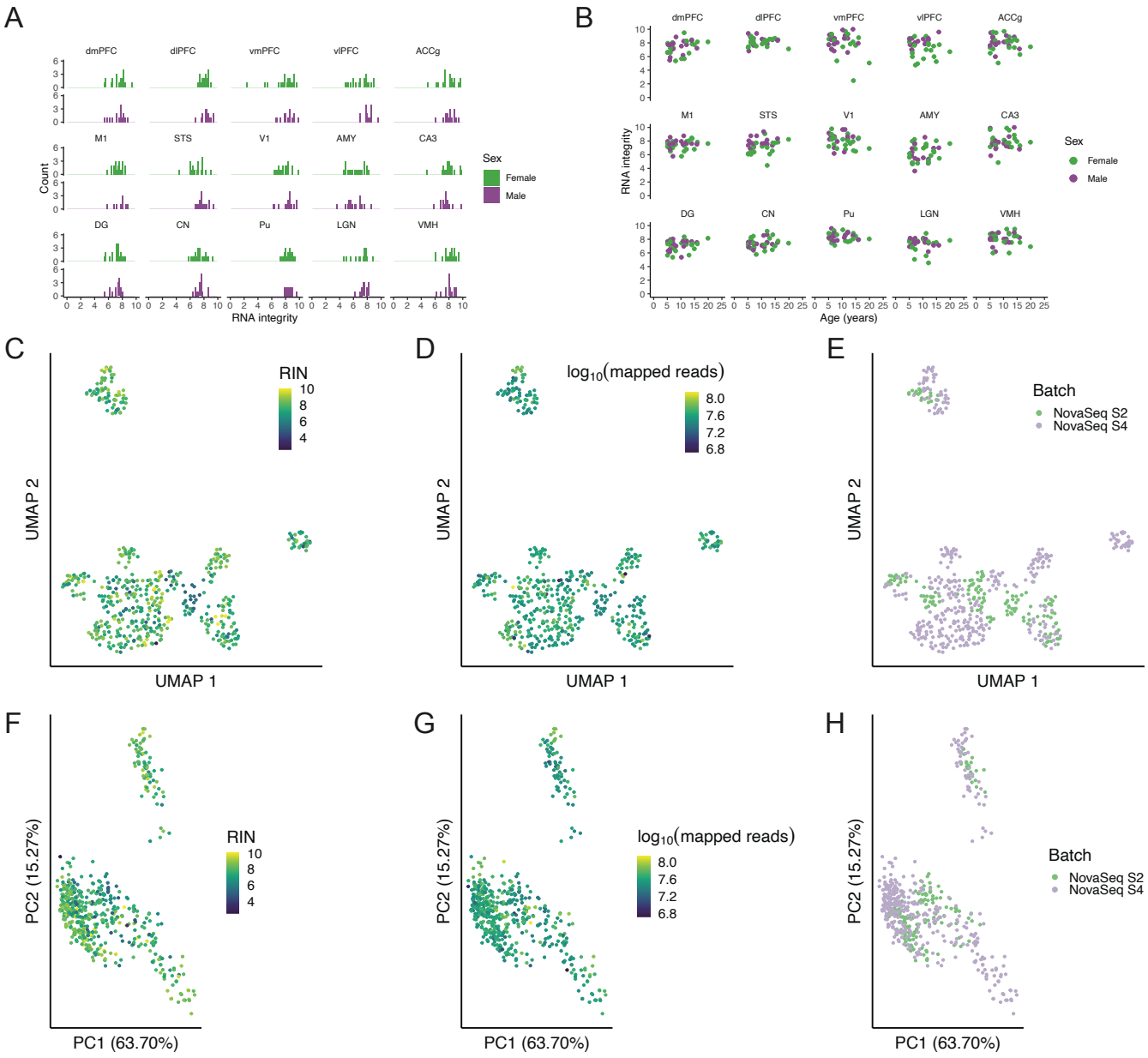
nature neuroscience

Resource

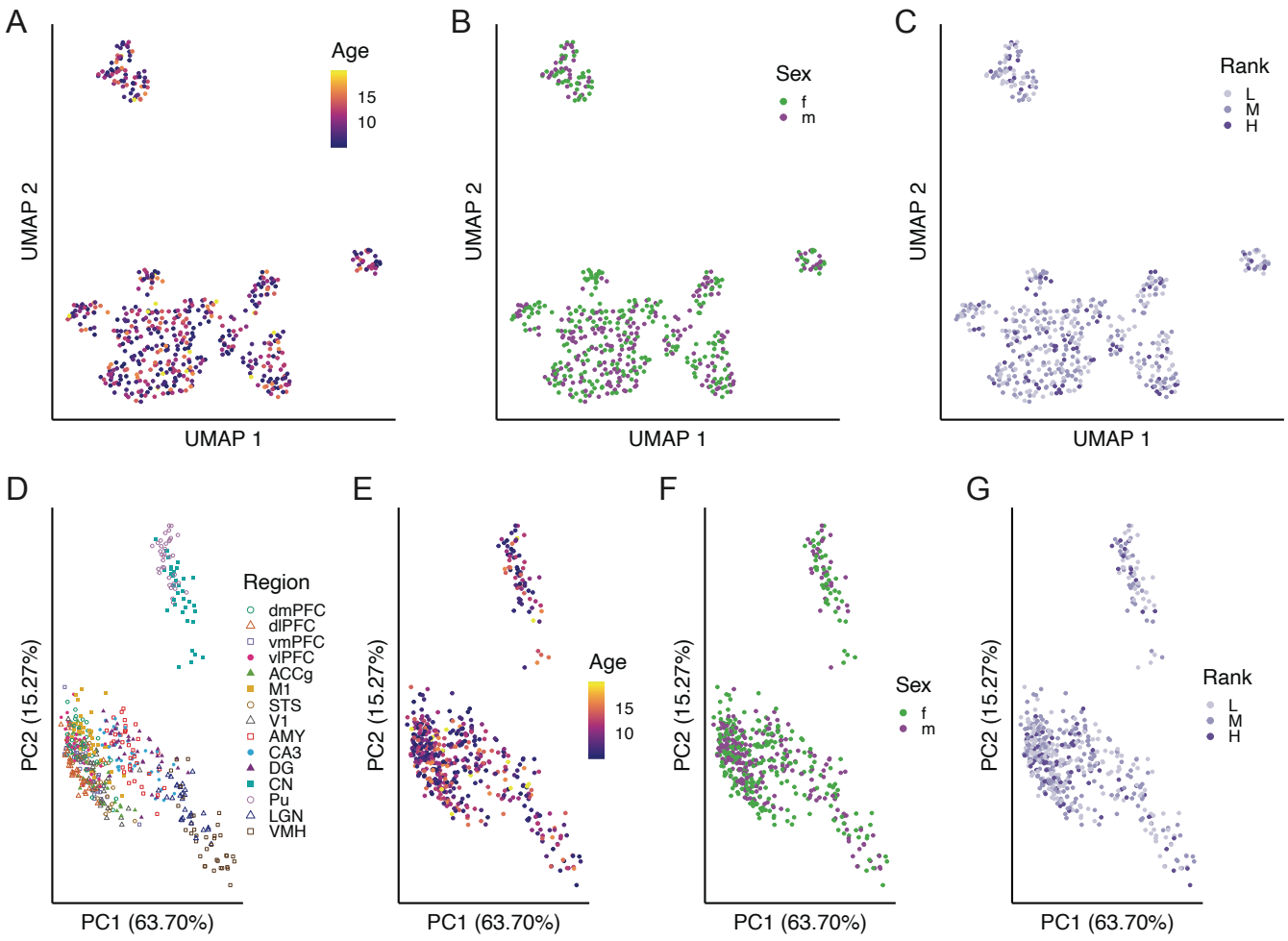
https://doi.org/10.1038/s41593-022-01197-0

Multiregion transcriptomic profiling of the primate brain reveals signatures of aging and the social environment

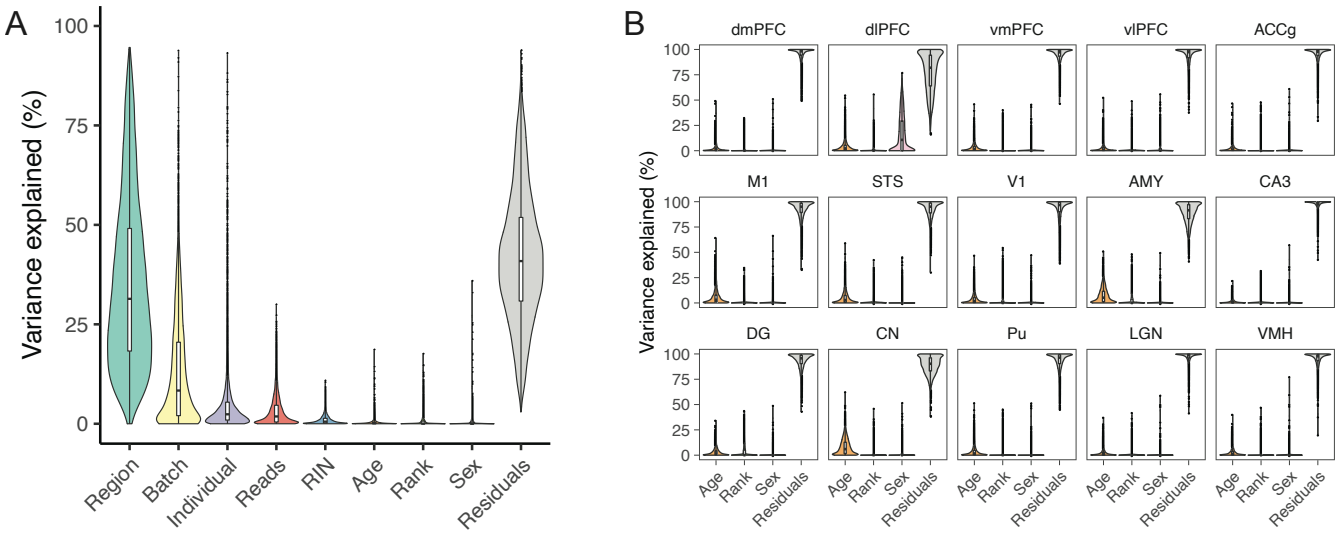
In the format provided by the authors and unedited



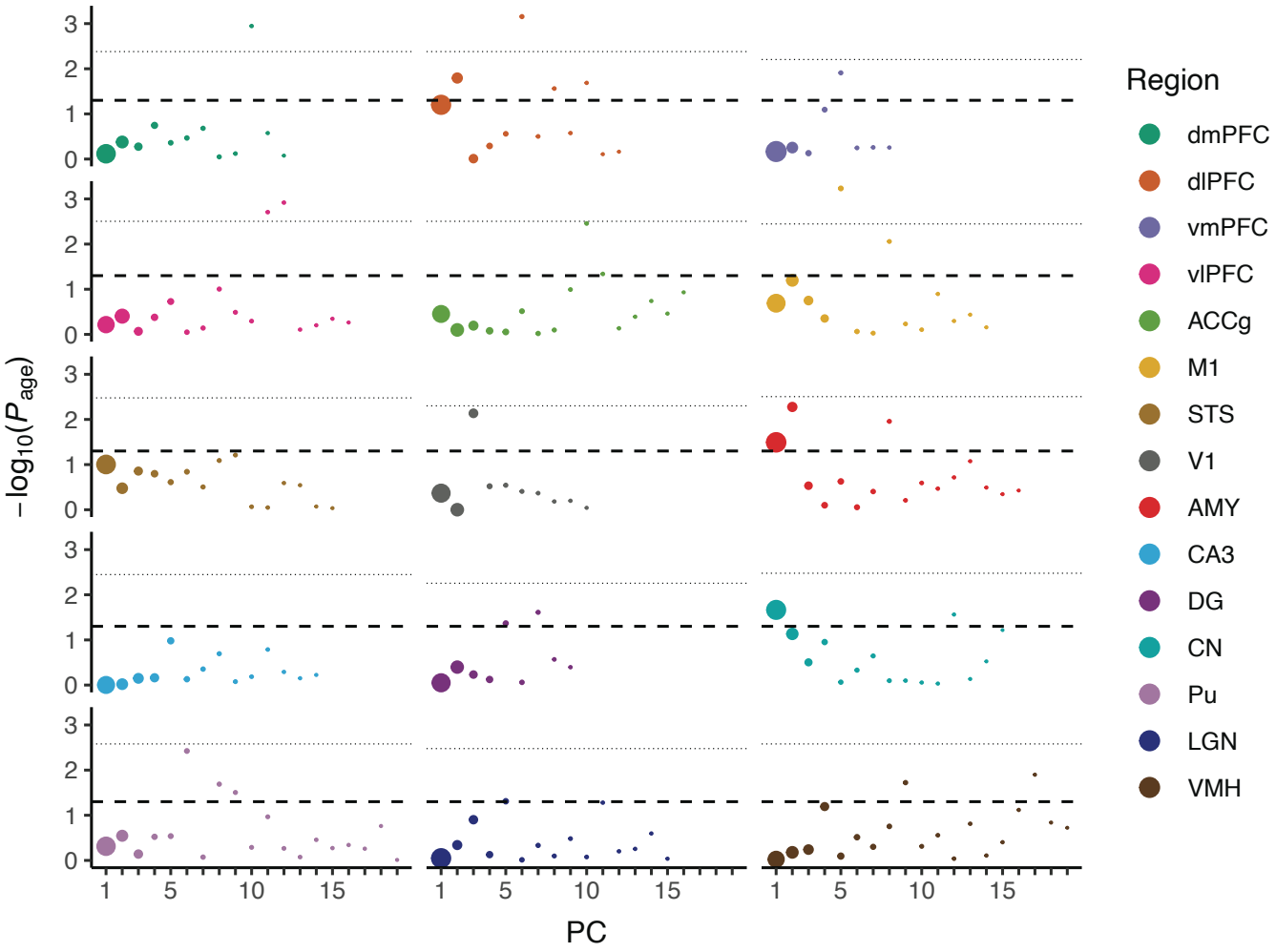
Supplementary Figure 1. Technical variation in the bulk tissue RNA sequencing data. (A) Distribution of RNA integrity (RIN) values are balanced between sexes and across brain regions. (B) RNA integrity (RIN) values are randomly distributed with respect to age. (C) RNA integrity (RIN) values, (D) number of reads mapping to the rhesus reference genome, and (E) sequencing batch visualized in UMAP dimensions. (F) RNA integrity (RIN) values, (G) number of reads mapping to the rhesus reference genome, and (H) sequencing batch visualized in the first two principal components.



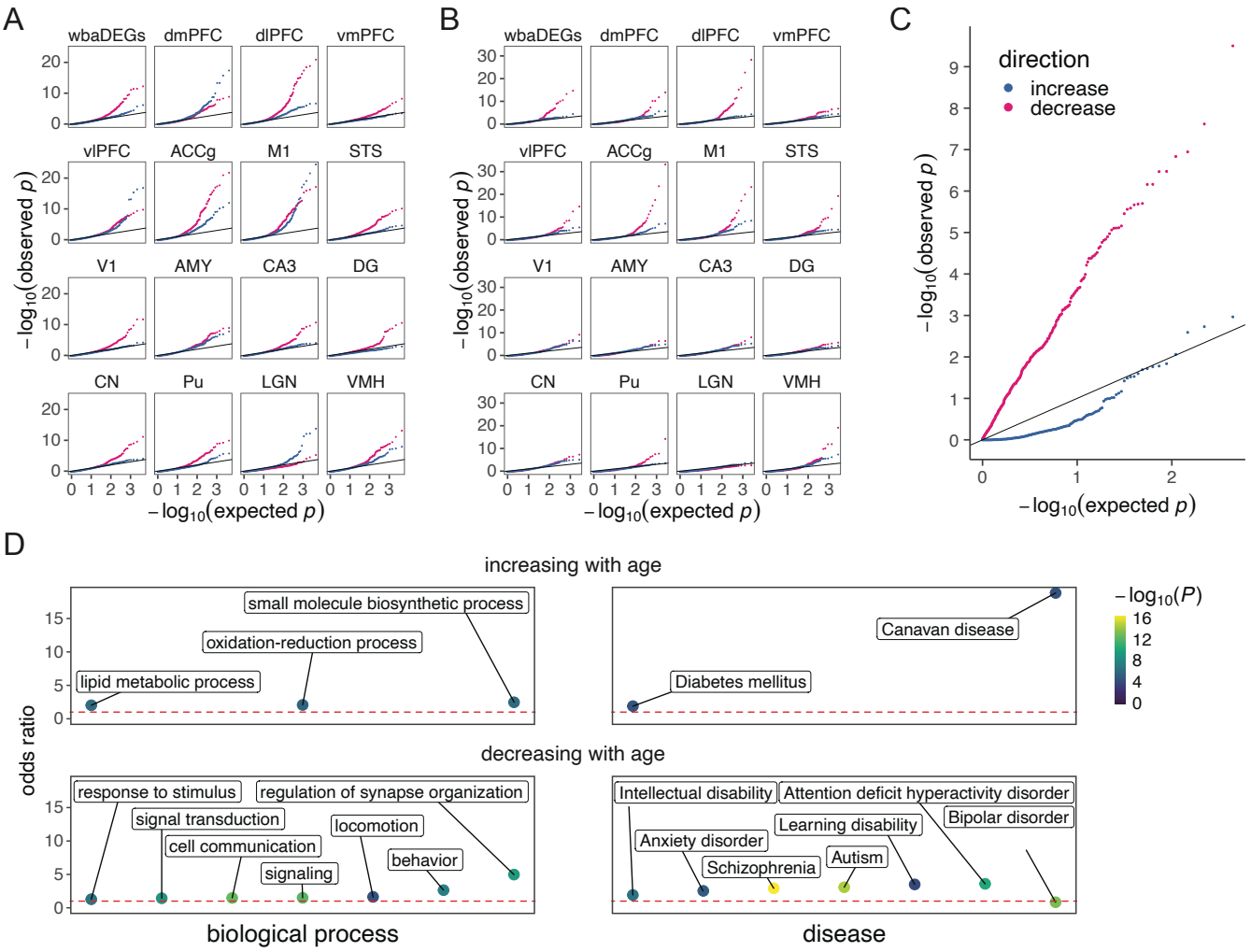
Supplementary Figure 2. Biological variation in the bulk tissue RNA sequencing data. (A) Age, (B) sex, and (C) dominance rank visualized in UMAP dimensions. (D) Brain region, (E) age, (F) sex, and (G) dominance rank visualized in the first two principal components.



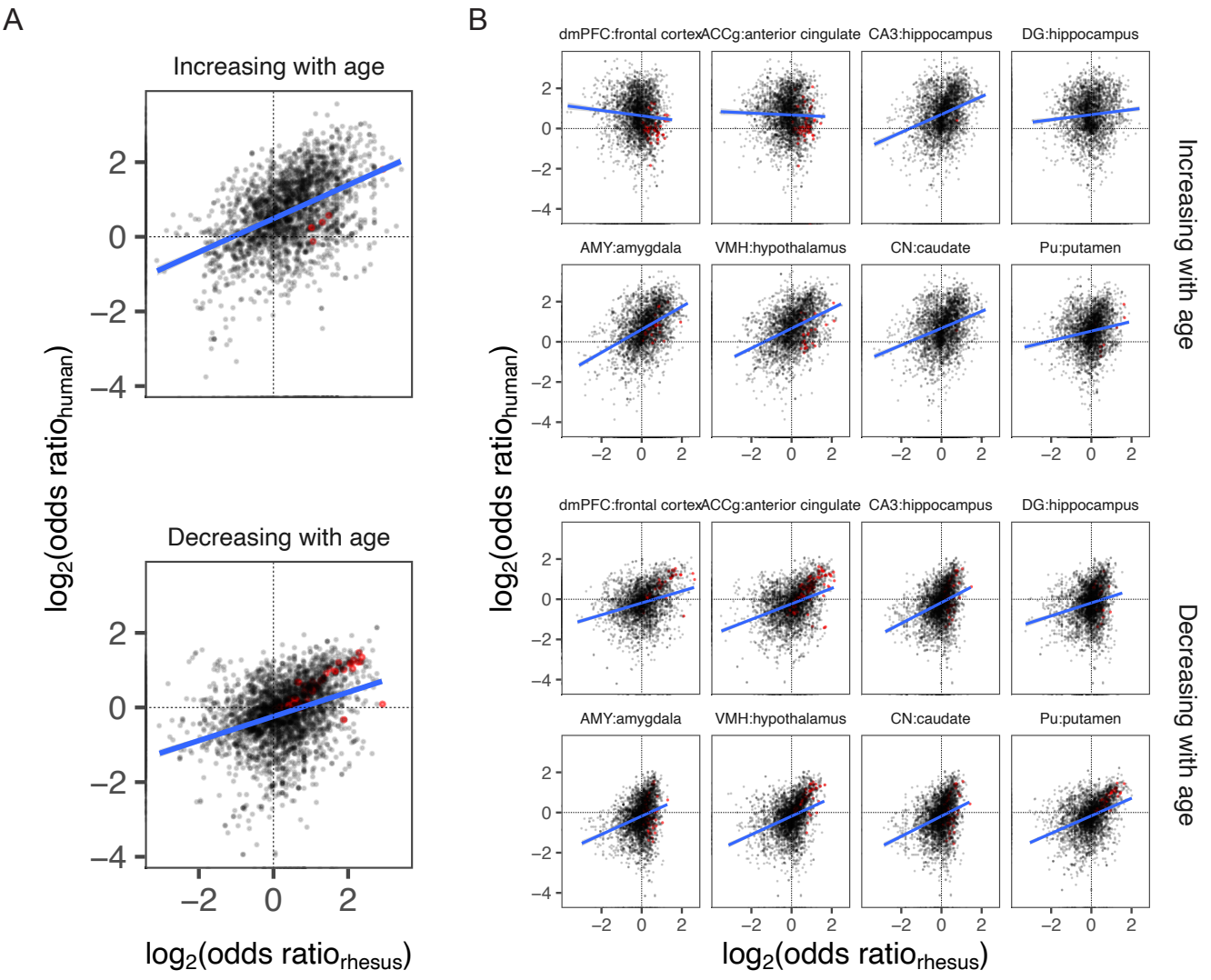
Supplementary Figure 3. Variance partitioning of brain gene expression. (A) Variance partitioning on global gene expression across all brain regions (*N*=527 biologically independent samples) reveals that, of all tested variables, brain region explains the greatest proportion (median ~30%) of variation. Variables are sorted by the proportion of variance explained. Variance partitioning was performed without first removing technical variation. Variance partitioning results from gene expression data with technical variation removed are shown in Fig. 1C. (B) Variance partitioning on region-specific gene expression (*N*=36 biologically independent animals)—with technical covariates regressed out—reveals that age explains the greatest proportion of variance of all tested covariates within single brain regions, though still a relatively small proportion (median 0.6–6.0%) of overall variation. Box plots depict the median (center), and interquartile range (IQR, bounds of box), with whiskers extending to either the maxima/minima or to the median $\pm 1.5 \times$IQR, whichever is nearest.



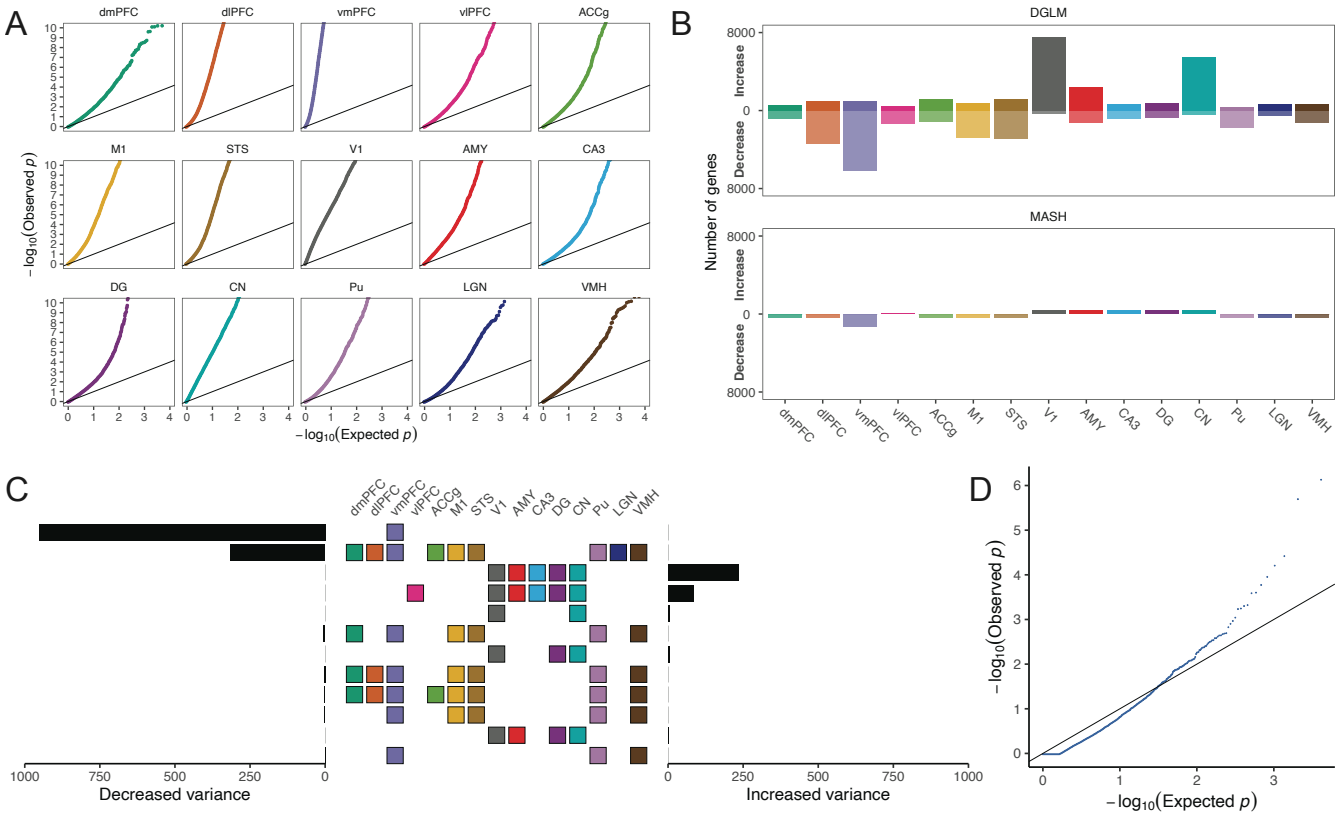
Supplementary Figure 4. Age effects on principal axes of variation. Principal components analyses (PCAs) were conducted on each brain tissue dataset separately. The loading on each principal component (PC) was regressed against age, with sex and dominance rank included as covariates. The -$\log_10$ *P* value from the age predictor (linear model, two-sided test, no correction for multiple comparisons) is plotted against each PC, with the size of each point proportional to the percentage variance explained. Only the first *N* principal components together explaining >99% of variance are included for each brain region.



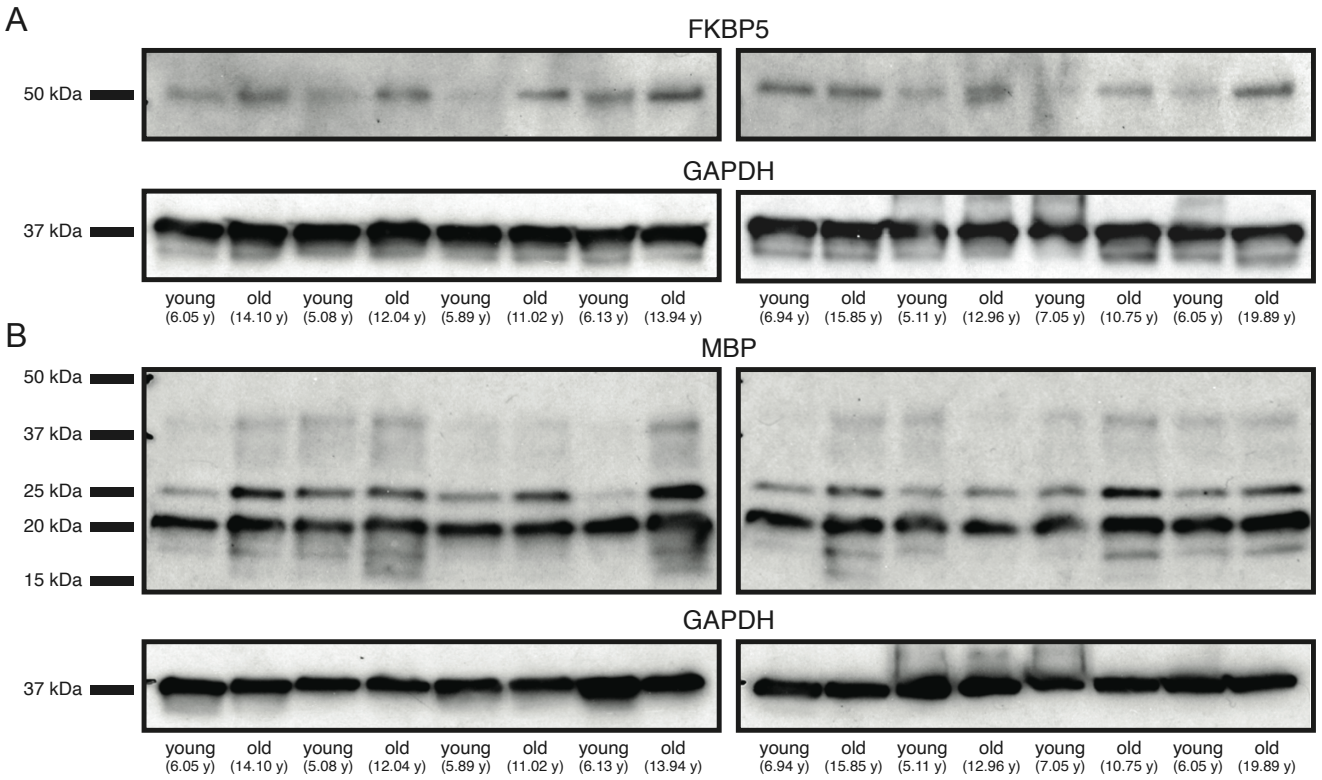
Supplementary Figure 5. Strong biological signals in whole-brain age-differentially expressed genes. (A) Quantile-quantile (QQ) plots from Gene Ontology (GO) hypergeometric enrichment tests, conducted for wbaDEGs as well as separately for each brain region, show enrichment of low P values, particularly for genes decreasing in expression with age. (B) QQ plots from hypergeometric enrichment tests conducted on disease associations from the DISEASES database and (C) QQ plot from transcription factor motif enrichment results, conducted on wbaDEGs using the HOMER software, similarly show strong enrichment signals. (D) Enriched GO biological processes and diseases among wbaDEGs (Supplementary Table 6). Odds ratios (ORs) are uncorrected for the GO graph structure, with the red dotted line representing a neutral OR = 1. All *P* values are based on one-sided Fisher's exact test with no correction for multiple comparisons.



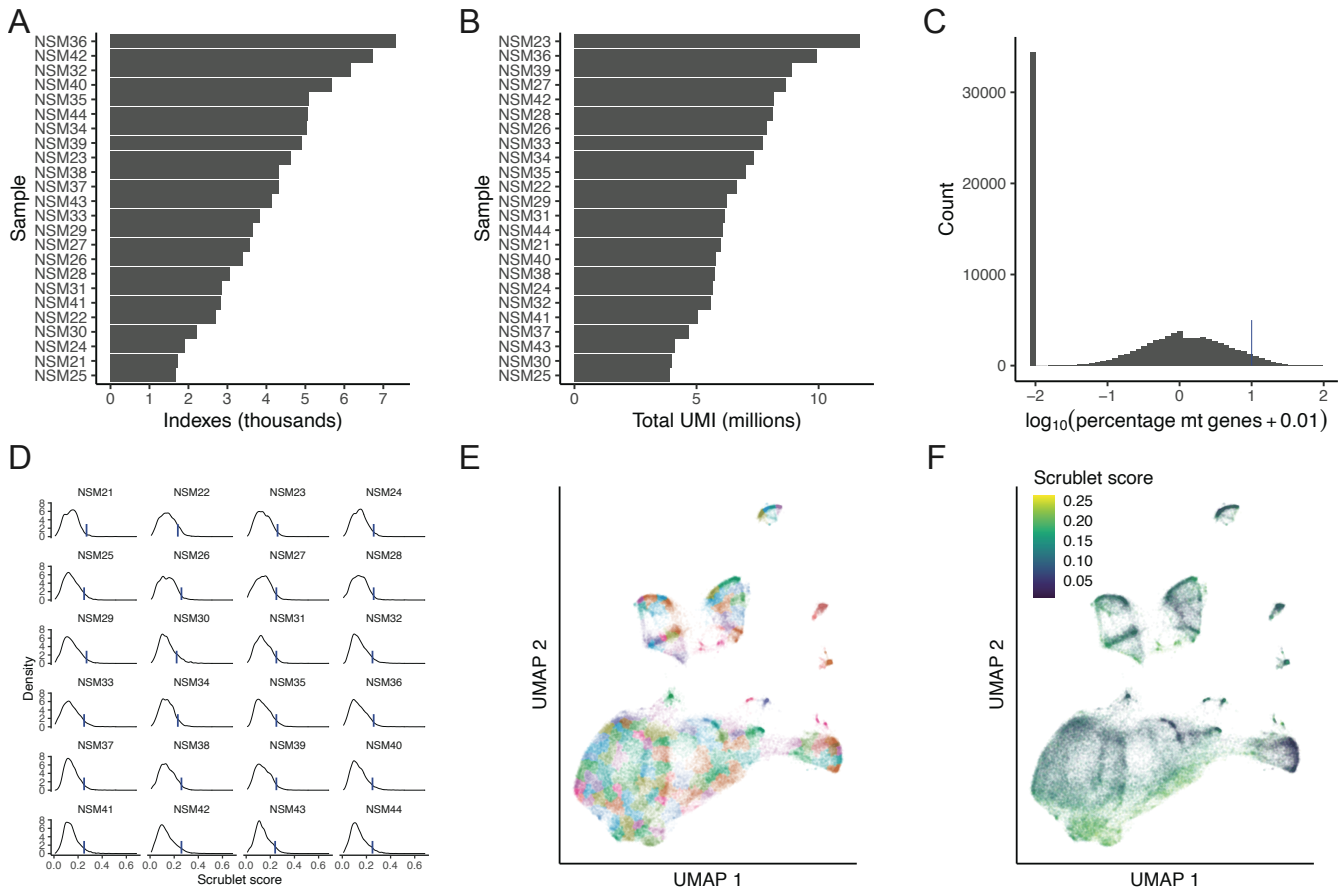
Supplementary Figure 6. Correlation between age enrichment effects between macaques and humans. Scatterplots depict positive relationships between age effects estimated from macaques from this study (x axis) and humans from the GTEx study (y axis) in overlapping brain regions. wbaDEGs, which show the strongest support for shared effects across macaque brain regions, are highlighted in red.



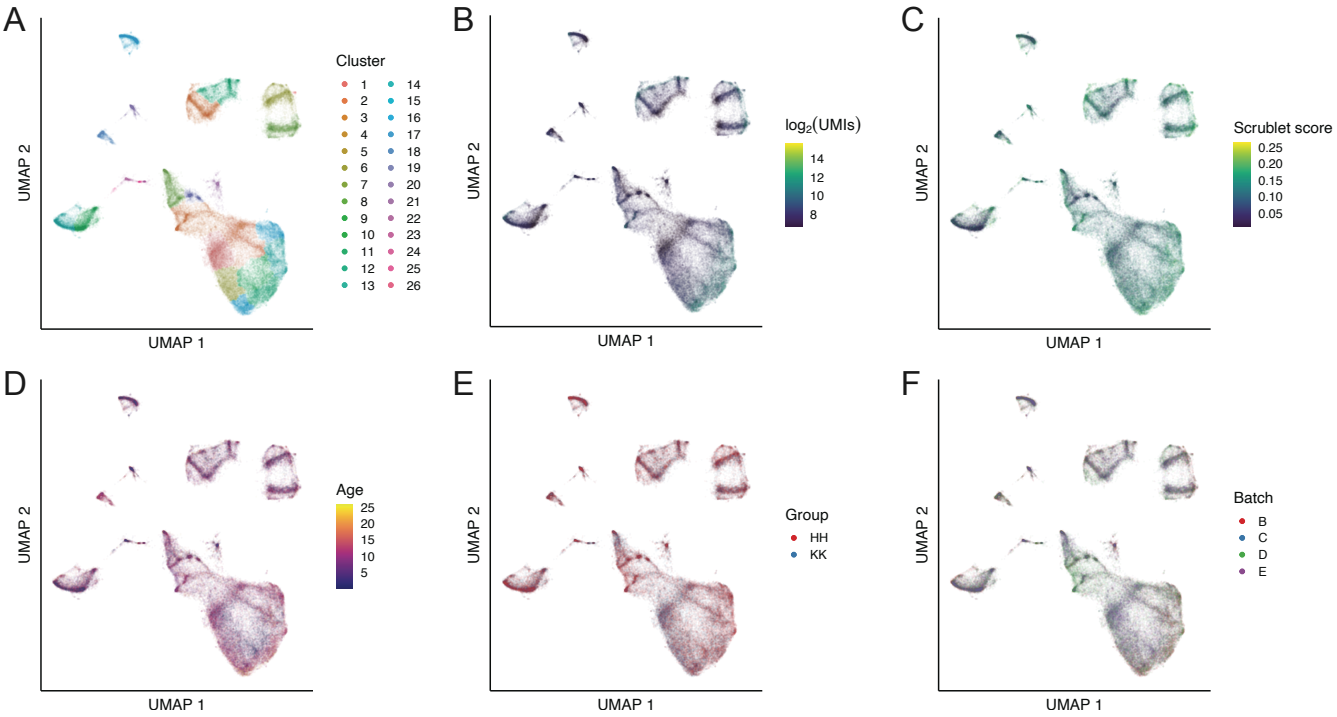
Supplementary Figure 7. Aging impacts the variability of gene expression. (A) QQ-plots of initial double generalized linear model (DGLM) results show enrichment of low *P* values (two-sided test, no correction for multiple comparisons). (B) A multivariate adaptive shrinkage (MASH) approach finds substantially less sharing between tissues compared to our analysis of mean effects, resulting in a more conservative set of significant genes. (C) Upset plot showing the rank order of brain regions with the greatest number of genes exhibiting significant age-associated changes in variance. Most genes with age-associated changes in variance were found across a group of regions including V1, AMY, CA3, DG, and CN. (D) QQ plot of GO hypergeometric enrichment test results shows a strong biological signal among genes increasing in dispersion with age (one-sided Fisher's exact test, no correction for multiple comparisons).



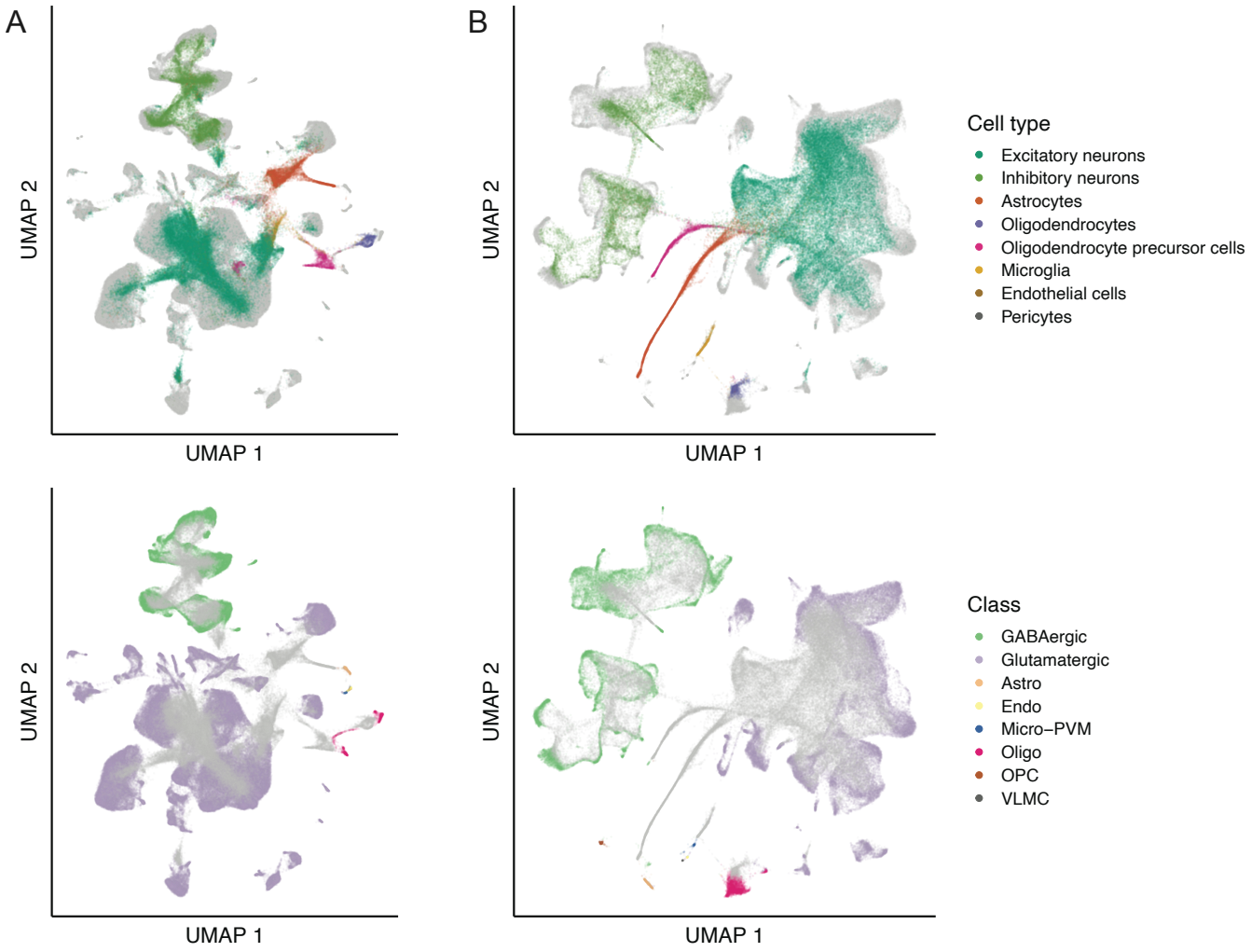
Supplementary Figure 8. Western blotting results. Representative western blot images for analysis of (A) FKBP5 and (B) MBP protein levels in macaque dlPFC. Experimenters were blinded to sample identities, which were alternated young (<10 years old) and old (>10 years old) across two gels. Both gels were run, transferred, blotted, exposed, and analyzed in parallel. Protein levels of target proteins were normalized to expression values of GAPDH. Relative quantities were measured and are reported in Supplementary Table 22. For FKBP5 analysis, two experiments were run, with similar results. For data analysis, normalized values were averaged across technical replicates for each animal. These average values are presented in Supplementary Table 22. For MBP analysis, a single experiment was run.



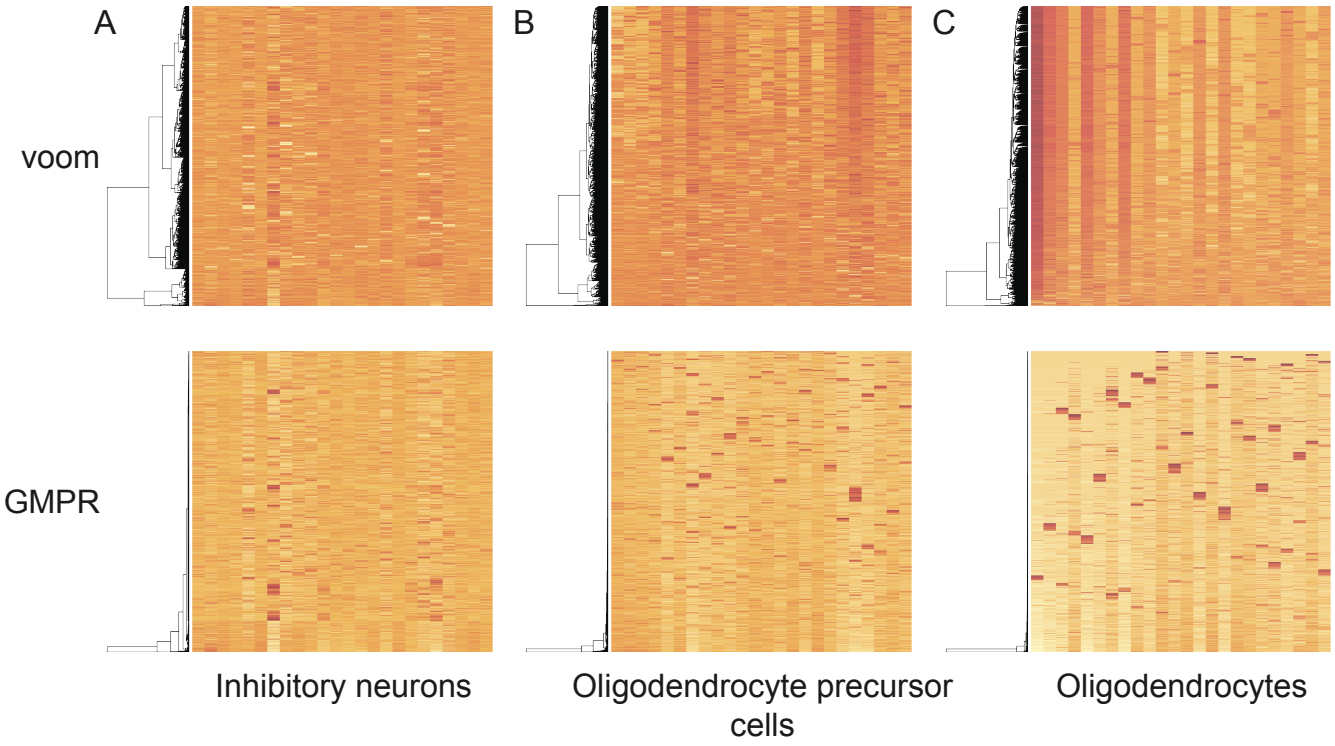
Supplementary Figure 9. Overview of single-nucleus RNA sequencing data quality control pipeline. (A) Unique indices (i.e., sci-RNA-seq3 barcode combinations, which primarily represent single nuclei with a small fraction of multiplets) per sample in an initial dataset of 96,906 total indices with > 100 unique molecular indices (UMIs). (B) Total UMIs per sample in the same initial dataset. (C) Nuclei with greater than 10% expression in mitochondrial genes were excluded. 93,143 total indices passed this filter. (D) The distribution of Scrublet *k*-nearest-neighbor (kNN) scores were visualized per sample and thresholds were manually set per sample to exclude indices with higher kNN scores. 88,491 indices passed this filter. (E) UMAP plot of 88,491 indices, colored by cluster. (F) Scrublet kNN scores visualized on UMAP plot. Clusters with high median kNN scores were manually removed. 71,863 total indices passed this filter (Supplementary Fig. 10).



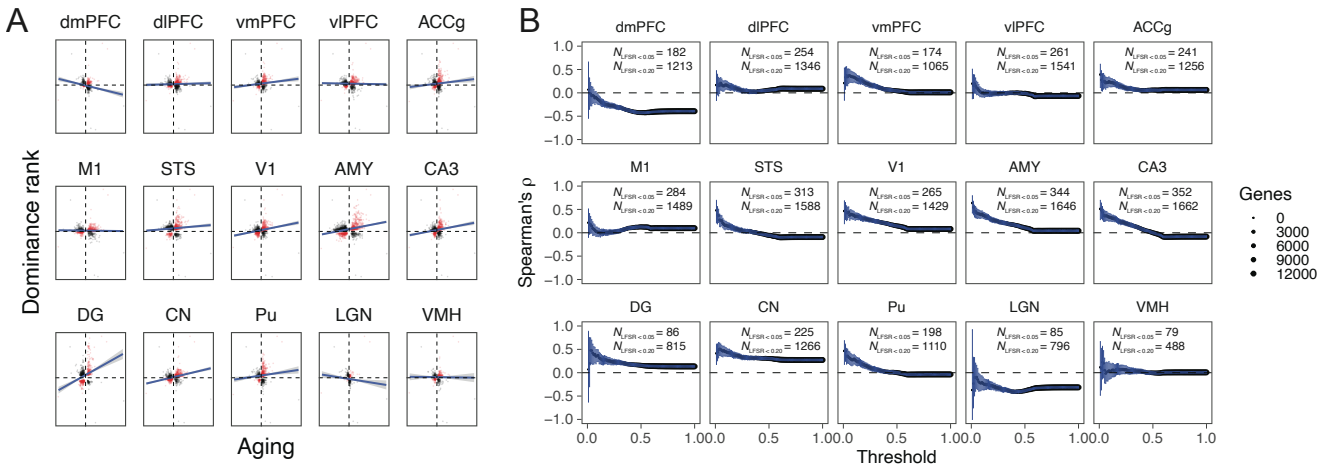
Supplementary Figure 10. Distribution of technical and biological variation in single-nucleus RNA sequencing data. (A) Unsupervised clusters, (B) UMI counts, (C) Scrublet kNN scores, (D) age, (E) social groups, and (F) nuclei isolation batches visualized in UMAP dimensions.



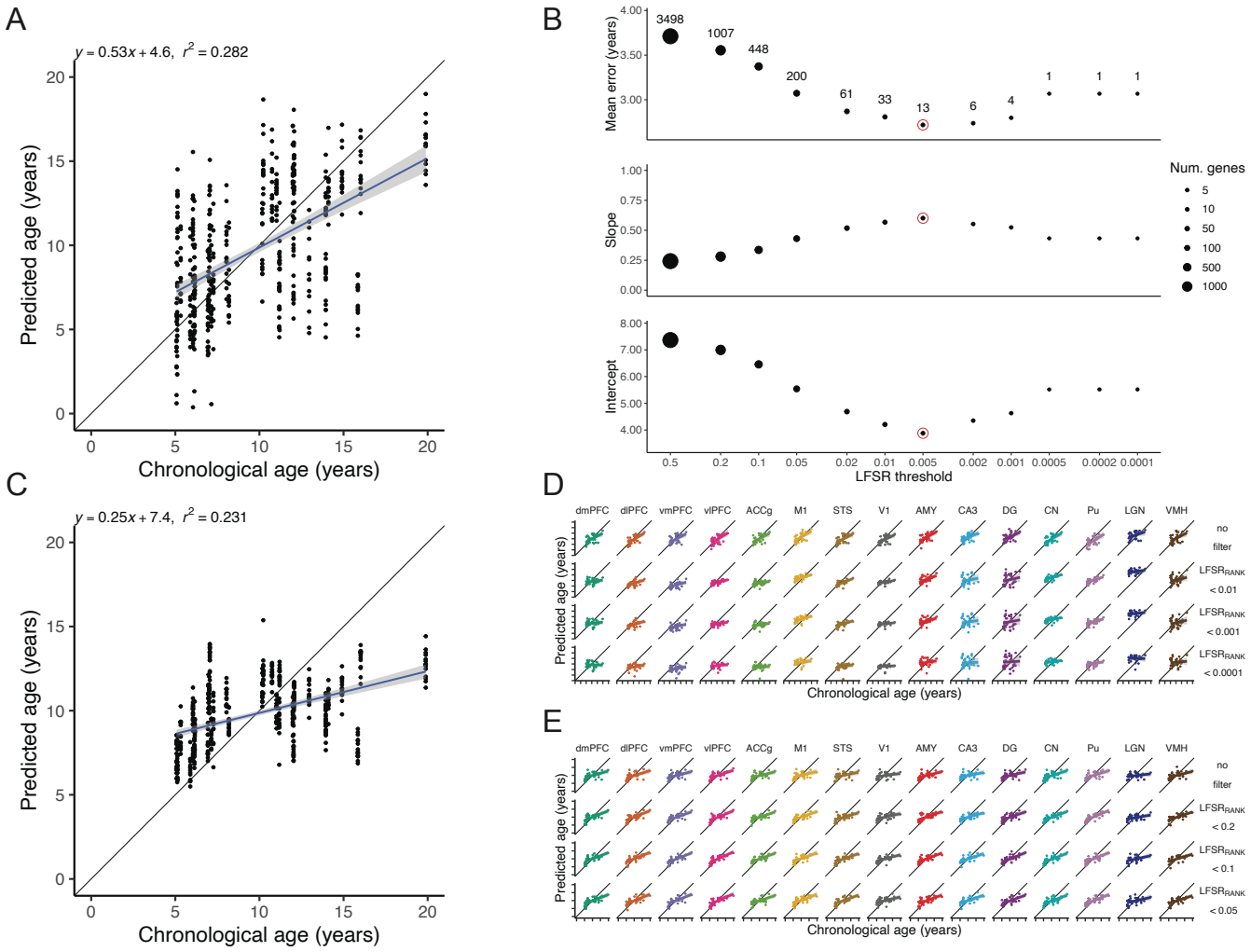
Supplementary Figure 11. Integration and label transfer on single-nucleus RNA sequencing data corroborate cell-type assignments. UMAP plots of 71,863 rhesus macaque single-nucleus dlPFC cells and reference Allen Brain Map cells are visualized on integrated data spaces. Rhesus macaque data are shown with (A) 200,000 mouse cerebral cortex+hippocampus cells; and (B) 76,533 human M1 cells. For both datasets, cells are colored by (top) rhesus macaque manual cell annotations (mouse cells shown in light gray) and (bottom) Allen Brain Map cell classes (rhesus cells shown in light gray). For non-neuronal cell types in the Allen reference datasets, the cell subclass is highlighted instead.



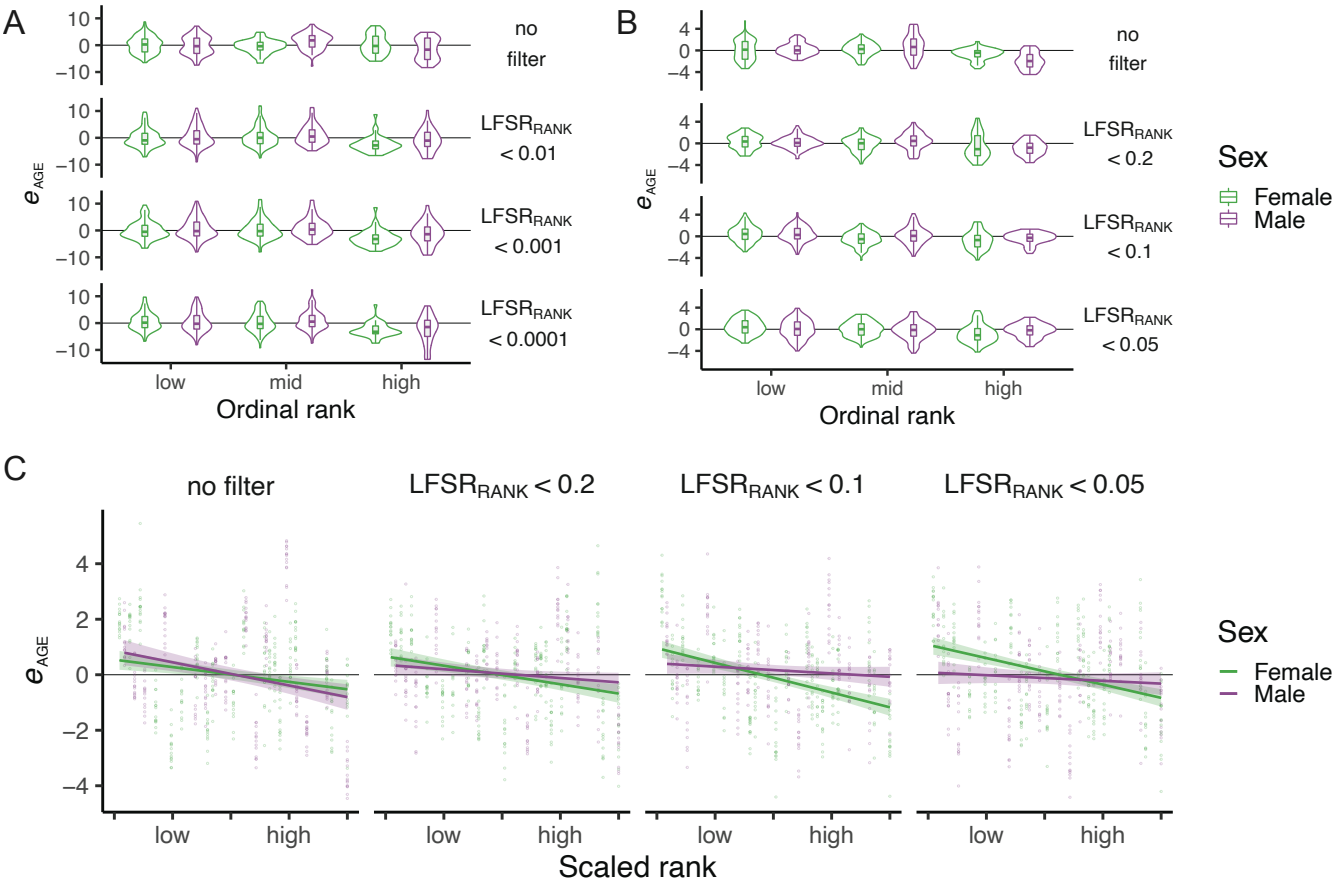
Supplementary Figure 12. Comparison of RNA count normalization strategies on pseudobulk data. Heatmaps compare (top) a standard RNA-seq normalization procedure (voom from the limma R package) with (bottom) a normalization procedure designed for zero-inflated count data (GMPR). The two methods produce qualitatively similar results on (A) a cell type (inhibitory neurons) exhibiting no age-associated differences in proportion (linear model, two-sided test, uncorrected $P = 0.900$). GMPR, however, eliminates age-associated artifacts on global expression in (B) a cell type (oligodendrocyte precursor cells) exhibiting moderate age-associated decreases in proportion (linear model, two-sided test, uncorrected $P = 0.029$, Bonferroni-adjusted $P = 0.234$) and (C) a cell type (oligodendrocytes) exhibiting significant age-associated increases in proportion (linear model, two-sided test, uncorrected $P = 0.005$, Bonferroni-adjusted $P = 0.043$), likely due to the correction for a higher prevalence of zeroes in libraries aggregated from fewer nuclei. For all plots, libraries on the x axis are arranged from youngest (left) to oldest (right).



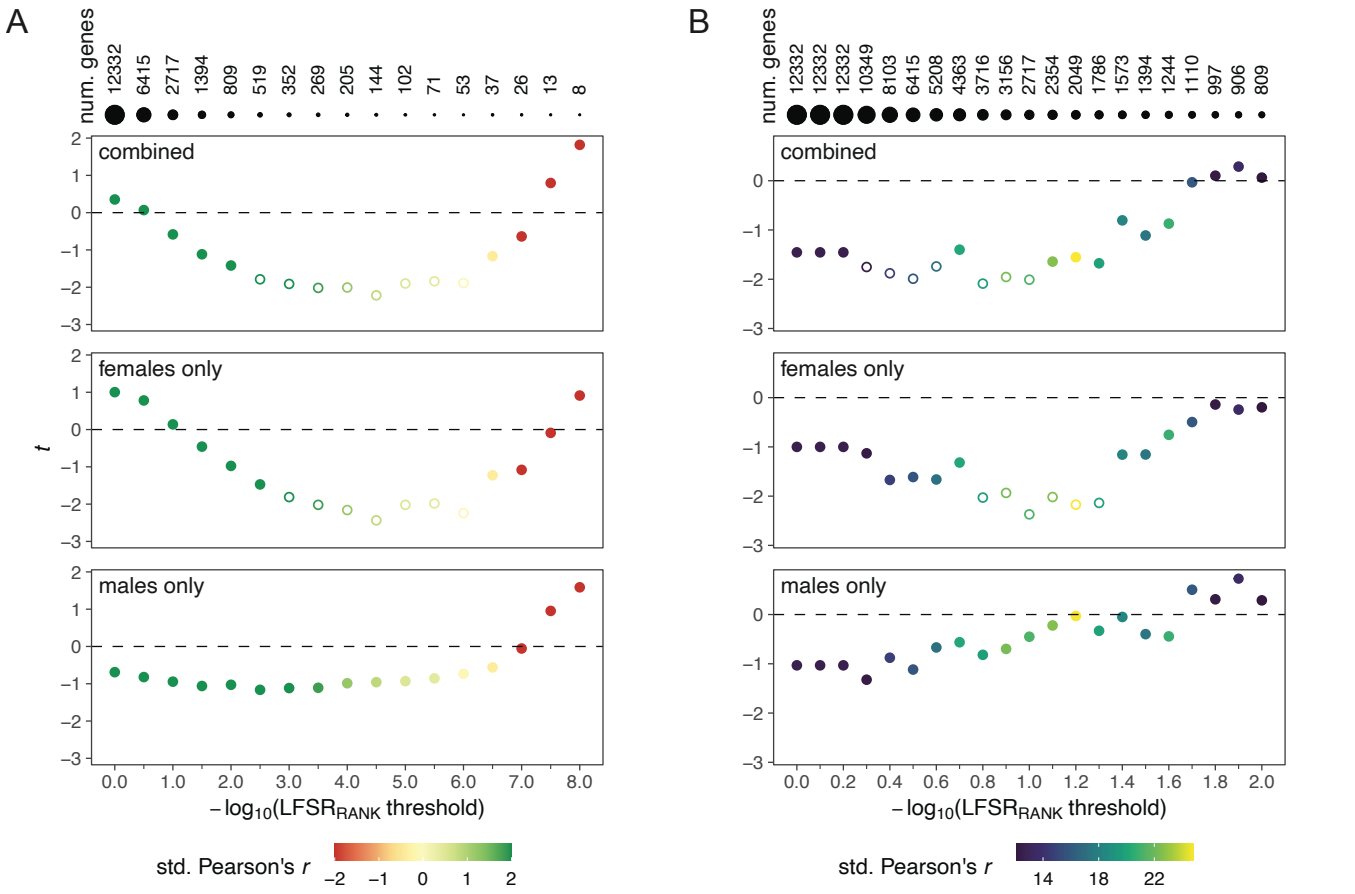
Supplementary Figure 13. Parallel signatures of aging and dominance rank on gene expression. Aging (x axis) and dominance rank (y axis) are positively correlated across most brain regions. For these plots, only genes with LFSR < 0.2 for both age and rank effects in a given brain region are included and effect size signs for dominance rank were reversed such that positive values indicate higher expression in lower ranking individuals. Red points indicate that genes share the sign of their effect sizes for both variables. Error bands represent the 95% confidence interval of linear model predictions. (B) As depicted in Fig. 6A, correlations between age and dominance rank effects robustly increase at more stringent statistical thresholds. 95% confidence intervals estimated through 1,000 bootstrap replicates are included. Dot sizes indicate the number of genes passing a given threshold for both age and dominance rank. In addition, the number of genes passing thresholds of LFSR < 0.05 and LFSR < 0.2 are annotated..



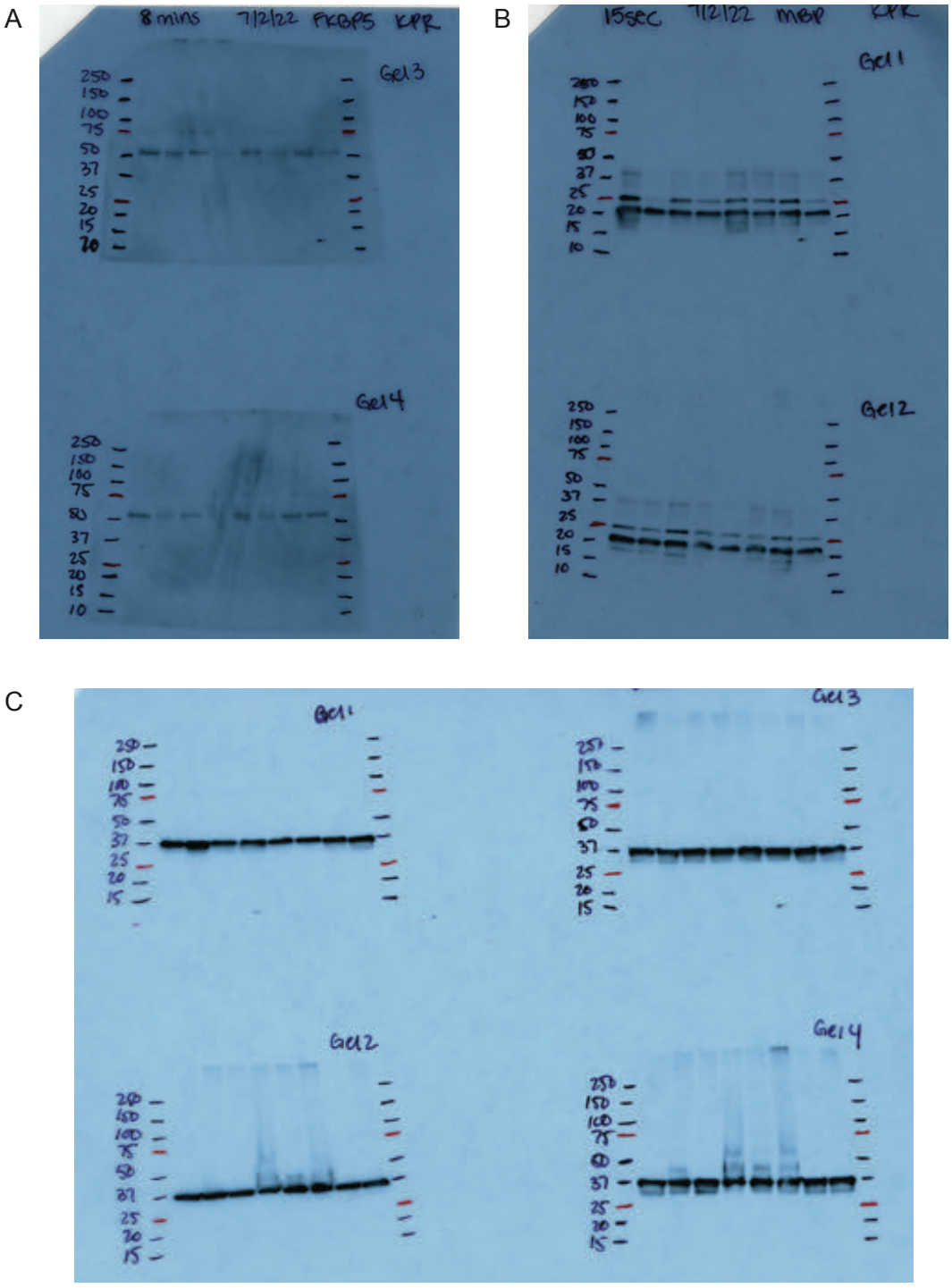
Supplementary Figure 14. Age predictions from gene expression robustly exhibit positive correlations with chronological age across brain regions. (A) Comparison of known chronological ages and predicted transcriptomic ages on all samples combined across regions. All predicted ages were estimated using leave-one-out cross validation (LOOCV) in which model coefficients were re-estimated following EMMA and MASH models omitting the test samples from an individual. Error bands represent the 95% confidence interval of linear model predictions. (B) Gene selection from wbaDEGs identified the optimal threshold (LFSR < 0.005) and number of genes (N = 13) that together minimized the mean error (mean absolute deviation), and produced a slope and intercept closest to 1 and 0, respectively. (C) Age predictions from elastic net regression models trained with glmnet showed a similarly strong overall correlation but with a more compressed range of predicted ages. Predicted ages were also estimated using an LOOCV procedure in which elastic net models were rerun after leaving out all samples from a test individual. Error bands represent the 95% confidence interval of linear model predictions. (D) wbaDEGs model predictions and (E) glmnet model predictions robustly maintain strong positive correlations and accuracy after filtering to genes showing increasingly strong support for dominance rank effects. Error bands represent the 95% confidence interval of linear model predictions.



Supplementary Figure 15. Dominance rank alters aging among a subset of age-associated genes. Comparison of dominance rank classes shows that high ranking females in particular exhibit age deceleration relative to other females and to males (*N*=527 biologically independent samples) based on both (A) wbaDEGs model predictions and (B) glmnet model predictions. The relationship is stronger at more stringent $LFSR_{RANK}$ thresholds. Box plots depict the median (center), and interquartile range (IQR, bounds of box), with whiskers extending to either the maxima/minima or to the median $\pm 1.5 \times$IQR, whichever is nearest. (C) Linear regression using percentage dominance ranks shows a similarly strengthening relationship between dominance rank and e_{AGE} at increasingly stringent $LFSR_{RANK}$ thresholds, indicating that dominance rank-based variation in aging is strongest in genes that underlie shared signatures of aging and dominance rank. Error bands represent the 95% confidence interval of linear model predictions.



Supplementary Figure 16. Evaluation of dominance rank/age relationship across a broad range of thresholds. For both wbaDEGs (A) and glmnet (B) models, iterating across a greater range of $LFSR_{RANK}$ thresholds reveals that statistical support for dominance rank/predicted age relationship generally improves while predictions remain positively correlated with chronological age. For both models, however, the improvements eventually reverse when very few genes remain. Open circles indicate that the relationship between dominance rank and predicted age was significant (linear model, two-sided test, uncorrected $P < 0.05$).



Supplementary Figure 17. Original scans of Western blot gels. Original scans of (A) FKBP5, (B) MBP, and (C) GAPDH Western blot experiments.