Cell Reports, Volume *42*

Supplemental information

The molecular genetic landscape of human brain size variation

Jakob Seidlitz, Travis T. Mallard, Jacob W. Vogel, Younga H. Lee, Varun Warrier, Gareth Ball, Oskar Hansson, Leanna M. Hernandez, Ayan S. Mandal, Konrad Wagstyl, Michael V. Lombardo, Eric Courchesne, Joseph T. Glessner, Theodore D. Satterthwaite, Richard A.I. Bethlehem, Joshua D. Bernstock, Lifespan Brain Chart Consortium, Shinya Tasaki, Bernard Ng, Chris Gaiteri, Jordan W. Smoller, Tian Ge, Raquel E. Gur, Michael J. Gandal, and Aaron F. Alexander-Bloch

The molecular genetic landscape of human brain size variation

Jakob Seidlitz[1,2,3,*], Travis T. Mallard[4,5,6], Jacob W. Vogel[3,7], Younga H. Lee[4,5,6], Varun Warrier[8,9], Gareth Ball[10,11], Oskar Hansson[12,13], Leanna M. Hernandez[14], Ayan S. Mandal[1,2,3], Konrad Wagstyl[15], Michael V. Lombardo[16], Eric Courchesne[17,18], Joseph T. Glessner[19,20], Theodore D. Satterthwaite[1,3,7], Richard A.I. Bethlehem[8], Joshua D. Bernstock[21,22,23], Lifespan Brain Chart Consortium, Shinya Tasaki[24], Bernard Ng[24], Chris Gaiteri[24], Jordan W. Smoller[4,5,6,25], Tian Ge[4,5,6,25], Raquel E. Gur[1,2,3], Michael J. Gandal[1,3,26,27], Aaron F. Alexander-Bloch[1,2,3,27]

1. Lifespan Brain Institute, The Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, 19104 USA
2. Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA
3. Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, 19104 USA
4. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 02114 USA
5. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA, 02142 USA
6. Department of Psychiatry, Harvard Medical School, Boston, MA, 02142 USA
7. Lifespan Informatics and Neuroimaging Center, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104 USA
8. Department of Psychiatry, University of Cambridge, Cambridge, CB2 1TN UK
9. Department of Psychology, University of Cambridge, Cambridge, CB2 1TN UK
10. Developmental Imaging, Murdoch Children's Research Institute, Melbourne, Victoria, 3052 Australia
11. Department of Paediatrics, University of Melbourne, Melbourne, Victoria, 3052 Australia
12. Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Malmö, P663+Q9 Sweden
13. Memory Clinic, Skåne University Hospital, Malmö, P663+Q9 Sweden
14. Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, 90024 USA
15. Wellcome Centre for Human Neuroimaging, University College London, London, WC1N 3AR UK
16. Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, 38068 Italy
17. Department of Neuroscience, University of California, San Diego, San Diego, CA, 92093 USA
18. Autism Center of Excellence, University of California, San Diego, San Diego, CA, 92093 USA
19. The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA
20. Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104 USA
21. Department of Neurosurgery, Brigham and Women's Hospital, Harvard University; Boston, MA, 02115 USA
22. Department of Neurosurgery, Boston Children's Hospital, Harvard University; Boston, MA, 02115 USA
23. David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA
24. Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612 USA

25. Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, 02114 USA
26. Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104 USA
27. These authors contributed equally

* Lead Contact: jakob.seidlitz@pennmedicine.upenn.edu, seidlitzj@chop.edu

Supplemental figures

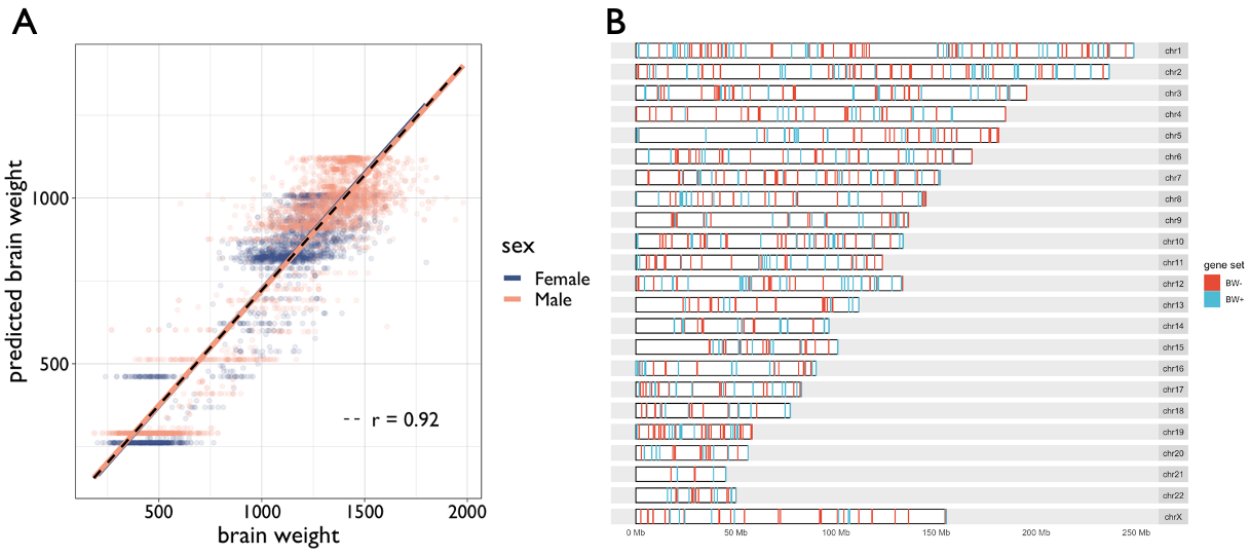

Figure S1. Validation of brain weight as a marker of overall brain size and genomic annotation of brain weight transcriptome association gene sets. A | Scatterplot showing the relationship between empirical brain weight (grams) and predicted brain weight (grams), based on lifespan neuroimaging models of brain size (total cerebrum volume) from a prior study (Total N=3,689; PsychENCODE N=1,670; GTEx N=227; ROSMAP N=634; independent aggregated dataset without associated transcriptomics N=1,158 [1]). Colors denote biological sex. Lines show linear fits. Black dotted line shows overall fit (Pearson r=0.92, P < 2.2e-16). **B |** Karyogram with colored lines indicating locations of genes with significant ($P_{Bonferroni}$ < 0.05) upregulated expression in smaller-brained (brain weight negative or "BW-") or larger-brained (brain weight positive or "BW+") individuals in the PsychENCODE dataset.

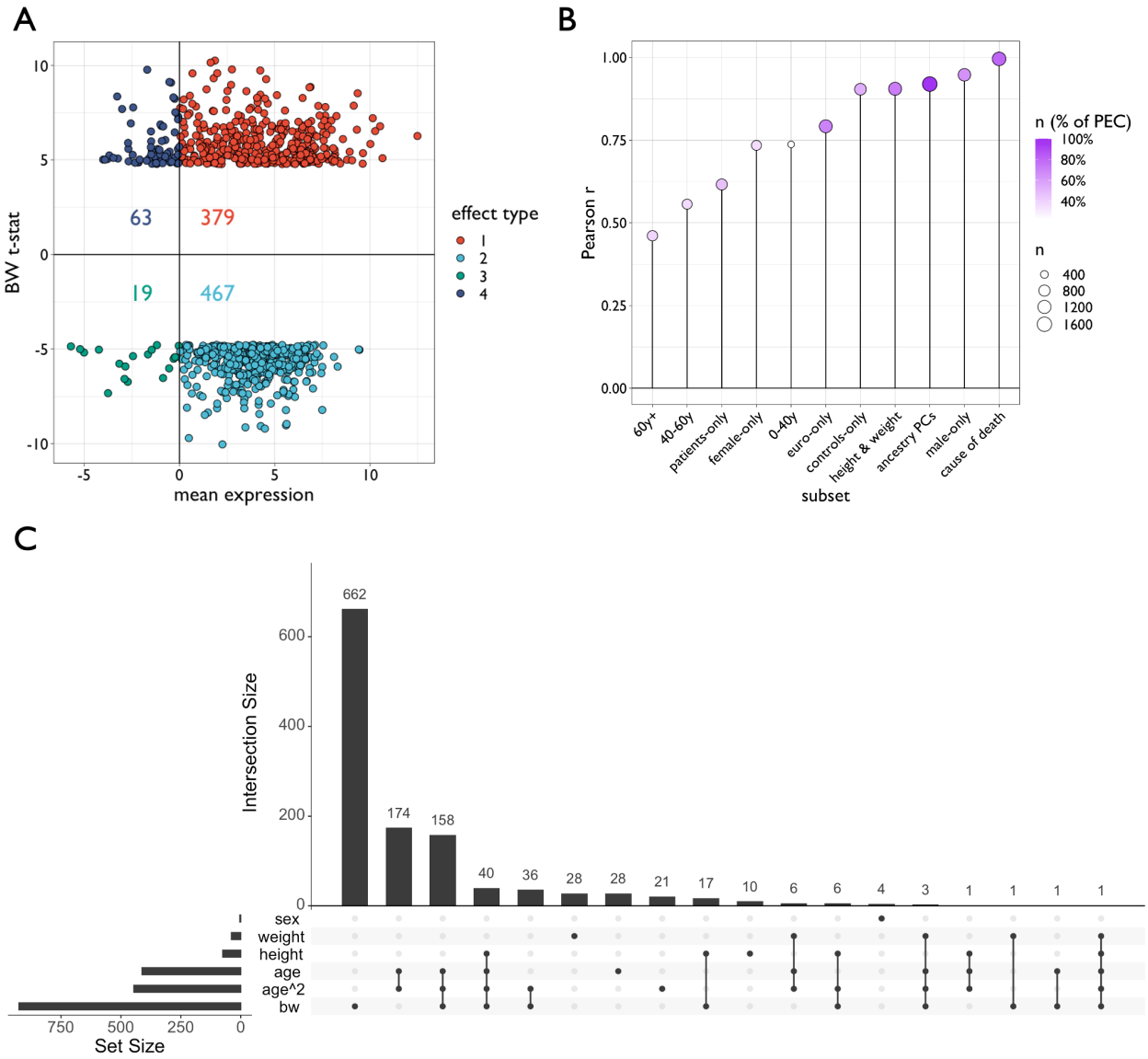

Figure S2. Sensitivity analyses of brain weight gene expression models. A | scatterplot showing the relationship between the average expression (z-score) and brain weight (BW) association (t-statistic) in the main PsychENCODE (PEC) dataset. Each dot represents a gene that was found to have a significant association ($P_{Bonferroni}$ < 0.05) between brain weight and expression across samples. Colors and numerical types are arbitrary. Both BW+ and BW- genes (positive and negative values on y-axis, respectively) disproportionately reflect genes with generally higher expression (more positive values on x-axis), which allows for the opposing interpretation of the two gene sets. However, a small percentage of these two gene sets shows the opposite pattern – BW+ and BW- genes having lower expression in higher or lower BW (63 genes and 19 genes, respectively). Numbers represent the counts of genes within each type. **B |** Lollipop plot showing the Pearson correlation coefficient across genes between the main PEC model output and output generated based on subsets of the PEC dataset. Circle color is the percentage of the total PEC sample in each subset, and size is the absolute sample size of each subset. 60+y=donors aged 60 years and older, 40-60y=donors aged 40-60 years,

patients-only=just donors with a documented psychiatric diagnosis, female-only=just female donors, 0-40y=donors aged 0-40 years, controls-only=just donors without a documented psychiatric diagnosis, height & weight=donors with height and weight information (height and weight included as covariates in models), male-only=just male donors, euro-only=just donors of european ancestry, ancestry PCs=inclusion of ancestry principal component scores based on genotype in gene models, cause of death=cause of death included as a covariate in models. **C |** Upset plot demonstrating the number and overlap of significant ($P_{Bonferroni} < 0.05$) genes for biological traits in the PsychENCODE discovery cohort. For brain weight (“bw”), sex, linear age (“age”) and squared age (“age^2”) the original model was used. Since only a subset of subjects had documented height and weight, those terms were included as additional covariates in separate gene-level models (e.g., as part of the analyses in panel B).

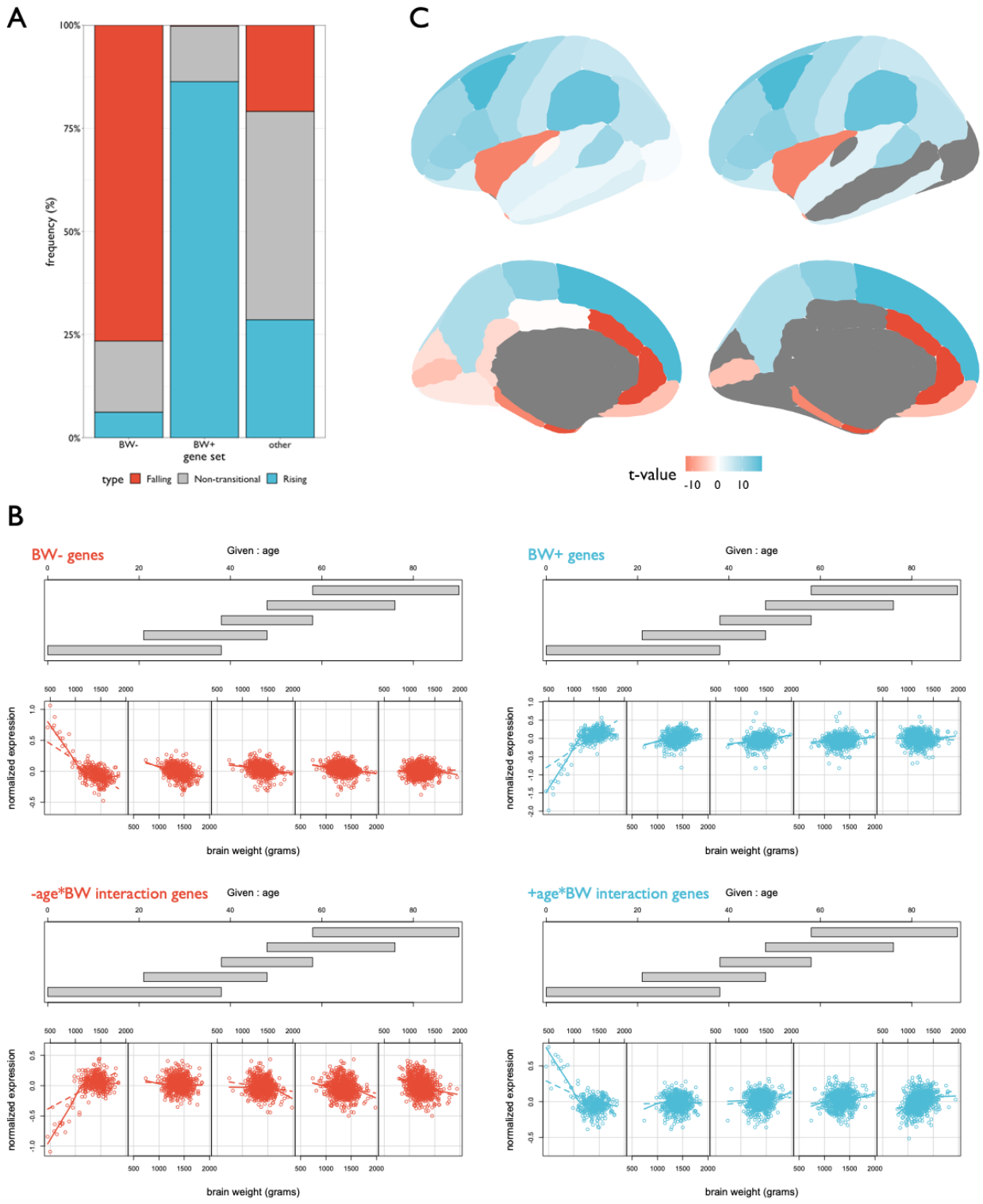

Figure S3. Spatiotemporal evaluation of the brain weight associated genes. A | Stacked bar plots showing the proportion of genes within each BW gene set that fall within predefined categories of perinatal expression trajectories (Werling et al., 2020). Colors were set based on qualitative assessment of the strong bias for “falling” (BW-, red) and “rising“ (BW+, blue) effects. The “other” gene set type refers to all genes not a part of the BW sets. **B |** Conditioning plots

showing the interaction effects of brain weight and age on gene expression. Age ranges (gray bars) were determined based on the distribution of age values. Top row: expression of brain weight gene sets across age, shown for reference. Bottom row: same plots shown for genes that showed brain weight by age interaction effects ($P_{Bonferroni} < 0.05$) that were not in the original brain weight sets – 92 genes had negative coefficients (red, left) and 130 genes had positive coefficients (blue, right). Normalized expression was averaged across genes for visualization of age-related trends. The number of age strata (N=5) and percent overlap (50%) were arbitrarily chosen for visualization. Dashed lines denote linear fits, while solid lines denote nonlinear (LOESS) fits. **C |** Brain plots showing the BW+/- relative expression differences (t-values), using a coarser cortical parcellation of the Allen Human Brain Atlas. The left column shows the complete set of effects for each region, and the right column shows only significant effects ($P_{Bonferroni} < 0.05$; gray values $P_{Bonferroni} > 0.05$). The top row shows the lateral view of the left hemisphere, and the bottom row shows the medial view.

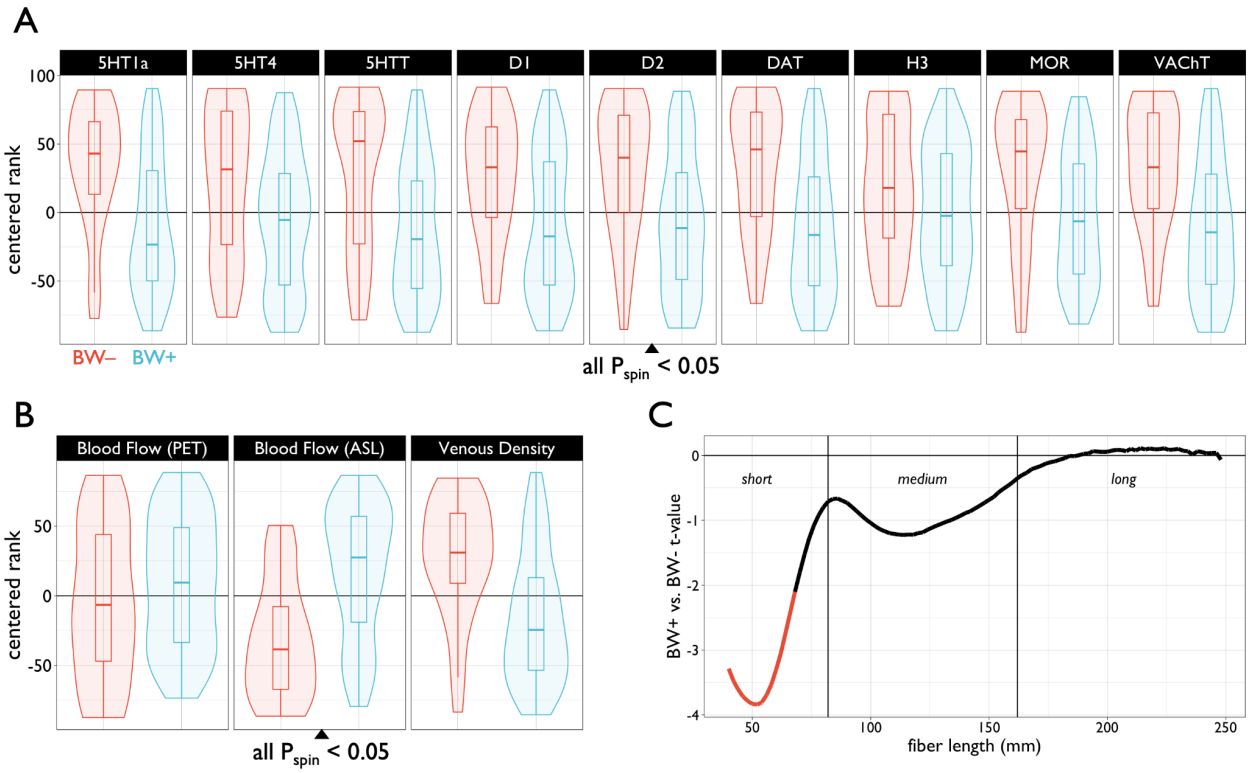

Figure S4. Heterotopic contextualization of brain weight (BW) differential expression. **A |** Comparison of independently-measured neurotransmitter densities in significant regions of BW+ (blue) or BW- (red) differential expression in the Allen Human Brain Atlas. Differences between BW+ and BW- were quantified with t-tests, and all comparisons shown were significantly different from statistics generated from spatially-permuted differential expression maps (P_{spin} < 0.05). Note that all relationships show greater BW- values. Centered rank values were computed by ranking (from low to high) regional estimates for each expression type and subtracting the median rank, which was used to aid in comparison across maps. **B |** Same as A, instead comparing areas of differential expression in terms of neuroimaging-derived estimates of cerebral blood flow and venous density. **C |** Comparison of prevalence of white matter fiber lengths across areas of BW differential expression. T-values were computed across BW+ versus BW- regions in terms of fiber density values at varying fiber lengths. The black line denotes fiber lengths where BW regional differences were insignificant ($P_{Bonferroni}$ > 0.05), red where BW- > BW+ ($P_{Bonferroni}$ < 0.05). PET: positron emission tomography, ASL: arterial spin labeling. See the Key Resources Table for further information on the neurotransmitter maps. All box-violin plots show median and interquartile range (IQR) with whiskers denoting 1.5 × IQR.

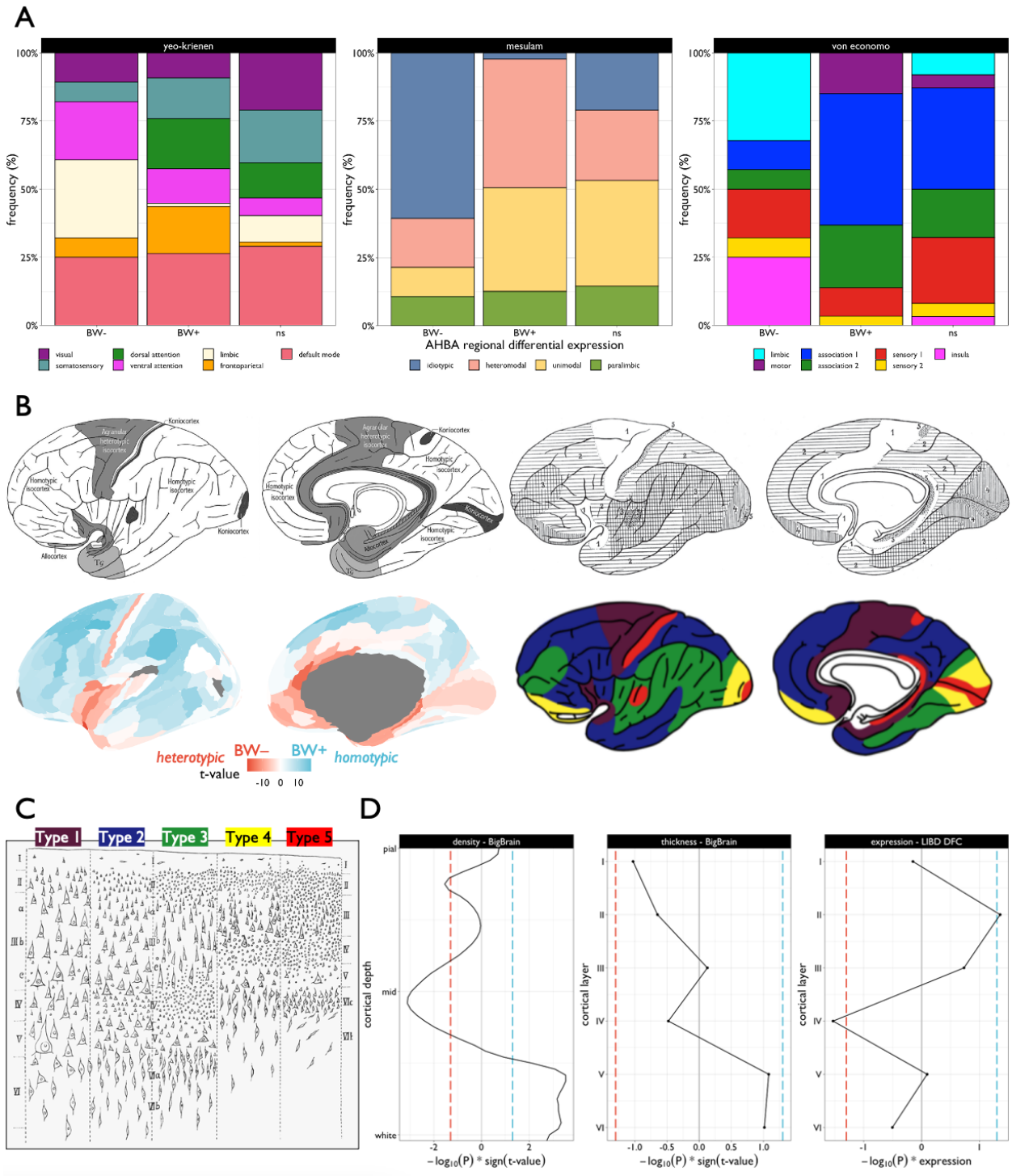

Figure S5. **Heterotopic categorization of regional brain weight relative expression. A |** Stacked barplots show the proportion of BW+ and BW- regions based on functional (left), hierarchical (middle), and cytoarchitectonic (right) classification. AHBA: Allen Human Brain Atlas, ns: non-significant regions of differential expression. **B |** Von Economo-Koskinas cortical atlases [2] based on cytoarchitectonic characteristics (top left), and subdivided based on one of

five cortical types (top right, colored in bottom right). The full map of the cortical region relative expression differences (t-statistics) of brain weight positive (BW+) versus brain weight negative (BW-) in the Allen Human Brain Atlas (similar to Figure 1, but unthresholded) is shown for qualitative comparison. **C |** The 5 described von Economo-Koskinas cortical types are shown in terms of their general profiles across layers. Notably, types 1 and 5 ("heterotypic isocortex" and "koniocortex", respectively) – which align with regions of BW- relative expression – are distinct in their agranular versus granular properties. However, both types 1 and 5 show more indistinguishable layer differentiation based on cytoarchitectonic features, compared to "homotypic isocortex" (mainly types 2 and 3) – which align with regions of of BW+ relative expression – contains more of the canonical 6-layer cortical cytoarchitecture. **D |** Laminar profile analysis of BW gene set relative expression differences in three independent datasets. Laminar density (left) and thickness (middle) based on 3D histological reconstruction in the BigBrain dataset characterizing areas of BW+ versus BW- relative expression differences across 100 cortical depths (left) and assigned layers (middle). The right-hand panel shows relative expression differences comparing BW+ versus BW- gene sets at each layer from the spatial RNA sequencing postmortem brain dataset (in the dorsolateral prefrontal cortex, DFC) from the Lieber Institute on Brain Development (LIBD). BigBrain data was parcellated according to the atlas used for the Allen Human Brain Atlas transcriptomic dataset and corresponding BW+ versus BW- relative expression map in **A**. Density values for each cortical depth and thickness values for each assigned layer within the BigBrain dataset were then compared across cortical areas (assigned BW+ or BW- based on the relative expression map in **A**). For the LIBD dataset, BW+ versus BW- relative expression was tested at each cortical layer. Colored dashed lines denote the negative log-scaled p-value significance thresholds (uncorrected $P < 0.05$, signed by the direction of effect) for BW- (red, negative) and BW+ (blue, positive) relative expression differences. See the Key Resources Table for more information on these datasets.

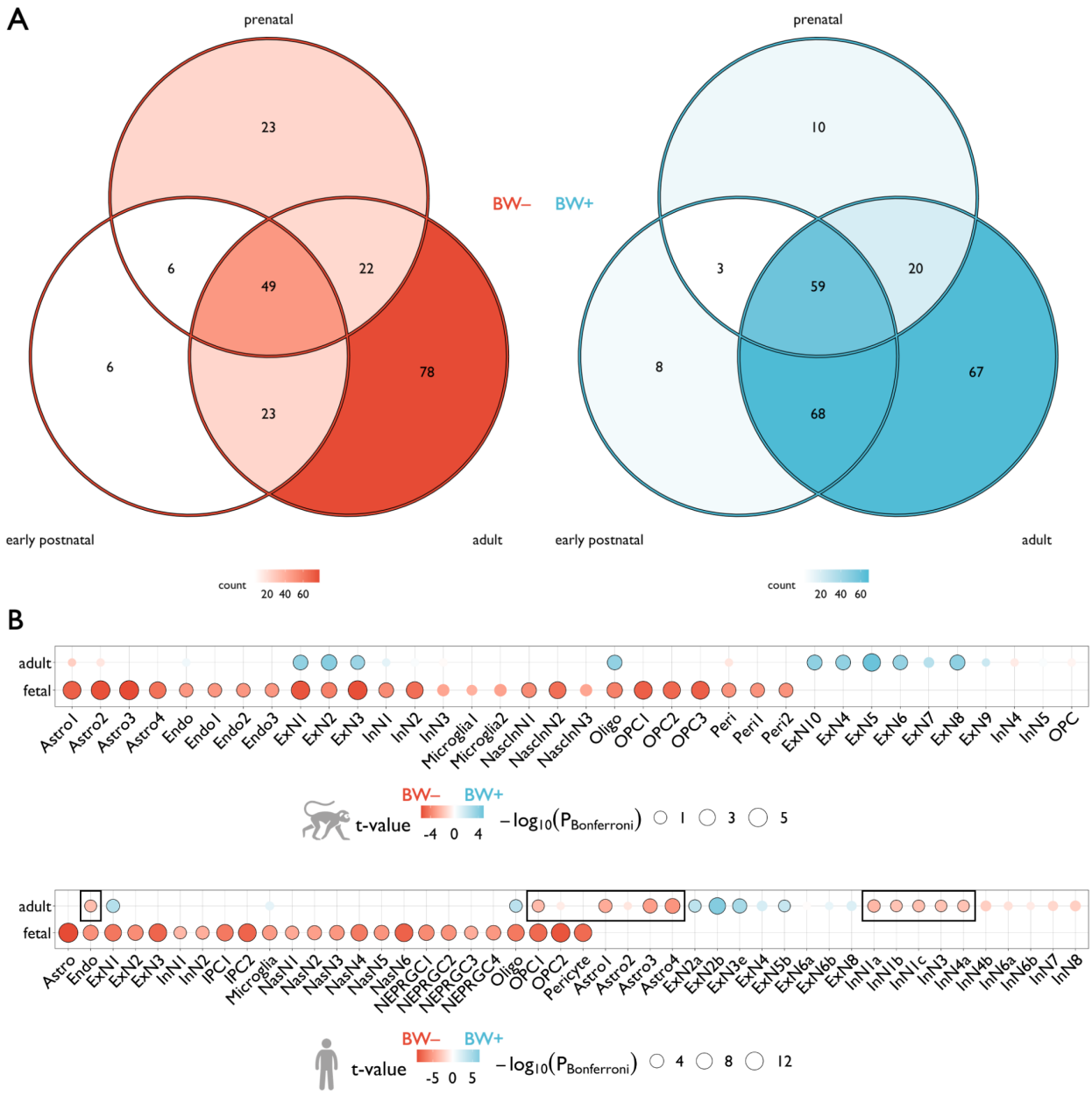

Figure S6. Cross-species, cell-type, and developmental bioinformatic contextualization of brain weight relative expression. A | Venn diagrams of brain weight negative (BW-) and brain weight positive (BW+) gene sets based on the significantly differentially expressed genes ($P_{Bonferroni}$ < 0.05) in humans relative to macaques in three developmental epochs. **B |** Relative expression of BW+ versus BW- genes across individual cell-types, using cell-specific RNA sequencing data in fetal and adult samples from macaques (top) and humans (bottom). BW- relative expression (red) indicates that BW- genes are more highly expressed in that cell type compared to BW+, whereas BW+ relative expression (blue) indicates the opposite effect. Black outlines denote significant effects ($P_{Bonferroni}$ < 0.05). Circles are scaled according to Bonferroni-corrected p-values. Black rectangles denote human-specific effects relative to macaques. X-axis labels are arbitrarily ordered based on alphabetized fetal cell subtype labels.

Alignment of adult cell subtype labels is purely based on shared naming, for visualization purposes. Astro: astrocytes, Endo: endothelial cells, ExN: excitatory neurons, InN: inhibitory neurons, NasN/NascInN: nascent neurons, Oligo: oligodendrocytes, OPC: oligodendrocyte precursor cells. Peri: pericytes.

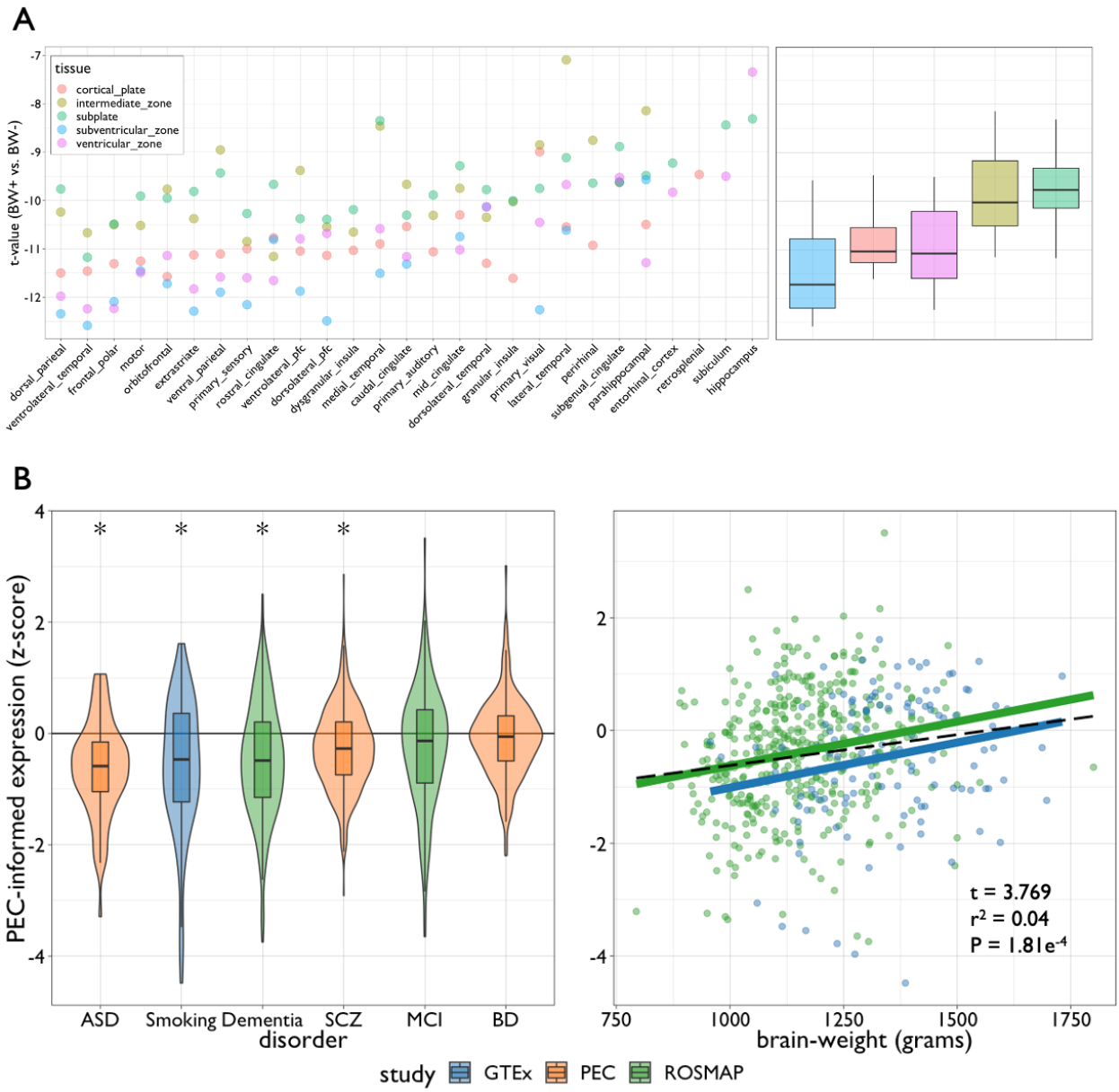

Figure S7. Prenatal spatial annotation of brain weight relative expression and validation of brain weight transcriptome scoring in independent datasets. A | Differential expression of brain weight positive (BW+) versus brain weight negative (BW-) gene sets in an independent prenatal developmental dataset (BrainSpan) across brain areas and zones during midgestation. T-statistics show all areas (circles, left) and zones (boxplots, right) show significant BW- differential expression, with notable gradated variation across zones – the subventricular zone showing greatest BW- differential expression relative to the other tissue types (t = -2.22, P = 0.029). All boxplots show median and interquartile range (IQR) with whiskers denoting 1.5 × IQR. **B |** (Left) distribution of brain weight (BW) transcriptome scores in patients with a diagnosed psychiatric or neurological disorder. BW transcriptome scores were computed using the PsychENCODE (PEC) BW model coefficients for all genes. Scores for each subject were generated by summing the weighted expression across all genes (dot product of all gene-wise

coefficients of the PEC brain weight models and gene expression), and normalized (z-scored) relative to their respective study controls. Smoking refers to smoking status, recorded as a binary variable upon autopsy. ASD: Autism Spectrum Disorder, AD: Alzheimer Disease, SCZ: Schizophrenia, MCI: Mild Cognitive Impairment, BD: Bipolar Disorder. Asterisks represent nominal significance (uncorrected $P < 0.05$) of disorder groupwise mean score compared to zero. All box-violin plots show median and interquartile range (IQR) with whiskers denoting $1.5 \times$ IQR. (Right) Scatterplot of the relationship between BW and BW transcriptome score. Black line and reported statistics represent the fit across datasets using a linear mixed effect model with fixed effects of age and sex, and random effects of diagnostic category (see left panel) and study.

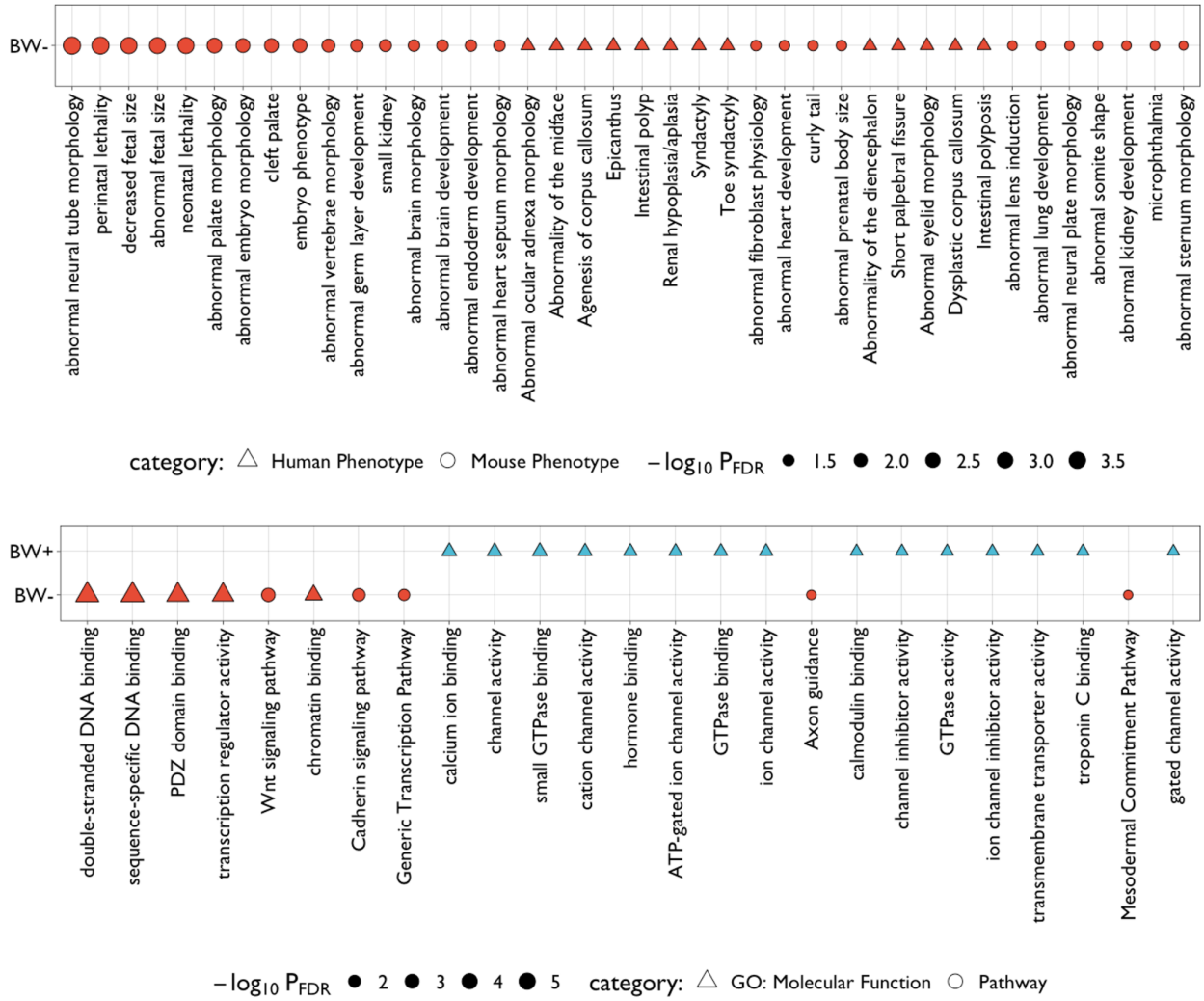

Figure S8. **Functional enrichment of brain weight gene sets.** Grid plot showing significant (P_{FDR} < 0.05, see Table S4) gene set enrichment of brain weight (BW) associated genes for Human and Mouse Phenotypes (top) and GO: Molecular Function and Pathway (bottom) using ToppGene. Shapes are sized according to adjusted negative log-scaled p-values.

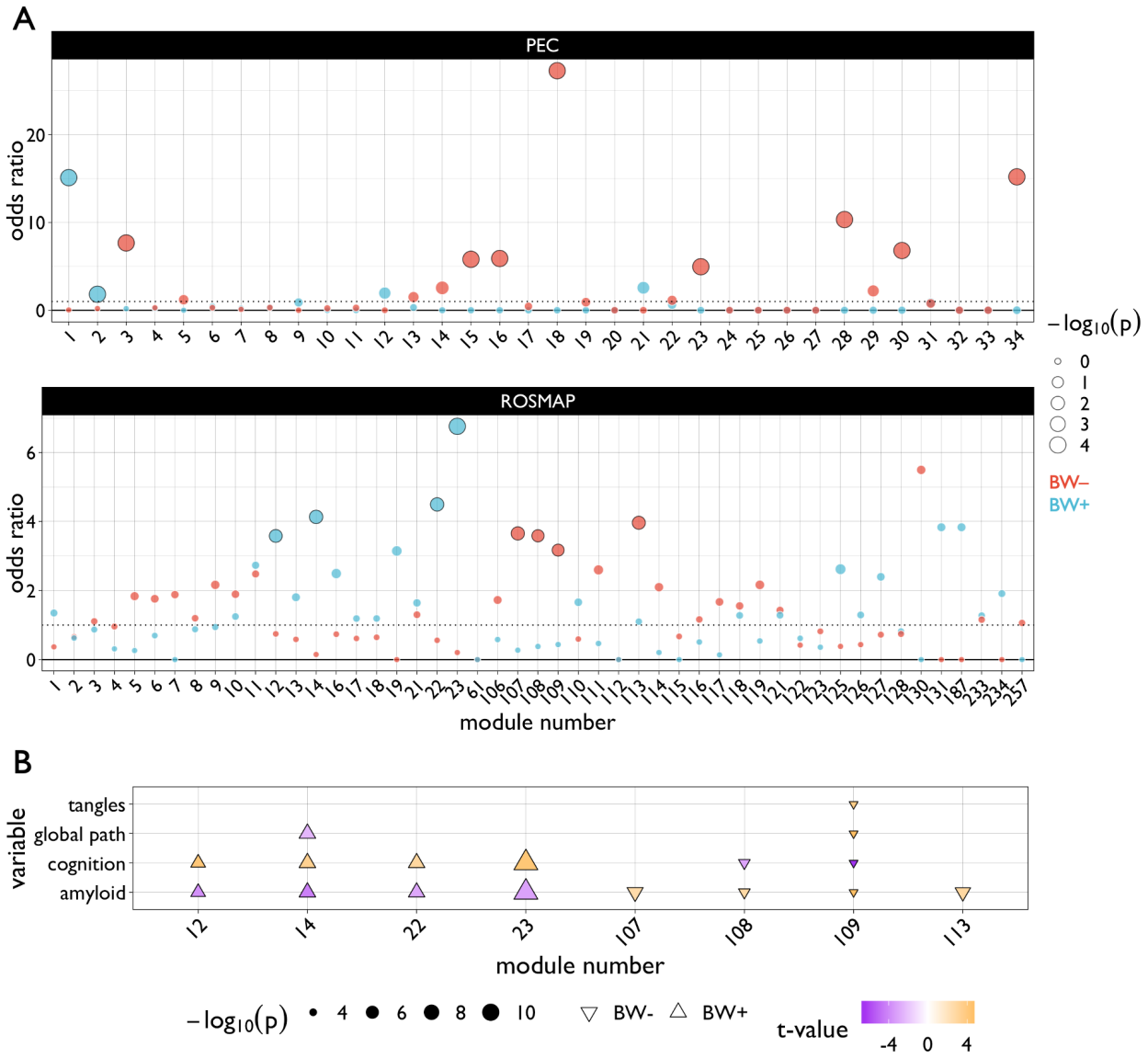

Figure S9. Enrichment of brain weight gene sets in co-expression modules across developmental and aging datasets, and in measures of neuropathology. A | Enrichment of brain weight (BW) gene sets in gene co-expression modules in the PsychENCODE (PEC) and ROSMAP datasets. Modules were derived in previous work [3,4], and gene set enrichment was calculated using hypergeometric tests. The size of the circles denotes negative log-scaled uncorrected p-values, with color corresponding to the BW gene set. Circles with black outlines are significant at $P_{FDR} < 0.05$ (corrected across number of modules). Dashed line denotes odds ratio = 1. See also Table S5 from the original PEC paper [3] which generated these module classifications for cell-type and disease enrichments. **B |** Annotations of significant ($P_{FDR} < 0.05$) BW-enriched modules based on metrics derived from postmortem neuropathology in the ROSMAP cohort. Shapes correspond to which BW gene set is enriched in a given module, and colors represent t-statistics from linear models comparing module eigengene scores to each neuropathology metric across individuals. Only uncorrected $P < 0.05$ effects are shown for visualization. See Table S6 in this paper for complete statistics in the ROSMAP cohort.

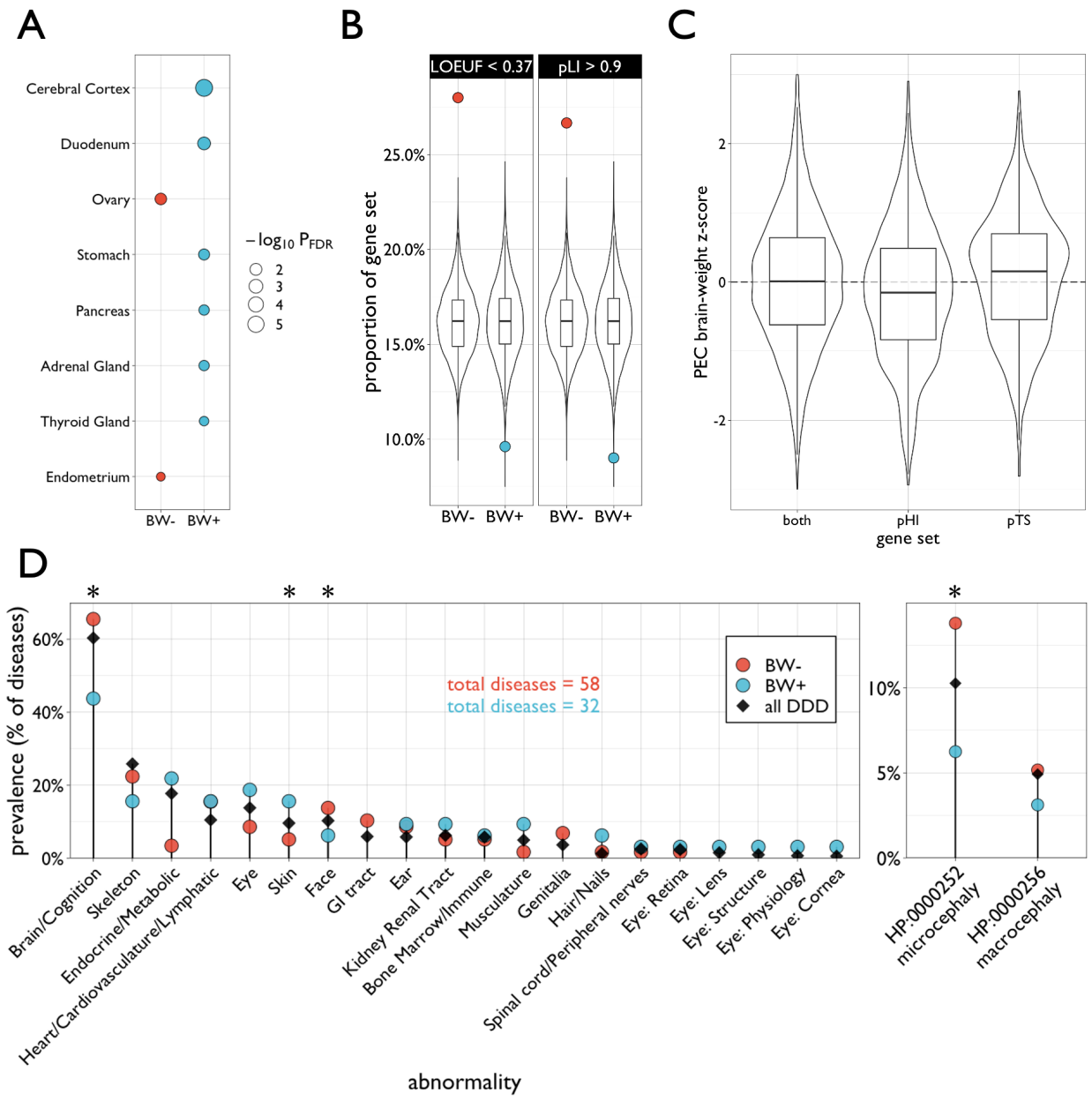

Figure S10. Tissue-specific and loss of function enrichment of brain weight gene sets. A | Gene set enrichment analysis (hypergeometric test) of brain weight (BW) gene sets across tissues in the Human Protein Atlas. For visualization only significant (P_{FDR} < 0.05) results are shown. Circles are colored according to BW gene sets and sized according to negative log-scaled p-values. **B |** Enrichment analysis of BW gene sets based on proportion of genes showing high loss-of-function intolerance based on two validated metrics derived from an external whole exome sequencing dataset (gnomAD v2.1.1). Recommended thresholds of each metric allowed for the LOEUF and pLI scores to be comparable. Colored circles represent the proportion of genes in BW gene sets meeting each established threshold (LOEUF < 0.37 or pLI > 0.9), and box-violin plots show the proportion overlap distribution across resampled 10,000

gene sets of similar size to the BW+ and BW- sets. **C |** Box-violin plots showing differential gene expression coefficients for the brain weight analysis in PsychENCODE (PEC) across gene sets based on dosage sensitivity probabilities for haploinsufficiency (pHI) and triplosensitivity (pTS) . Gene sets were established using predefined dosage sensitivity thresholds (pHI > 0.86 and pTS > 0.94). The pHI and pTS gene sets were further refined to exclude genes meeting both probability thresholds (constituting the “both” category). **D |** Prevalence of associated phenotypic symptomatology in diseases with de novo mutations in overlapping BW genes in the Deciphering Developmental Disorders (DDD) dataset (DECIPHER v11). Colored circles represent the prevalence of respective phenotypes within the 58 developmental disorders of overlapping BW- genes and 32 developmental disorders of overlapping BW+ genes. Black diamond represents the prevalence of respective phenotypes across all documented disorders in the database. Cephalic conditions were separated out based on identification via associated Human Phenotype (HP) documentation. Red and/or blue circles may be obscured for some low prevalence phenotypes (e.g., “Eye: Lens”). Asterisks denote significant enrichment for a BW gene set above expected based on prevalence across all disorders (black diamond) and permutation testing ($P_{permutation} < 0.05$). All box-violin plots show median and interquartile range (IQR) with whiskers denoting 1.5 × IQR. See the Key Resources Table for more information on these datasets.

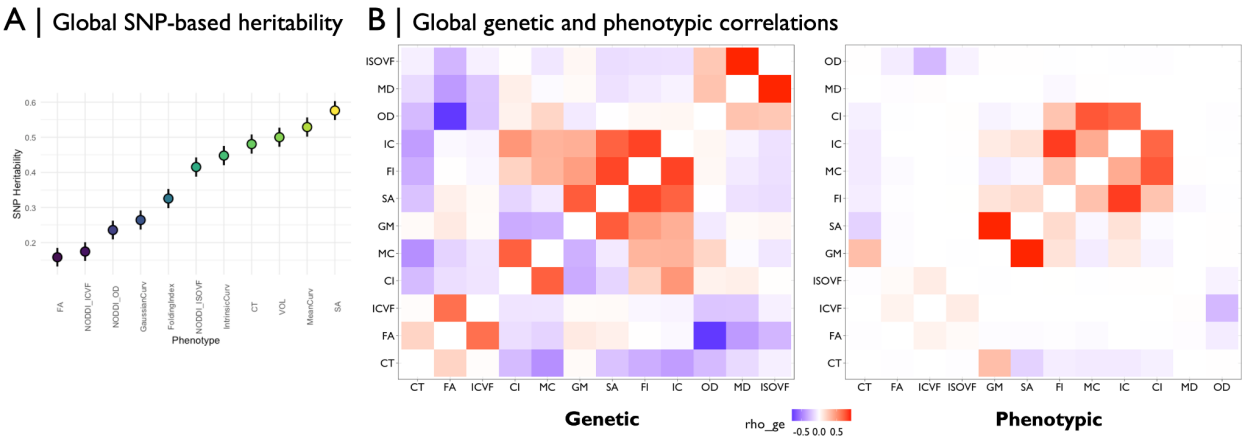

Figure S11. Multimodal neuroimaging transcriptome-wide association study (TWAS) results in UK Biobank. A | SNP-based heritability of global multimodal neuroimaging phenotypes (averaged or summed across the brain) in the UK Biobank. **B |** Genetic (left) or phenotypic (right) correlations ("rho_ge", using the 'RHOGE' package; https://github.com/bogdanlab/RHOGE) across the transcriptome-wide association study summary statistics for each global neuroimaging phenotype. FA=fractional anisotropy, NODDI_ICVF=intracellular volume fraction derived from neurite orientation dispersion and density imaging (NODDI), NODDI_OD=orientation dispersion derived from NODDI, GaussianCurv=Gaussian curvature, IntrinsicCurv=intrinsic curvature, NODDI_ISOVF=isotropic volume fraction derived from NODDI, CT=cortical thickness, VOL=gray matter volume, SA=surface area.

References

1. Borzage, M., Blüml, S., and Seri, I. (2014). Equations to describe brain size across the continuum of human lifespan. Brain Struct. Funct. *219*. 10.1007/s00429-012-0490-6.

2. Triarhou, L.C. (2007). The Economo-Koskinas Atlas Revisited: Cytoarchitectonics and Functional Context. SFN *85*, 195–203.

3. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*. 10.1126/science.aat8127.

4. Mostafavi, S., Gaiteri, C., Sullivan, S.E., White, C.C., Tasaki, S., Xu, J., Taga, M., Klein, H.-U., Patrick, E., Komashko, V., et al. (2018). A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat. Neurosci. *21*, 811–819.