PTSD, Major Depression, and Advanced Transcriptomic Age in Brain Tissue

Supplementary Materials

Xiang Zhao, M.S.[1,2], Mark W. Logue, Ph.D.[1,2,3,4], Sage E. Hawn, Ph.D.[1,2], Zoe E. Neale, Ph.D.[1,2], Zhenwei Zhou, M.S.[4], Bertrand R. Huber, Ph.D.[1,5,6], Traumatic Stress Brain Research Group, Mark W. Miller, Ph.D.[1,2] & Erika J. Wolf, Ph.D.[1,2]

[1]National Center for PTSD at VA Boston Healthcare System, Boston, MA, USA

[2] Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA

[3]Biomedical Genetics, Boston University School of Medicine, Boston, MA, USA

[4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

[5]Pathology and Laboratory Medicine, VA Boston Healthcare System, Boston, MA, USA

[6]Department of Neurology, Boston University School of Medicine, Boston, MA, USA

Supplementary Methods

Genotyping and DNA Methylation Data

DNA was extracted and isolated from all three brain regions using Qiagen Blood & Cell Culture kits. DNA quantification was proceeded with PicoGreen dsDNA fluorescent assays (Invitrogen). DNA quality and quantity were assessed by TaqMan® RNase P Detection assay (Applied Biosystems Assay, Life Technologies, Carlsbad, CA) with fluorescence detection on a 7900 Fast Real Time PCR System (Applied Biosystems, Life Technologies, Carlsbad, CA).

Genotypes were assessed from the motor cortex samples using Illumina HumanOmni2.5-8 BeadChips by the Pharmacogenomics Analysis Laboratory at the Central Arkansas Veterans Healthcare System (PAL). DNA was whole-genome amplified, fragmented, precipitated, resuspended, hybridized to the BeadChips, stained, and imaged by Illumina iScan System. Illumina. GenomeStudio v2011.1 software (Genotyping v1.9.4 module) was used for processing the results. The resulting data were cleaned using PLINK [1] by filtering missing data, checking sex mismatch between reported sex and X chromosome homozygosity, and screening for cryptic relatedness across all samples to detect potentially related or duplicated samples and suspicious samples swaps. Genotypes were imputed by Ricopili [2]. Covariates for ancestry were computed using Principal Components Analysis (PCA) of 100,000 randomly chosen common (minor allele frequency>0.05) single nucleotide polymorphisms (SNPs).

DNA methylation (DNAm) was assessed using Illumina Infinium EPIC BeadChips (EPIC). DNA from each of the three brain regions was bisulfite-modified, whole-genome amplified and hybridized to EPIC arrays, single-based extended, and stained by the Automated Protocol for the Illumina Infinium HD Methylation Assay. Chip positions were balanced based on brain regions and PTSD diagnosis. GenomeStudio projects were generated for each batch. Following the Psychiatric Genomics Consortium (PGC)-EWAS quality control pipeline [3], DNAm data was cleaned using the CpGassoc [4] and ChAMP [5,6] packages in R by the following steps. First, CpG sites that failed to achieve a detection p-value of 0.001 were set to missing, and probes that can cross hybridize to sex chromosomes or have more than

10% missingness were dropped. Then, samples that had more than 5% missing data or didn't meet the probe intensity threshold (> 50% of the experiment-wide mean or with intensity >2,000 arbitrary units) were excluded. Next, the beta mixture quantile dilation (BMIQ) method was used for DNAm data normalization via the wateRmelon R package [7]. After that, missing data were imputed using a k-nearest neighbor method by the Bioconductor impute package [8]. Finally, batch effects were removed using an empirical-Bayes batch-correction method (ComBat) by the Bioconductor package sva [9]. Duplicated samples were screened, and one of the duplicates with the lowest missing rate was retained. Horvath DNAm age [10] was calculated on the raw beta values using the R script supplied by Dr. Horvath. The script automatically performs normalization and imputation and outputs the DNAm age estimates based on methylation levels at 353 450K probes [10]. Choi DNAm age was generated using the normalized and imputed DNAm data by computing the product of the corresponding effect size estimates times the beta values from 230 brain-specific age-associated probes [11]. Shireby DNAm age was assessed using the R script provided by Dr. Shireby at https://github.com/gemmashireby/CorticalClock. Normalized with imputed beta values were used as the input; the cortical DNAm age was computed based on methylation levels at 347 probes [12].

RNA Sequencing Methods and Gene Expression Data

RNA was extracted from 25mg of tissue from each of the three brain regions using Qiagen RNeasy Fibrous Tissue Minikit. Library preparation was conducted using the Illumina TruSeq Stranded total RNA kit with globin depletion. The libraries were sequenced by a Hiseq 2500 which produced paired-end 75 base pair (bp) reads. To avoid empty lanes, the Hiseq was performed in both the "high output" mode (flow cells are run over eight lanes that contain unique library pools) and "rapid" mode (single cell run over two lanes). Trimmomatic [13] was used to eliminate adapters and remove short or low-quality reads followed by mapping the trimmed reads to the hg38 human reference genome [14] via STAR [15] using the two-pass mode. The quality of aligned reads was evaluated using FastQC [16], RseQC [17], and MultiQC [18], and samples with <50% uniquely mapped reads were excluded. Transcripts were quantified using Kallisto [19], and gene-level counts were assessed by collapsing the

Kallisto transcript abundance estimates via tximport [20]. Regularized log transformed (rlog) expression values were generated based on gene-level counts using DESeq2 [21]. Rlog expression values of X and Y chromosome genes were examined to confirm correspondence of the data to reported sex. To detect outliers, PCs were computed from the rlog values. Samples were considered as outliers and excluded if more than 6 SDs away from the group mean on any of the first 10 PCs. To evaluate RNA degradation, quality surrogate variables (qSVs) were computed using quality surrogate variable analysis (qSVA) [22] in the Bioconductor package sva [8], and the first three qSVs were used as covariates in downstream analyses. The relative balance of seven cell types (astrocytes, endothelial cells, microglia, mural cells, neurons, oligodendrocytes, and red blood cells) were estimated from rlog values in each brain region using BrainInABlender [23].

RNA Age Calculation

RNAAgeCalc has multiple versions of pre-trained calculators based on training samples from different ancestral groups (all races or Caucasians) and tissues, and choices in the candidate gene set used to compute RNA age [24]. We used the universal (calculator trained on samples from all race groups) brain-specific calculator, and the DESeq2 candidate gene set (n = 472 age-associated genes in the brain across all races identified by DESeq2 using GTEx data [24]) to calculate RNA age in our sample. We inputted raw read counts of gene expression from dlPFC, vmPFC, and motor cortex to the RNAAgeCalc *predict_age* function. 457 genes were covered in the Brain Bank data for all three brain regions and 15 genes that were not produced via sequencing were imputed by the RNAAgeCalc algorithm before calculating the RNA age.

Transcriptome-wide Analysis

Transcriptome-wide analyses of gene expression vs. psychiatric variables and cell types were only conducted in vmPFC, as stated in the main text. The analyses were preformed using DESeq2 package [21] on genes with more than 1 read count in at least 30 subjects. The number of genes that met this threshold was 34,205 in vmPFC, 448 of which were included in the RNAAgeCalc algorithm. For the analysis of the psychiatric disorder(s), the following covariates were included: sex, age at the time of

death, post-mortem interval (PMI), sequencing run IDs, first three qSVs, and seven cell types. We then extracted the results of the vmPFC genes in RNAAgeCalc from the transcriptome-wide analysis (n = 448) from the DESeq2 output and generated a corrected p-values across them using a False Discovery Rate (FDR) correction [25]. For cell type analysis, sequencing run IDs and first three qSVs were covaried. Results of RNA age genes were extracted and multiple testing correction was conducted by adjusting p-values across the number of cell types and the number of RNA age genes (7 x 448) via FDR.

Cell Type Enrichment Analysis

We used a set of brain cell type marker genes (http://resource.psychencode.org/DER-21_Single_cell_markergenes_UMI.xlsx) from PsychENCODE [26] to test the enrichment of 25 cell type layers from 8 cell types (astrocytes, endothelial cells, microglia, oligodendrocytes, oligodendrocyte precursor cells, pericytes, excitatory neurons, and inhibitory neurons) in RNAAgeCalc genes. This was achieved by running a hypergeometric test to calculate the probability of a set of cell type marker genes being present in RNAAgeCalc algorithm. P-values were adjusted using the FDR correction for the 25 cell type layers.

CIBERSORTx Estimation and Cell Type Association Analysis

In addition to the BrainInABlender cell type analysis, we estimated another set of proportions of eight cell types (excitatory neurons, inhibitory neurons, astrocytes, endothelial cells, oligodendrocytes, microglia, pericytes, and oligodendrocyte progenitor cells) using CIBERSORTx [27]. CIBERSORTx cell type estimates are proportions ranging between 0 and 1. We transformed the cell types proportions using the logit function into cell type scores for statistical analysis. We used these estimates to test for replication of the associations between the BrainInABlender endothelial and mural cell scores with PTSD/MDD and RNA age residuals. While CIBERSORTx estimation doesn't yield mural cell estimates, it does yield estimates of pericytes, which are a subtype of mural cells located in blood microvessels. We also modeled the associations between CIBERSORTx cell scores and age at death and RNA age. These replication analyses were conducted through multiple regression controlling for sex, sequencing run IDs, and first three qSVs (results below).

Supplementary Results

Associations between Psychopathology and RNA Age Residuals: Sensitivity Analysis

The effect for PTSD/MDD was still significant in vmPFC when additionally controlling for body mass index (BMI) ($p_{PTSD/MDD}$ = .006, p_{BMI} =.64), PMI ($p_{PTSD/MDD}$ = .004, p_{PMI} =.62), the top three ancestry PCs ($p_{PTSD/MDD}$ = .004, smallest p_{PC} = .12), seven estimated cell types ($p_{PTSD/MDD}$ = .018, smallest $p_{cell\text{-}type}$ = .022, which was for oligodendrocytes), manner of death ($p_{PTSD/MDD}$ = .018, smallest $p_{manner\text{-}of\text{-}death}$ = .27), anti-depressant use at time-of-death ($p_{PTSD/MDD}$ = .003, $p_{anti\text{-}depressants}$ = .334), and total trauma exposure ($p_{PTSD/MDD}$ = .014, p_{trauma} = .29). The standardized coefficients (βs) for the PTSD/MDD effects in these follow-up models with additional covariates included differed on average by an absolute value of .02 (absolute value Δ β range: 0 - .06) from the β = .39 effect reported for the PTSD/MDD term in the primary model (as listed in Table 2 in the main text).

RNA Age Residuals and Comorbid Psychiatric Disorders Group

We examined if the joint PTSD/MDD association with RNA Age residuals in vmPFC was driven by PTSD, MDD, or their combination. To do so, we first tested a model that included variables reflecting PTSD without MDD, MDD without PTSD, and PTSD + MDD, controlling for AUD, smoking, three qSVs, and sex (consistent with the primary analysis). Of these variables, we found significant effects for PTSD + MDD (β = .50, p = .003) and MDD without PTSD (β = .41, p = .016). As reported elsewhere in the manuscript, only n = 4 individuals met criteria for PTSD without MDD, making it difficult to observe effects specific to PTSD in this cohort and underscoring the preliminary nature of these supplementary analyses.

We examined these models with additional covariates included, analogous to the sensitivity analyses described above. We found that PTSD + MDD and MDD without PTSD remained significantly associated with RNA age residuals in vmPFC when additionally adjusting for BMI ($p_{PTSD+MDD}$ = .005, $p_{MDDnoPTSD}$ = .019, p_{BMI} = .715), PMI ($p_{PTSD+MDD}$ = .004, $p_{MDDnoPTSD}$ = .015, p_{PMI} = .747), the top three ancestry PCs ($p_{PTSD+MDD}$ = .004, $p_{MDDnoPTSD}$ = .012, smallest p_{PC} = .184), seven estimated cell types ($p_{PTSD+MDD}$ = .012, $p_{MDDnoPTSD}$ = .040, smallest $p_{celltype\text{-}oligodendrocyte}$ = .021), anti-depressant use at time-of-

death ($p_{PTSD+MDD}$ = .002, $p_{MDDnoPTSD}$ = .009, $p_{anti\text{-}depressant}$ = .274), and manner of death ($p_{PTSD+MDD}$ = .013, $p_{MDDnoPTSD}$ = .029, smallest $p_{manner\text{-}suicide}$ = .239). The only covariate which impacted the significance of either of these variables was the inclusion of total trauma exposure in the model, which resulted in a model with a non-significant association for the PTSD + MDD group (p = .125), though a significant one for the MDD no PTSD group (p = .015), with no significant effect for trauma count (p = .237). Further examination of the relationship between the PTSD + MDD variable and total trauma count suggested that these variables were multicollinear (r = .72, p < .001), meaning that it was not possible to differentiate the effects of one variable from the other due to their overlapping variance. Similar to the primary model, the standardized coefficients (βs) for the PTSD + MDD effects in the follow-up models with additional covariates included differed on average by an absolute value of .05 from the β = .51 reported above and the MDD without PTSD standardized βs differed by a mean absolute value of .02 from the βs = .41 reported above.

We also evaluated if the number of psychiatric diagnoses (i.e., ranging from 0 -2) was associated with RNA age residuals in vmPFC. This analysis revealed that RNA age residuals increased with each additional psychiatric diagnosis (β = .44, p = .006), controlling for sex, cigarette use, and three qSVs. We evaluated this model with the additional covariates included and found that the count of psychiatric diagnoses continued to be significantly associated with RNA age residuals in each model, including when additionally adjusting for BMI ($p_{\#dx}$ = .009, p_{BMI} = .694), PMI ($p_{\#dx}$ = .007, p_{PMI} = .843), the top three ancestry PCs ($p_{\#dx}$ = .010, smallest p_{PC} = .181), seven estimated cell types ($p_{\#dx}$ = .019, smallest $p_{celltype\text{-}oligodendrocyte}$ = .023), anti-depressant use at time of death ($p_{\#dx}$ = .005, $p_{anti\text{-}depressant}$ = .351), manner of death ($p_{\#dx}$ = .030, smallest $p_{manner\text{-}suicide}$ = .317), and total trauma count ($p_{\#dx}$ = .038, p_{trauma} = .804). The mean difference in the absolute value of the βs for the count of psychiatric diagnoses effect with these covariates included in the model was .02 compared to the magnitude of this effect for the model without these covariates.

Associations with Sample Cell Type Composition in vmPFC

The BrainInABlender cell type association analyses revealed significant associations between: (1) endothelial cell scores and PTSD/MDD (B=-0.294, p = 0.004, $p\text{-}adj$ = 0.017; Table S4); (2) endothelial cell scores and RNA age residuals (B = -0.022, p = 0.008, $p\text{-}adj$ = 0.028; Table 4); (3) mural cell scores and PTSD/MDD (B = -0.234, p = 0.005, $p\text{-}adj$ = 0.017; Table S4); and (4) mural cell scores and RNA age residuals (B = -0.019, p = 0.005, $p\text{-}adj$ = 0.028; Table 4). Follow-up RNA age residual and cell type score associations in dlPFC and motor cortex are shown in Table S9. There was a nominally significant association between oligodendrocyte scores and raw (not residual) RNA age in vmPFC (B = 0.027, p = 0.013, $p\text{-}adj$ = 0.092; Table S6), but no significant associations between raw RNA age estimates and endothelial or mural cell scores. There were no BrainInABlender vmPFC cell score associations with age-at-death (Table S7). We then tested if the significant associations in vmPFC replicated when examining cell type scores estimated using the CIBERSORTx approach. PTSD/MDD was associated with decreased endothelial cell scores in vmPFC (B=-1.674, p = 0.018) using this second estimation approach, but was not associated with CIBERSORTx pericytes scores (p = 0.925). RNA age residuals just missed the threshold for statistical significance in association with CIBERSORTx endothelial scores (B = -0.111, p = 0.054), but evidenced no association with pericytes scores (p = 0.250). Both RNA age and age at death were associated with CIBERSORTx vmPFC oligodendrocyte scores ($B_{\text{RNA age}}$ = 0.171, $p_{RNA\ Age}$ = 0.001; $B_{\text{Age at death}}$ = 0.063, $p_{Age\ at\ death}$ = 0.048).

Supplementary Tables

Table S1

Full result of PTSD/MDD transcriptome-wide analysis in vmPFC – see separately uploaded excel file (SuppTableS1.csv)

Table S1 *note*: baseMean = mean of normalized counts of all samples; lfcse = the standard error of the log2 fold change; stat = Wald test statistics; *p*adj = adjusted *p*-values by FDR across the number of RNAAgeCalc genes included in the transcriptome-wide analysis (*n* = 448); RNAAge_EffectSize = weights of RNA age genes contributing to the RNA age.

Table S2

Corrected Significant GO Terms Enrichment for Genes in the RNAAgeCalc Algorithm Associated with PTSD/MDD in vmPFC

Category	Ontology	Term	numDEInCat	numInCat	*p*	*p*-adj	Gene
GO:0008285	BP	negative regulation of cell proliferation	10	20	6.844E-05	0.045	CEBPB;PKN1;DLL1;VSIG4;ATOH8;HSPA1A;GAL;CDKN2;TGFB1I1;RFPL1
GO:0045321	BP	leukocyte activation	12	32	0.0002	0.045	HYAL2;CEBPB;PKN1;DLL1;UNG;VSIG4;CCND3;HSPA1A;DPP7;MT1G;GAL;RPS3
GO:0002520	BP	immune system development	8	17	0.0003	0.045	HYAL2;CEBPB;PKN1;DLL1;UNG;HSPA1A;MT1G;MKNK2
GO:0032879	BP	regulation of localization	15	50	0.0004	0.045	KCNJ16;HYAL2;CEBPB;PKN1;DLL1;ATOH8;IL17RC;FXYD1;HSPA1A;TBC1D1;THADA;GAL;P2RY2;STX10;PIM3
GO:0008134	MF	transcription factor binding	7	14	0.0004	0.045	CEBPB;PKN1;DLL1;ATOH8;RPS3;CDKN2A;TGFB1I1
GO:0008283	BP	cell proliferation	13	42	0.001	0.045	CEBPB;PKN1;DLL1;VSIG4;CCND3;ATOH8;HSPA1A;GAL;PLCD1;RPS3;CDKN2A;TGFB1I1;RFPL1
GO:0042127	BP	regulation of cell proliferation	12	36	0.001	0.045	CEBPB;PKN1;DLL1;VSIG4;ATOH8;HSPA1A;GAL;PLCD1;RPS3;CDKN2A;TGFB1I1;RFPL1
GO:0002376	BP	immune system process	16	62	0.001	0.045	HYAL2;CEBPB;PKN1;DLL1;UNG;VSIG4;IFITM3;CCND3;APOBEC3C;IL17RC;HSPA1A;DPP7;MT1G;MKNK2;GAL;RPS3
GO:0080134	BP	regulation of response to stress	9	25	0.001	0.045	HYAL2;CEBPB;PKN1;VSIG4;IL17RC;HSPA1A;HIC1;RPS3;CDKN2A
GO:0006915	BP	apoptotic process	14	45	0.001	0.045	HYAL2;CEBPB;PKN1;DLL1;UNG;HSPA1A;ARHGEF3;MKNK2;GAL;HIC1;RPS3;CDKN2A;PIM3;RFPL1

GO:0001775	BP	cell activation	12	37	0.001	0.045	HYAL2;CEBPB;PKN1;DLL1;UNG;VSIG4;CCND3;HSPA1A;DPP7;MT1G;GAL;RPS3
GO:0012501	BP	programmed cell death	14	46	0.001	0.047	HYAL2;CEBPB;PKN1;DLL1;UNG;HSPA1A;ARHGEF3;MKNK2;GAL;HIC1;RPS3;CDKN2A;PIM3;RFPL1
GO:0048534	BP	hematopoietic or lymphoid organ development	7	16	0.001	0.047	HYAL2;CEBPB;PKN1;DLL1;HSPA1A;MT1G;MKNK2

Note. The GOSeq analysis tested for enriched GO terms in the nominally significant PTSD/MDD-associated genes in the RNAAgeCalc algorithm (n = 43) compared to the background gene list of all the RNAAgeCalc genes (n = 448). BP = biological process; MF = molecular function; CC = cellular component; numDEInCat = number of PTSD/MDD associated RNAAgeCalc genes included in the corresponding GO category; numInCat = number of RNAAgeCalc genes included in the corresponding GO category; Padj = FDR adjusted p-value across the number of GO categories examined.

Table S3

Full result of GO terms analysis for PTSD/MDD associated RNA age genes - see separately uploaded csv file (SuppTableS3.csv)

Table S3 note: over_represented_pvalue = *p*-value of overrepresentation; numDEInCat = number of PTSD/MDD associated RNAAgeCalc genes included in the corresponding GO category; numInCat = number of RNAAgeCalc genes included in the corresponding GO category; over_represented_*p*adj = adjusted *p*-value of overrepresentation by FDR across the number of GO categories examined; sig.genes.col = PTSD/MDD associated RNAAgeCalc genes included in the corresponding GO category.

Table S4

Associations between PTSD/MDD and Cell Type Content Estimates in vmPFC

Cell type	B	SE	*p*	*p*-adj
Astrocytes	0.012	0.102	0.909	0.909
Endothelial	**-0.294**	**0.098**	**0.004**	**0.017**
Microglia	-0.261	0.154	0.095	0.221
Mural	**-0.234**	**0.081**	**0.005**	**0.017**
Neurons	0.109	0.117	0.354	0.619
Oligodendrocytes	-0.109	0.176	0.536	0.626
RBC	0.087	0.116	0.453	0.626

Note. Regressions controlled for sex, sequencing run IDs, and first three qSVs. Cell type content scores were estimated via BrainInaBlender. Significant effects are shown in bold font. B = unstandardized coefficient; SE = standard error; *P*-adj = FDR adjusted *p*-value across the number of cell types; Endothelial = endothelial cells; Mural = mural cells; RBC = red blood cells.

Table S5

Cell Type Enrichment Analysis of Genes Included in the RNAAgeCalc Algorithm in vmPFC (n=448)

Cell type	*N*	Count	Gene ratio (%)	Gene	*p*	*p*-adj
Ex1	305	2	0.446	PRKCB, PPP2R2C	0.992	1
Ex2	219	2	0.446	ATXN1, SLC39A10	0.949	1
Ex3e	372	4	0.893	RPL35, AKAP11, SLC39A10, PRKCB	0.964	1
Ex4	293	4	0.893	BRINP1, HIVEP2, ZMYND8, PRKCB	0.872	1
Ex5b	240	4	0.893	SLC39A10, PRKCB, AFF3, PPP2R2C	0.744	1
Ex6a	233	6	1.339	ROBO3, HIVEP2, STRBP, CADPS, RANBP17, GALNT14	0.342	1
Ex6b	218	6	1.339	HIVEP2, STRBP, CADPS, PRKCB, AFF3, NLGN4Y	0.267	1
Ex8	299	7	1.563	RPL35, MGAT5, CADPS, MEF2A, PRKCB, PPP2R2C, GALNT14	0.420	1
Ex9	200	2	0.446	SLC39A10, WSB2	0.940	1
In1a	257	6	1.339	IGF1, MGAT5, NRIP3, HMBOX1, GALNTL6, KIAA1211	0.439	1
In1b	240	3	0.670	GALNTL6, RANBP17, NLGN4Y	0.889	1
In1c	194	3	0.670	IGF1, GALNTL6, KIAA1211	0.776	1
In3	195	2	0.446	GALNTL6, RANBP17	0.921	1
In4a	215	3	0.670	GALNTL6, KIAA1211, PRKCB	0.834	1
In4b	289	4	0.893	KIT, NRIP3, KIAA1211, RANBP17	0.870	1
In6a	219	4	0.893	GALNTL6, CADPS, KIAA1211, AFF3	0.665	1
In6b	231	3	0.670	ESRRG, CADPS, FAR2	0.876	1
In7	232	4	0.893	ARHGEF3, CADPS, KIAA1211, RCAN2	0.737	1
In8	197	4	0.893	NRIP3, GALNTL6, CADPS, RCAN2	0.612	1
Endo	**83**	**8**	**1.786**	**IFITM3, HLA-B, HIGD1B, HERC2P3, IGFBP7, SLC39A10, SLC7A5, GPCPD1**	**0.0002**	**0.005**
Per	64	2	0.446	HIGD1B, IGFBP7	0.387	1
Astro	159	1	0.223	LGI4	0.970	1
Oligo	179	0	0		1.000	1
OPC	143	0	0		1.000	1
Microglia	97	2	0.446	MEF2A, FAM49B	0.602	1

Note. Significant effects are shown in bold font. Ex=excitatory neurons; In=inhibitory neurons; Endo=endothelial cells; Per=pericytes; Astro=astrocytes; Oligo=oligodendrocytes; OPC=Oligodendrocyte progenitor cells; N=number of genes annotated by cell type markers. Count=number of genes in the RNA age algorithm that are annotated by the respective cell type markers. *P*-adj= FDR adjusted *p*-value across number of cell type markers examined (n = 25).

Table S6

Associations between RNA Age and Cell Type Content Estimates in vmPFC

Cell type	B	SE	*p*	*p*-adj
Astrocytes	-0.004	0.006	0.579	0.810
Endothelial	-0.010	0.006	0.144	0.352
Microglia	-0.013	0.010	0.197	0.352
Mural	-0.007	0.005	0.201	0.352
Neurons	0.002	0.007	0.796	0.819
Oligodendrocytes	**0.027**	**0.011**	**0.013**	**0.092**
RBC	-0.002	0.007	0.819	0.819

Note. Regressions controlled for sex, sequencing run IDs, and first three qSVs. Cell type content scores were estimated via BraininaBlender. Significant effect is shown in bold font. B = unstandardized coefficient; SE = standard error; *P*-adj = FDR adjusted *p*-value across the number of cell types; Endothelial = endothelial cells; Mural = mural cells; RBC = red blood cells.

Table S7

Associations between Age at Death and Cell Type Content Estimates in vmPFC

Cell type	B	SE	*p*	*p*-adj
Astrocytes	0.004	0.004	0.294	NA
Endothelial	0.003	0.004	0.438	NA
Microglia	<0.001	0.006	0.970	NA
Mural	0.004	0.003	0.223	NA
Neurons	-0.005	0.004	0.300	NA
Oligodendrocytes	0.011	0.007	0.099	NA
RBC	-0.001	0.004	0.877	NA

Note. Regressions controlled for sex, sequencing run IDs, and first three qSVs. Cell type content scores were estimated via BrainInaBlender. B = unstandardized coefficient; SE = standard error; *P*-adj = FDR adjusted *p*-value across the number of cell types; Endothelial = endothelial cells; Mural = mural cells; RBC = red blood cells.

Table S8

Top 5 Most Significant Associations between Genes in the RNAAgeCalc Algorithm and Cell Type Content Estimates in vmPFC

Cell Type	Gene	# Corrected Significant Association	log2FoldChange	*p*	*p*-adj
Astrocyte	BBOX1	134	1.423	1.243E-26	6.272E-24
	FKBP10		0.816	2.461E-23	7.017E-21
	UNG		0.624	5.368E-16	5.043E-14
	PLCD1		0.727	5.160E-15	4.258E-13
	CADPS		-0.528	9.664E-14	6.735E-12
Endothelial	TINAGL1	162	1.513	5.866E-27	3.679E-24
	TLN1		0.939	1.233E-22	3.222E-20
	HLA-B		1.125	1.332E-18	2.321E-16
	IFITM1		1.636	3.422E-18	5.365E-16
	ENG		1.144	4.552E-17	5.437E-15
Microglia	VSIG4	167	2.435	6.273E-36	9.836E-33
	MAN2B1		0.747	3.530E-28	2.768E-25
	MS4A6A		1.918	1.503E-21	3.142E-19
	HLA-B		0.716	5.093E-18	7.260E-16
	MS4A4A		1.747	1.001E-17	1.308E-15
Mural	AEBP1	129	1.877	2.031E-16	2.123E-14
	TINAGL1		1.615	6.555E-16	5.873E-14
	SYDE1		1.233	4.956E-14	3.701E-12
	TLN1		0.979	1.910E-13	1.223E-11

	IFITM1		1.732	1.070E-11	5.082E-10
Neuron	PRKCB	199	0.939	1.160E-48	3.639E-45
	PTPN3		0.937	2.326E-33	2.431E-30
	ARHGEF3		0.656	1.400E-26	6.272E-24
	HIVEP2		0.544	1.649E-24	6.466E-22
	C1QL3		1.194	9.178E-24	2.878E-21
Oligodendrocyte	AQP1	137	2.214	4.047E-24	1.410E-21
	C14orf132		-0.460	1.252E-21	2.971E-19
	CRB2		0.766	1.326E-21	2.971E-19
	KIF19		1.991	1.199E-20	2.212E-18
	KIAA1324L		0.411	1.442E-18	2.380E-16
RBC	HBG2	88	1.767	4.542E-09	1.228E-07
	CDC42EP4		0.434	7.914E-06	8.392E-05
	ZNF711		-0.342	8.390E-06	8.764E-05
	ORC4		-0.400	1.483E-05	1.422E-04
	BRWD1		-0.237	1.492E-05	1.427E-04

Note. Log2FoldChange = log2 fold change; *p*-adj = FDR adjusted *p*-value across number of genes examined and cell types (*n* = 448 x 7). Endothelial = endothelial cells; Mural = mural cells; RBC = red blood cells;

Table S9

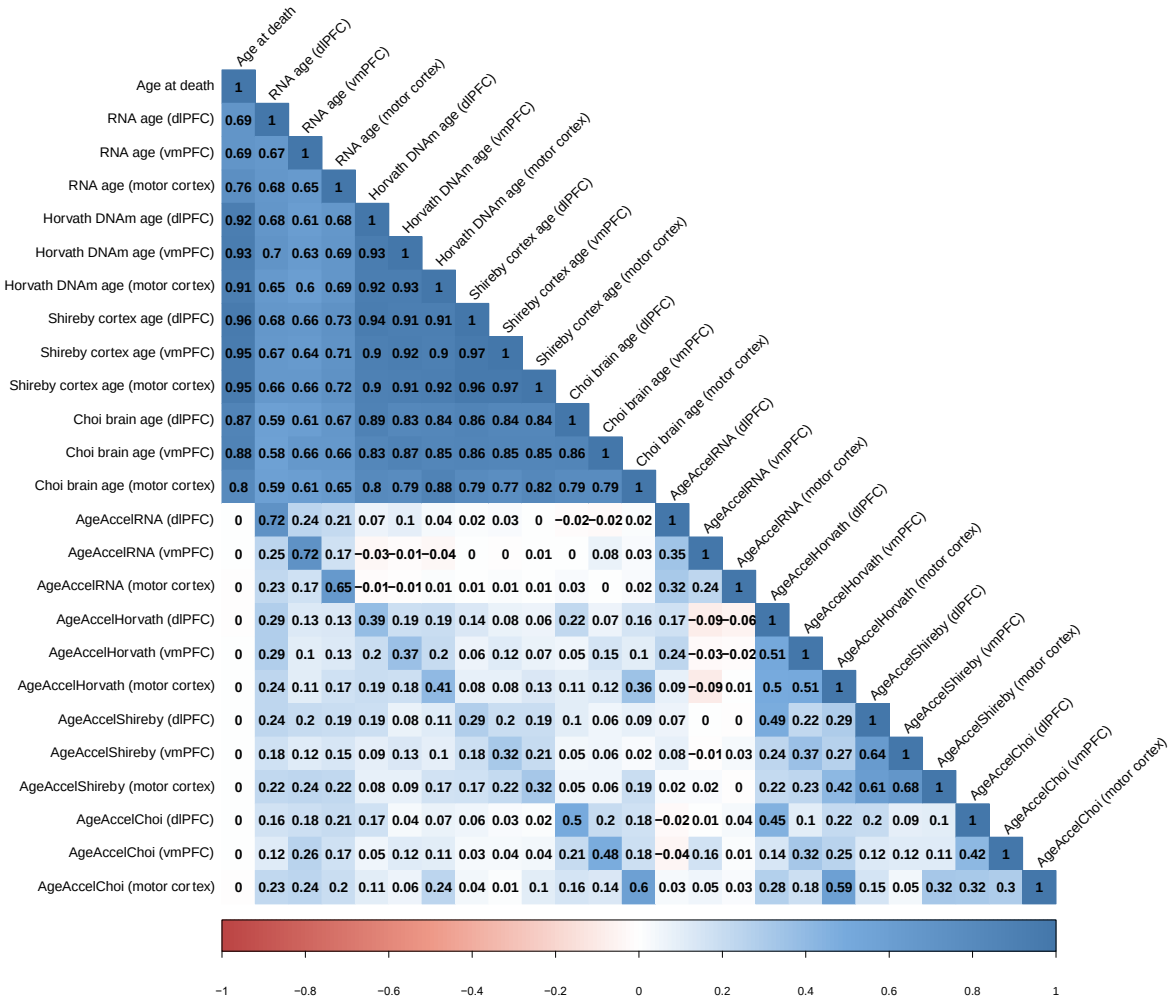
Associations between RNA Age Residuals and Cell Type Content Estimates in dlPFC and motor cortex

Cell type	dlPFC				Motor cortex			
	B	SE	*p*	*p*-adj	B	SE	*p*	*p*-adj
Astrocytes	**-0.024**	**0.009**	**0.007**	**0.016**	0.007	0.010	0.508	0.711
Endothelial	**-0.029**	**0.007**	**0.0002**	**0.001**	-0.014	0.008	0.094	0.282
Microglia	**-0.023**	**0.011**	**0.031**	**0.044**	-0.016	0.013	0.215	0.376
Mural	**-0.027**	**0.007**	**0.0003**	**0.001**	-0.010	0.006	0.121	0.282
Neurons	0.012	0.006	0.059	0.069	0.002	0.008	0.800	0.800
Oligodendrocytes	**0.021**	**0.009**	**0.023**	**0.040**	0.017	0.010	0.076	0.282
RBC	0.012	0.008	0.136	0.136	0.003	0.008	0.661	0.771

Note. Regressions controlled for sex, sequencing run IDs, and first three qSVs. Cell type content scores were estimated via BraininaBlender. *P*-values were adjusted across the number of cell types within each brain region. Significant effects are shown in bold font. B = unstandardized coefficient; SE = standard error; *p*-adj = FDR adjusted *p*-value across the number of cell types; Endothelial = endothelial cells; Mural = mural cells; RBC = red blood cells.

Supplementary Figures

Supplementary Figure S1



The figure shows a correlation heatmap illustrating the correlations among age at the time of death, RNA age, Horvath DNAm age, Choi brain-specific DNAm age, Shireby cortical DNAm age, RNA age residuals, Horvath age residuals, Shireby age residuals, and Choi age residuals across dlPFC, vmPFC, and motor cortex. AgeAccelRNA = RNA age residuals. AgeAccelHorvath = Horvath DNAm age residuals. AgeAccelShireby = Shireby DNAm age residuals. AgeAccelChoi = Choi DNAm age residuals.

Supplementary References

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.

2. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics*. 2020 Feb 1;36(3):930–3.

3. Ratanatharathorn A, Boks MP, Maihofer AX, Aiello AE, Amstadter AB, Ashley-Koch AE, et al. Epigenome-wide association of PTSD from heterogeneous cohorts with a common multi-site analysis pipeline. *Am J Med Genet Part B Neuropsychiatr Genet*. 2017 Sep;174(6):619–30.

4. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics*. 2012 May 1;28(9):1280–1.

5. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017 Dec 15;33(24):3982–4.

6. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics*. 2014 Feb 1;30(3):428–30.

7. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013 May 1;14(1):293.

8. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: impute: Imputation for microarray data [Internet]. Bioconductor version: Release (3.14); 2021. Available from: https://bioconductor.org/packages/impute/

9. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012 Mar 15;28(6):882–3.

10. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.

11. Choi H, Joe S, Nam H. Development of Tissue-Specific Age Predictors Using DNA Methylation Data. *Genes*. 2019 Nov 4;10(11):888.

12. Shireby GL, Davies JP, Francis PT, Burrage J, Walker EM, Neilson GWA, et al. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain*. 2020 Oct 29;143(12):3763–75.

13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114–20.

14. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res*. 2017 May 1;27(5):849–64.

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15–21.

16. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

17. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012 Aug 15;28(16):2184–5.

18. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016 Oct 1;32(19):3047–8.

19. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016 May;34(5):525–7.

20. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521.

21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.

22. Jaffe AE, Tao R, Norris AL, Kealhofer M, Nellore A, Shin JH, et al. qSVA framework for RNA quality correction in differential expression analysis. *Proc Natl Acad Sci U S A*. 2017 Jul 3;114(27):7130–5.

23. Hagenauer MH, Schulmann A, Li JZ, Vawter MP, Walsh DM, Thompson RC, et al. Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. *PLOS ONE*. 2018 Jul 17;13(7):e0200003.

24. Ren X, Kuan PF. RNAAgeCalc: A multi-tissue transcriptional age calculator. *PloS One*. 2020;15(8):e0237006.

25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300.

26. Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, et al. The PsychENCODE project. *Nat Neurosci*. 2015 Dec;18(12):1707–12.

27. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell-type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019 Jul;37(7):773–82.