised AHBA components C1-C3 aligned with previously published PCs.** AHBA components C1 and C2 are consistent with previously reported PCs computed from 1000 genes in single AHBA brains. In the initial release of the AHBA that included just two donor brains, Hawrylycz et al. showed that principal components of the expression of the 1000 genes with greatest spatial variation computed in a 56-region parcellation of a single brain were consistent between the two donor brains. Here we show that the gene weights of those PCs have similar alignment with the C1 and C2 that we derive using DME on 7973 genes across six donor brains in the HCP-MMP parcellation for different variations of data processing.

## S2: Proportion of variance explained for C1-C3 was proportional to the number of strongly weighted genes



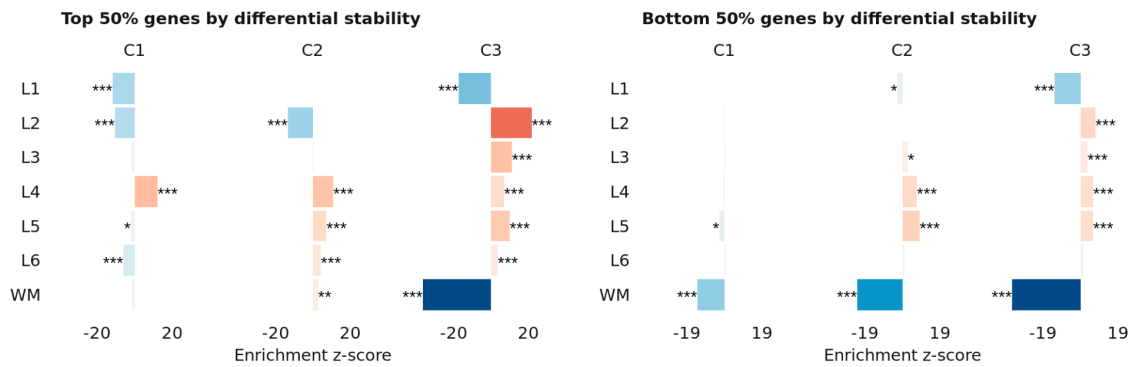
**Supplementary Figure S2: Proportion of variance explained for C1-C3 was proportional to the number of strongly weighted genes.** **a**, For each of C1-C3, the % of genes with absolute correlation  $|r| > 0.5$  is shown alongside the % variance explained (with the number of  $|r| > 0.5$  genes in parentheses), demonstrating that the % variance explained by higher-order components is related to the number of strongly weighted genes. **b**, The scatter plot shows the linear relationship between % variance explained and % strongly weighted genes for a range of threshold values ( $|r| > 0.4, 0.5$ , or  $0.6$ ). Size of points and their annotations display the number of genes above the given  $|r|$  threshold. For each  $|r|$  threshold, the slope of the linear relationship is  $>1$ , indicating that higher-order components C2 and C3 explain relatively more variance than expected given the number of strongly weighted genes alone (i.e. that C2 and C3 genes have relatively higher spatial variation in expression).

## S3: Enrichments for C1 aligned with prior results



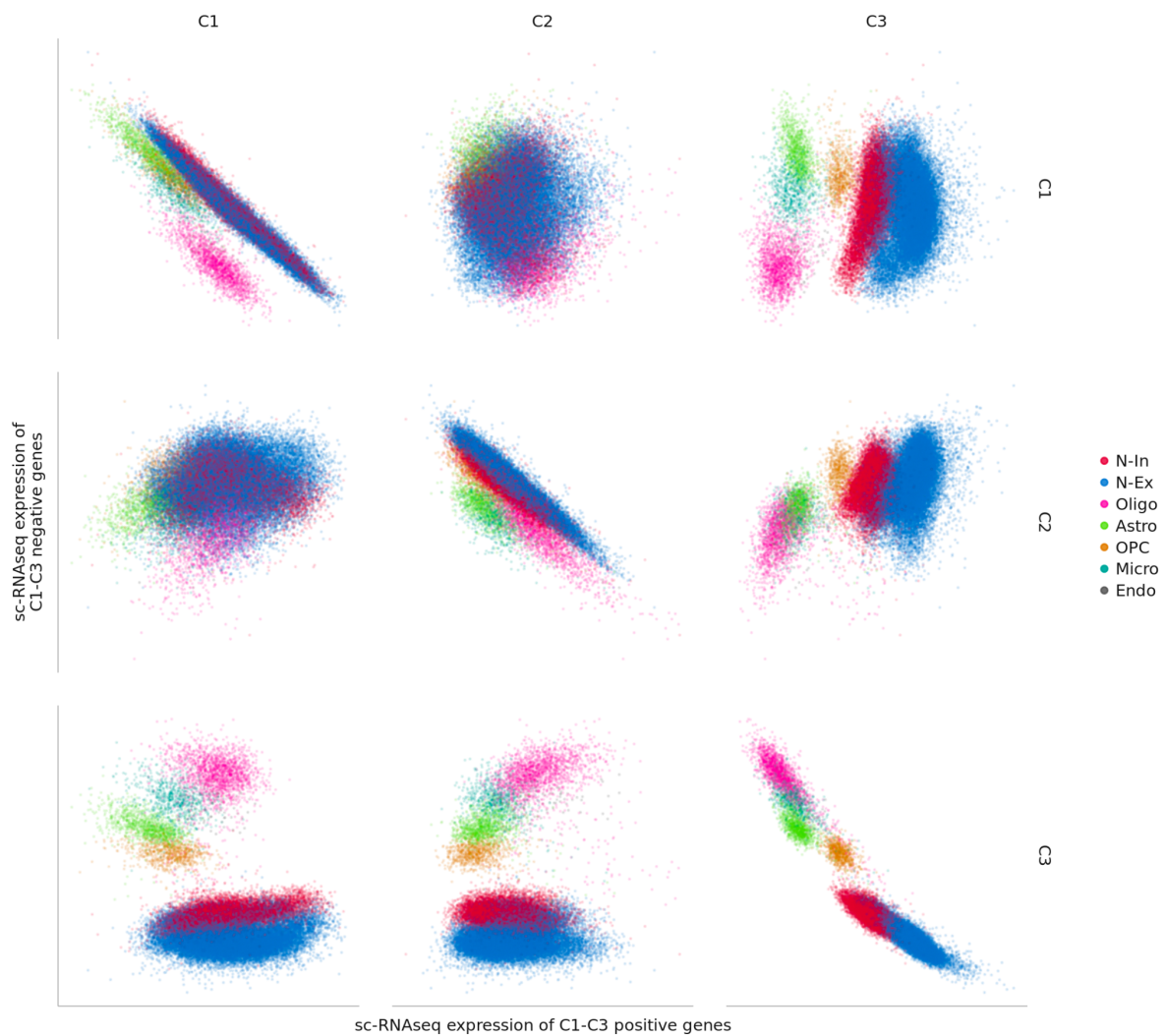
**Supplementary Figure S3: Enrichments for C1 aligned with prior results.** **a**, z-score of gene weight of key interneuron genes analysed for PC1 by Burt et al.<sup>5</sup>. **b**, as in **a** for key glutamatergic genes. **c**, enrichment (computed as z-score of mean gene weight of marker genes relative to permutations) of marker genes for neuron subclasses<sup>6</sup>. Significance was computed by two-sided permutation tests (Methods) and FDR-corrected across all tests; \*, \*\*, \*\*\* respectively indicate FDR-corrected two-sided p-value < 0.05, 0.01, 0.001.

## S4: Enrichments were not distinct for the less stable 50% of genes



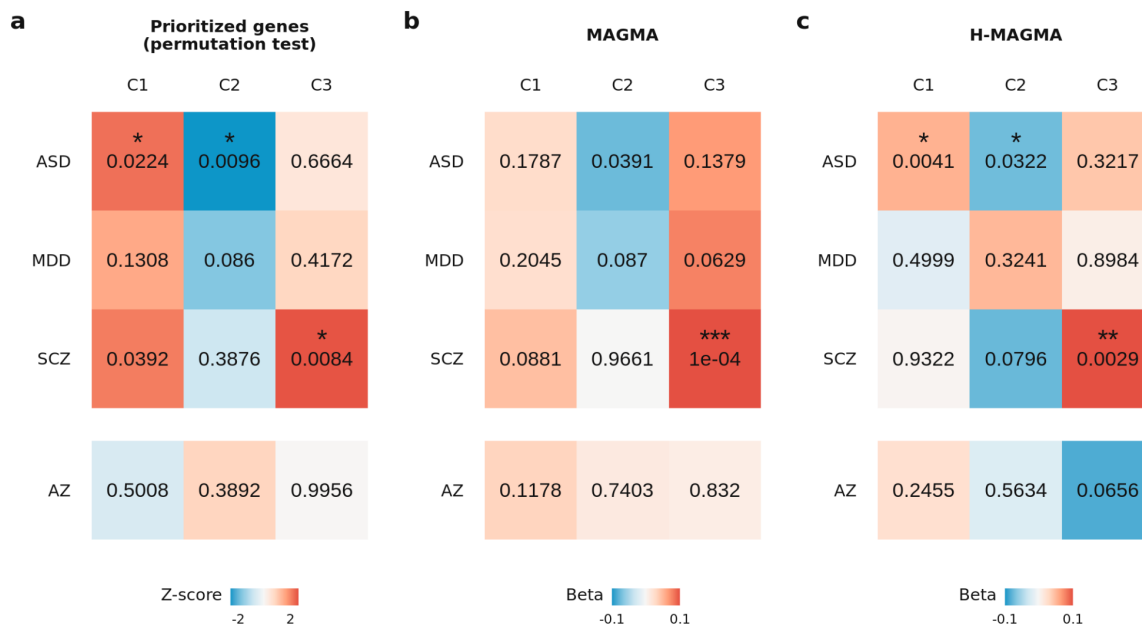
**Supplementary Figure S4: Enrichments were not distinct for the less stable 50% of genes.** Laminar enrichments from C1-C3 computed from the top 50% most stable genes (as in Fig. 1c-d), and the bottom 50% least stable genes. This demonstrates that the distinct laminar enrichments for C1-C3 depend on the gene stability filtering used to optimise the AHBA processing for generalisability. Significance was computed by two-sided permutation tests (Methods) and FDR-corrected across all tests; \*, \*\*, \*\*\* respectively indicate FDR-corrected two-sided  $p$ -value < 0.05, 0.01, 0.001.

S5: Covariation in single-cells between positive and negative genes was only consistent for the same component, not between different components



**Supplementary Figure S5: Covariation in single-cells between positive and negative genes was only consistent for the same component, not between different components.** This demonstrates that the observed anticorrelations are not a trivial result of e.g. total overall expression.

## S6: C1-C3 associations to autism and schizophrenia GWAS were consistent when not thresholded for prioritised genes



**Supplementary Figure S6: C1-C3 associations to autism and schizophrenia GWAS were consistent when not thresholded for prioritised genes.** **a**, Enrichment results for genes prioritised in the latest published GWAS studies of ASD<sup>7</sup>, MDD<sup>8</sup>, SCZ<sup>9</sup> (as in **Figure 4d**), and also Alzheimer's disease (AZ)<sup>10</sup>, assessed using the aggregate fold change method<sup>11,12</sup>. Colour indicates the Z-score of the true mean C1-C3 weight of the prioritised genes relative to 5000 random permutations; numbers show the uncorrected two-sided p-values of this permutation test; \*, \*\*, \*\*\* indicate FDR-corrected two-sided p-value < 0.05, 0.01, 0.001, respectively. **b**, Associations for ASD, MDD, SCZ, and AZ from the same GWAS studies assessed using MAGMA<sup>13</sup>, showing a strong association between SCZ and C3 consistent with the prioritised genes, but a weaker (though directionally consistent) association of ASD with C1 and C2. **c**, as in **b**, but assessed using H-MAGMA<sup>14</sup>, which accounts for trans-regulatory genetic effects using Hi-C chromatin data from postmortem brain tissue. The use of H-MAGMA reproduced the significant association of ASD with C1 and C2, and the association of SCZ with C3. H-MAGMA was also the method used by Matoba et al. (2022) in the GWAS study to prioritise ASD-associated genes that were significantly associated with C1 and C2 (**a**). This suggests that trans-regulatory effects may play a significant role in how the spatially-expressed transcriptional programmes reflected in C1 and C2 are altered in ASD. Significance was computed by the MAGMA method<sup>13</sup> and FDR-corrected across all tests; \*, \*\*, \*\*\* respectively indicate FDR-corrected two-sided p-value < 0.05, 0.01, 0.001.

## References

1. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
2. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
3. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
4. Hodges, A. Computer science. Beyond Turing’s machines. *Science* **336**, 163–164 (2012).
5. Burt, J. B. *et al.* Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
6. 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).
7. 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).
8. 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).
9. Trubetsky, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
10. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
11. Yu, C. *et al.* A strategy for evaluating pathway analysis methods. *BMC Bioinformatics* **18**, 453 (2017).
12. 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).

13. 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).
14. 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).

## Consortium Members

Members in the Lifespan Brain Chart Consortium as of March 2024 are listed below. The updated list can be found at <https://github.com/brainchart/Lifespan>.

|           |      |                 |
|-----------|------|-----------------|
| Chris     |      | Adamson         |
| Sophie    |      | Adler           |
| Aaron     | F.   | Alexander-Bloch |
| Evdokia   |      | Anagnostou      |
| Kevin     | M.   | Anderson        |
| Ariosky   |      | Areces-Gonzalez |
| Duncan    | E.   | Astle           |
| Bonnie    |      | Auyeung         |
| Muhammad  |      | Ayub            |
| Jong Bin  |      | Bae             |
| Gareth    |      | Ball            |
| Simon     |      | Baron-Cohen     |
| Richard   |      | Beare           |
| Saashi    | A.   | Bedford         |
| Vivek     |      | Benegal         |
| Richard   | A.I. | Bethlehem       |
| Frauke    |      | Beyer           |
| John      |      | Blangero        |
| Manuel    |      | Blesa Cábez     |
| James     | P.   | Boardman        |
| Matthew   |      | Borzage         |
| Jorge     | F.   | Bosch-Bayard    |
| Niall     |      | Bourke          |
| Edward    | T.   | Bullmore        |
| Vince     | D.   | Calhoun         |
| Mallar    | M.   | Chakravarty     |
| Christina |      | Chen            |
| Casey     |      | Chertavian      |
| Gaël      |      | Chetelat        |
| Yap       | S.   | Chong           |
| Aiden     |      | Corvin          |
| Manuela   |      | Costantino      |
| Eric      |      | Courchesne      |

|             |    |              |
|-------------|----|--------------|
| Fabrice     |    | Crivello     |
| Vanessa     | L. | Cropley      |
| Jennifer    |    | Crosbie      |
| Nicolas     |    | Crossley     |
| Marion      |    | Delarue      |
| Richard     |    | Delorme      |
| Sylvane     |    | Desrivieres  |
| Gabriel     |    | Devenyi      |
| Maria       | A. | Di Biase     |
| Ray         |    | Dolan        |
| Kirsten     | A. | Donald       |
| Gary        |    | Donohoe      |
| Lena        |    | Dorfschmidt  |
| Katharine   |    | Dunlop       |
| Anthony     | D  | Edwards      |
| Jed         | T. | Elison       |
| Cameron     | T. | Ellis        |
| Jeremy      | A. | Elman        |
| Lisa        |    | Eyler        |
| Damien      | A. | Fair         |
| Paul        | C. | Fletcher     |
| Peter       |    | Fonagy       |
| Carol       | E. | Franz        |
| Lidice      |    | Galan-Garcia |
| Ali         |    | Gholipour    |
| Jay         |    | Giedd        |
| John        | H. | Gilmore      |
| David       | C. | Glahn        |
| Ian         | M. | Goodyer      |
| P.          | E. | Grant        |
| Nynke       | A. | Groenewold   |
| Shreya      |    | Gudapati     |
| Faith       | M. | Gunning      |
| Raquel      | E. | Gur          |
| Ruben       | C. | Gur          |
| Christopher | F. | Hammill      |
| Oskar       |    | Hansson      |

|            |    |             |
|------------|----|-------------|
| Trey       |    | Hedden      |
| Andreas    |    | Heinz       |
| Richard    | N. | Henson      |
| Katja      |    | Heuer       |
| Jacqueline |    | Hoare       |
| Bharath    |    | Holla       |
| Avram      | J. | Holmes      |
| Hao        |    | Huang       |
| Kiho       |    | Im          |
| Jonathan   |    | Ipser       |
| Clifford   | R. | Jack Jr     |
| Andrea     | P. | Jackowski   |
| Tianye     |    | Jia         |
| David      | T. | Jones       |
| Peter      | B. | Jones       |
| Rene       | S  | Kahn        |
| Hasse      |    | Karlsson    |
| Linnea     |    | Karlsson    |
| Ryuta      |    | Kawashima   |
| Elizabeth  | A. | Kelley      |
| Silke      |    | Kern        |
| Ki-Woong   |    | Kim         |
| Manfred    | G. | Kitzbichler |
| William    | S. | Kremen      |
| François   |    | Lalonde     |
| Brigitte   |    | Landeau     |
| Jason      |    | Lerch       |
| John       | D. | Lewis       |
| Jiao       |    | Li          |
| Wei        |    | Liao        |
| Conor      |    | Liston      |
| Michael    | V. | Lombardo    |
| Jinglei    |    | Lv          |
| Travis     | T. | Mallard     |
| Machteld   |    | Marcelis    |
| Samuel     | R. | Mathias     |
| Bernard    |    | Mazoyer     |

|           |    |                |
|-----------|----|----------------|
| Philip    |    | McGuire        |
| Michael   | J. | Meaney         |
| Andrea    |    | Mechelli       |
| Bratislav |    | Misic          |
| Sarah     | E  | Morgan         |
| David     |    | Mothersill     |
| Cynthia   |    | Ortinou        |
| Rik       |    | Ossenkoppele   |
| Minhui    |    | Ouyang         |
| Lena      |    | Palaniyappan   |
| Leo       |    | Paly           |
| Pedro     | M. | Pan            |
| Christos  |    | Pantelis       |
| Min Tae   | M. | Park           |
| Tomas     |    | Paus           |
| Zdenka    |    | Pausova        |
| Deirel    |    | Paz-Linares    |
| Alexa     |    | Pichet Binette |
| Karen     |    | Pierce         |
| Xing      |    | Qian           |
| Anqi      |    | Qiu            |
| Armin     |    | Raznahan       |
| Timothy   |    | Rittman        |
| Amanda    |    | Rodrigue       |
| Caitlin   | K. | Rollins        |
| Rafael    |    | Romero-Garcia  |
| Lisa      |    | Ronan          |
| Monica    | D. | Rosenberg      |
| David     | H. | Rowitch        |
| Giovanni  | A. | Salum          |
| Theodore  | D. | Satterthwaite  |
| H. Lina   |    | Schaare        |
| Jenna     |    | Schabdach      |
| Russell   | J. | Schachar       |
| Michael   |    | Schöll         |
| Aaron     | P. | Schultz        |
| Jakob     |    | Seidlitz       |

|             |    |                 |
|-------------|----|-----------------|
| David       |    | Sharp           |
| Russell     | T. | Shinohara       |
| Ingmar      |    | Skoog           |
| Christopher | D. | Smyser          |
| Reisa       | A. | Sperling        |
| Dan         | J. | Stein           |
| Aleks       |    | Stolicyn        |
| John        |    | Suckling        |
| Gemma       |    | Sullivan        |
| Benjamin    |    | Thyreau         |
| Roberto     |    | Toro            |
| Nicolas     |    | Traut           |
| Kamen       | A. | Tsvetanov       |
| Nicholas    | B. | Turk-Browne     |
| Jetro       | J. | Tuulari         |
| Christophe  |    | Tzourio         |
| Étienne     |    | Vachon-Presseau |
| Mitchell    | J. | Valdes-Sosa     |
| Pedro       | A. | Valdes-Sosa     |
| Sofie       | L. | Valk            |
| Therese     |    | van Amelsvoort  |
| Simon       | N. | Vandekar        |
| Lana        |    | Vasung          |
| Petra       | E  | Vértes          |
| Lindsay     | W. | Victoria        |
| Sylvia      |    | Villeneuve      |
| Arno        |    | Villringer      |
| Jacob       | W. | Vogel           |
| Konrad      |    | Wagstyl         |
| Yin-Shan    | S. | Wang            |
| Simon       | K. | Warfield        |
| Varun       |    | Warrier         |
| Eric        |    | Westman         |
| Margaret    | L. | Westwater       |
| Heather     | C. | Whalley         |
| Simon       | R. | White           |
| A. Veronica |    | Witte           |

|             |    |            |
|-------------|----|------------|
| Ning        |    | Yang       |
| B.T. Thomas |    | Yeo        |
| Hyuk Jin    |    | Yun        |
| Andrew      |    | Zalesky    |
| Heather     | J  | Zar        |
| Anna        |    | Zettergren |
| Juan        | H. | Zhou       |
| Hisham      |    | Ziauddeen  |
| Dabriel     |    | Zimmerman  |
| Andre       |    | Zugman     |
| Xi-Nian     | N. | Zuo        |