Supplemental for:

Divergent landscapes of A-to-I editing in postmortem and living human brain

Miguel Rodriguez de los Santos[1*], Brian H. Kopell[1*], Ariela Buxbaum Grice[1], Gauri Ganesh[1], Andy Yang[1], Pardis Amini[1], Lora E. Liharska[1], Eric Vornholt[1], John Fullard[1], Pengfei Dong[1], Eric Park[1], Sarah Zipkowitz[1], Deepak A. Kaji[1], Ryan C. Thompson[1], Donjing Liu[1], You Jeong Park[1], Esther Cheng[1], Kimia Ziafat[1], Emily Moya[1], Brian Fennessy[1], Lillian Wilkins[1], Hannah Silk[1], Lisa M. Linares[1], Brendan Sullivan[1], Vanessa Cohen[1], Prashant Kota[1], Claudia Feng[1], Jessica S. Johnson[1], Marysia-Kolbe Rieder[1], Joseph Scarpa[1], Girish N. Nadkarni[1], Minghui Wang[1], Bin Zhang[1], Pamela Sklar[1], Noam D. Beckmann[1], Eric E. Schadt[1], Panos Roussos[1*], Alexander W. Charney[1*], Michael S. Breen[1*]

Affiliations:

[1]Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

[*]These authors contributed equally

Correspondence to:

Michael S. Breen PhD, michael.breen@mssm.edu

CONTENTS:

Supplemental Note

A deep atlas of neuronal and non-neuronal A-to-I editing sites across five postmortem cortical areas

While global *Alu* editing rates can be accurately measured via snRNA-seq[1], it is challenging to ascertain individual cell-specific sites with high confidence due to individual differences in cell type composition and low sequencing depth, with reads covering only a fraction of the entire transcriptome[1,2] (**Figure S8**). A-to-I editing levels are known to be elevated in GABAergic and glutamatergic neurons relative to oligodendrocytes, and these differences are mostly, but not fully, explained by increased expression of *ADAR* and *ADARB1*[1]. Similar profiles have been observed in mice and *Drosophila*[3,4], which further support our findings. These contrasting profiles between neurons and non-neurons prompted us to integrate a deeply sequenced resource of neuronal and non-neuronal nuclei isolated from ten biological replicates across five postmortem cortical regions[5] (**Figure S9A**). Here, the depth of read coverage and the major cell types and anatomical regions profiled, provide an optimal framework cataloguing cell-type specific A-to-I sites in the human cortex – with the caveat that these are derived from postmortem tissues. Fluorescence activated nuclei sorting (FANS) was applied to an antibody against the well-established neuronal marker NeuN (*i.e.* RNA-binding protein RBFOX3) was used to isolate neuronal from non-neuronal nuclei followed by bulk RNA-sequencing. The AEI was computed for each sample and confirmed higher levels of global *Alu* editing in neuronal versus non-neuronal nuclei (p=2.0×10^{-9}) (**Figure S9B**). No significant differences were observed in global *Alu* editing between cortical areas within each cellular population. The expression of *ADAR* and *ADARB1* was significantly higher in neuronal nuclei (p=8.4×10^{-15}, p=2.1×10^{-18}, respectively) and positively correlated with the AEI (r=0.47, r=0.52, respectively), while *ADARB2* was more highly expressed in non-neuronal nuclei (p=4.2×10^{-15}) and negatively correlated with the AEI (r=-0.51) (**Figure S9C**). A linear mixed model quantified the variance of the AEI explained by known factors, and cell type differences explained the largest median fraction of *Alu* editing variability (~48%), followed by donor as a repeated measure (~40%) (**Figure S9D**).

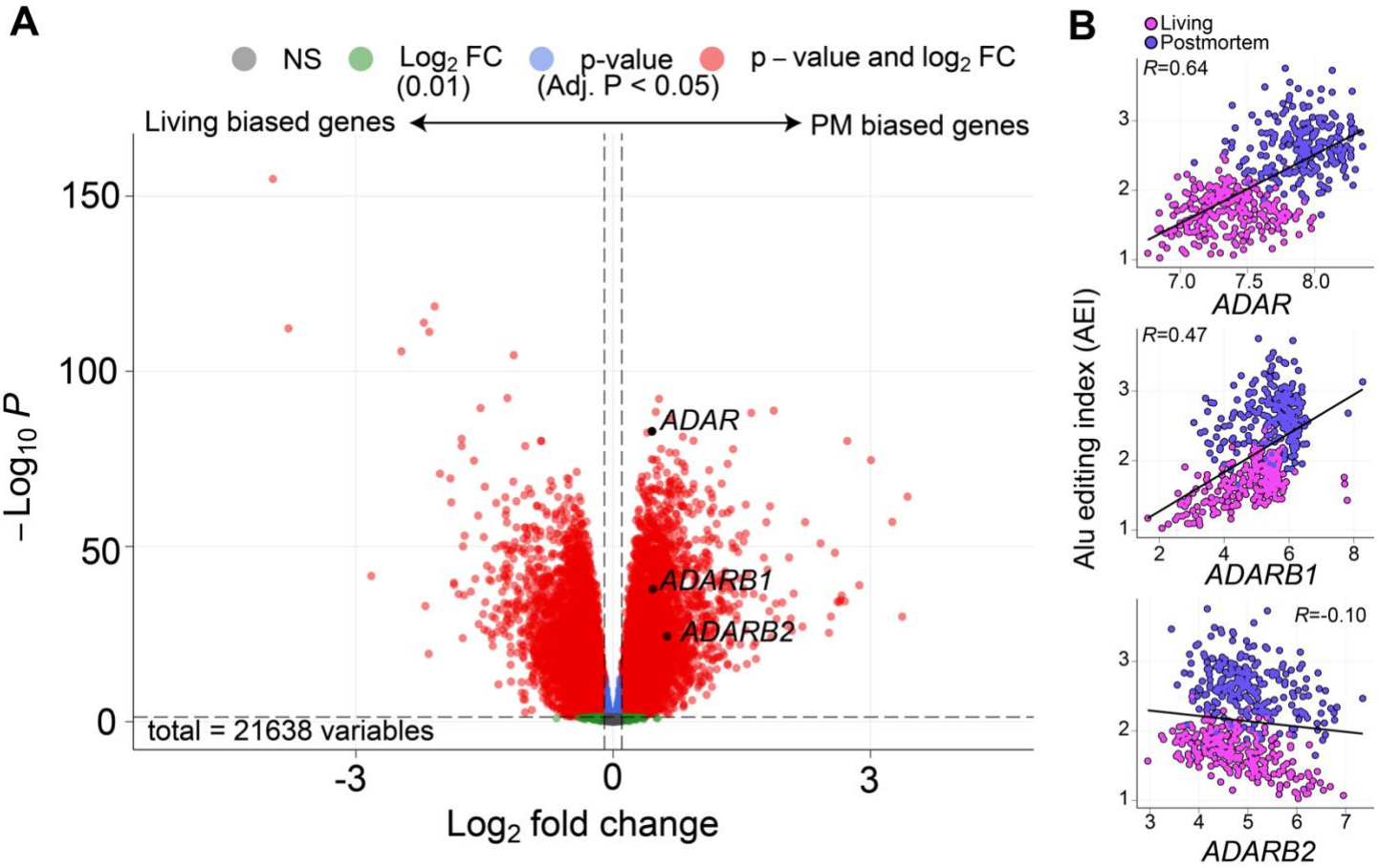
To quantify individual cell-specific sites, we applied the same pipeline as above to query high-confidence A-to-I sites across neuronal and non-neuronal nuclei (**Supplemental Data 4**). A mean of 1,036,579 editing sites were detected per sample across neuronal nuclei and 925,440 sites were detected per sample across non-neuronal nuclei (**Figure S9E**), which showed similar hallmarks of ADAR-mediated RNA editing, in that the majority: (1) were A-to-I sites (~84% neuronal, ~88% non-neuronal); (2) mapped to *Alu* repeats (~75% neuronal, ~75% non-neuronal); (3) were predominately known sites cataloged in editing databases (~78% neuronal, ~80% non-neuronal). The enrichment of A-to-I sites in non-coding regions is consistent with existing reports that intronic regions are over-represented in RNA-seq and snRNA-seq generated from the nuclear fraction relative to a combined cytoplasmic and nuclear fraction[6-8]. In general, site detection was convergent between neuronal and non-neuronal populations, with little-to-no regional specificity (**Figure S10**). PCA applied to the editing levels of 290,495 sites detected across all nuclei and regions confirmed a significant cell-type effect along PC1 and a donor effect along PC2 (**Figure S9F**). Differential editing tested for differences in mean editing levels between neuronal and non-neuronal nuclei: 157,546 sites displayed higher editing levels in neurons ('neuronal-biased') and 44,485 sites displayed higher editing levels in non-neuronal cells (non-neuronal biased) (**Figure S9G**). Covarying for *ADARB1* and *ADAR* expression explained a large fraction of differentially edited sites (**Figure S9H**). Most differentially edited sites mapped to non-coding regions (**Figure S9I**). A total of 82 recoding sites were differentially edited; 71 were neuronal-biased, which were commonly evolutionary conserved (phastCons) and mapped to genes with elevated pLI, underscoring their putative functionality (**Figure S9J**). Finally, we replicated these findings and significantly expand the scope of existing catalogues of cell-specific RNA editing

in the cortex[1] (**Figure S11**), offering 201,941 A-to-I sites specific to either neuronal or non-neuronal nuclei. All results are presented in **Supplemental Data 4**.

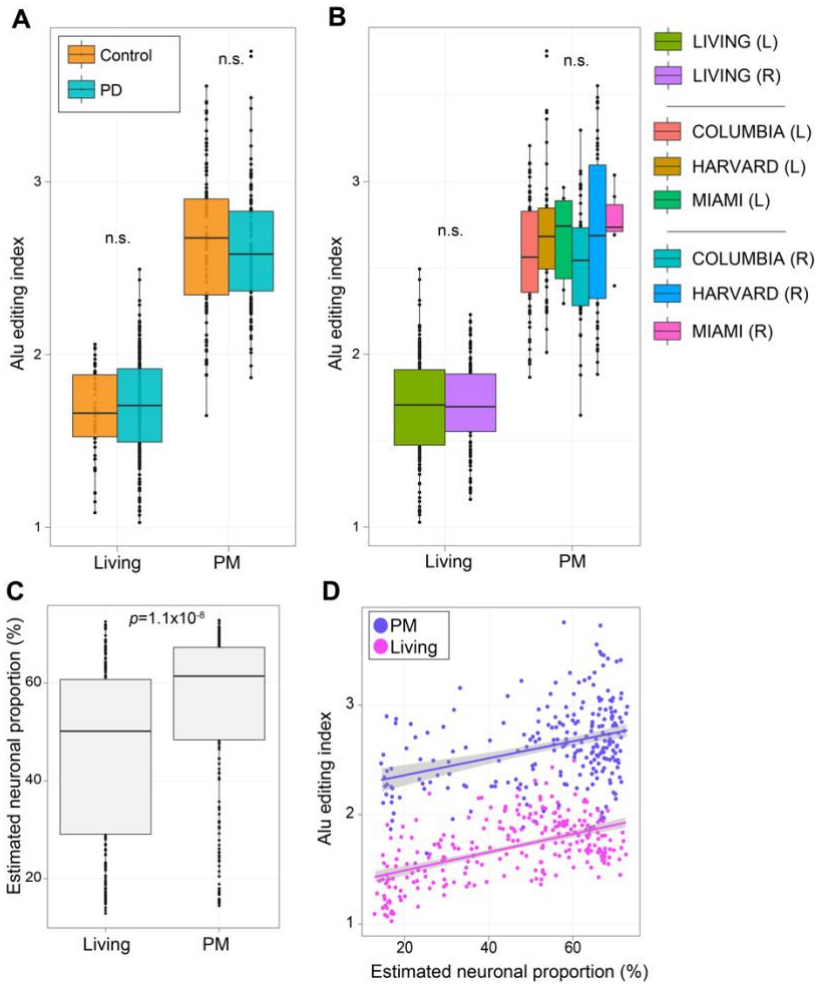
Supplemental References Related to Supplemental Note

1. Cuddleston WH, Li J, Fan X, Kozenkov A, Lalli M, Khalique S, Dracheva S, Mukamel EA, Breen MS. Cellular and genetic drivers of RNA editing variation in the human brain. Nature Communications. 2022 May 30;13(1):2997.
2. Ansell BR, Thomas SN, Bonelli R, Munro JE, Freytag S, Bahlo M. A survey of RNA editing at single-cell resolution links interneurons to schizophrenia and autism. RNA. 2021 Dec 1;27(12):1482-96
3. Gal-Mark N, Shallev L, Sweetat S, Barak M, Billy Li J, Levanon EY, Eisenberg E, Behar O. Abnormalities in A-to-I RNA editing patterns in CNS injuries correlate with dynamic changes in cell type composition. Scientific reports. 2017 Mar 7;7(1):43421.
4. Sapiro AL, Shmueli A, Henry GL, Li Q, Shalit T, Yaron O, Paas Y, Billy Li J, Shohat-Ophir G. Illuminating spatial A-to-I RNA editing signatures within the Drosophila brain. Proceedings of the National Academy of Sciences. 2019 Feb 5;116(6):2318-27.
5. Dong P, Hoffman GE, Apontes P, Bendl J, Rahman S, Fernando MB, Zeng B, Vicari JM, Zhang W, Girdhar K, Townsley KG. Population-level variation in enhancer expression identifies disease mechanisms in the human brain. Nature Genetics. 2022 Oct;54(10):1493-503.
6. Price AJ, Hwang T, Tao R, Burke EE, Rajpurohit A, Shin JH, Hyde TM, Kleinman JE, Jaffe AE, Weinberger DR. Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex uncovers new insight into psychiatric disease gene regulation. Genome research. 2020 Jan 1;30(1):1-1.
7. Zaghloul HA, Hice RH, Arensburger P, Bideshi DK, Federici BA. Extended in vivo transcriptomes of two ascoviruses with different tissue tropisms reveal alternative mechanisms for enhancing virus reproduction in hemolymph. Scientific reports. 2021 Aug 12;11(1):16402.
8. Lee S, Zhang AY, Su S, Ng AP, Holik AZ, Asselin-Labat ML, Ritchie ME, Law CW. Covering all your bases: incorporating intron signal from RNA-seq data. NAR Genomics and Bioinformatics. 2020 Sep;2(3):lqaa073.

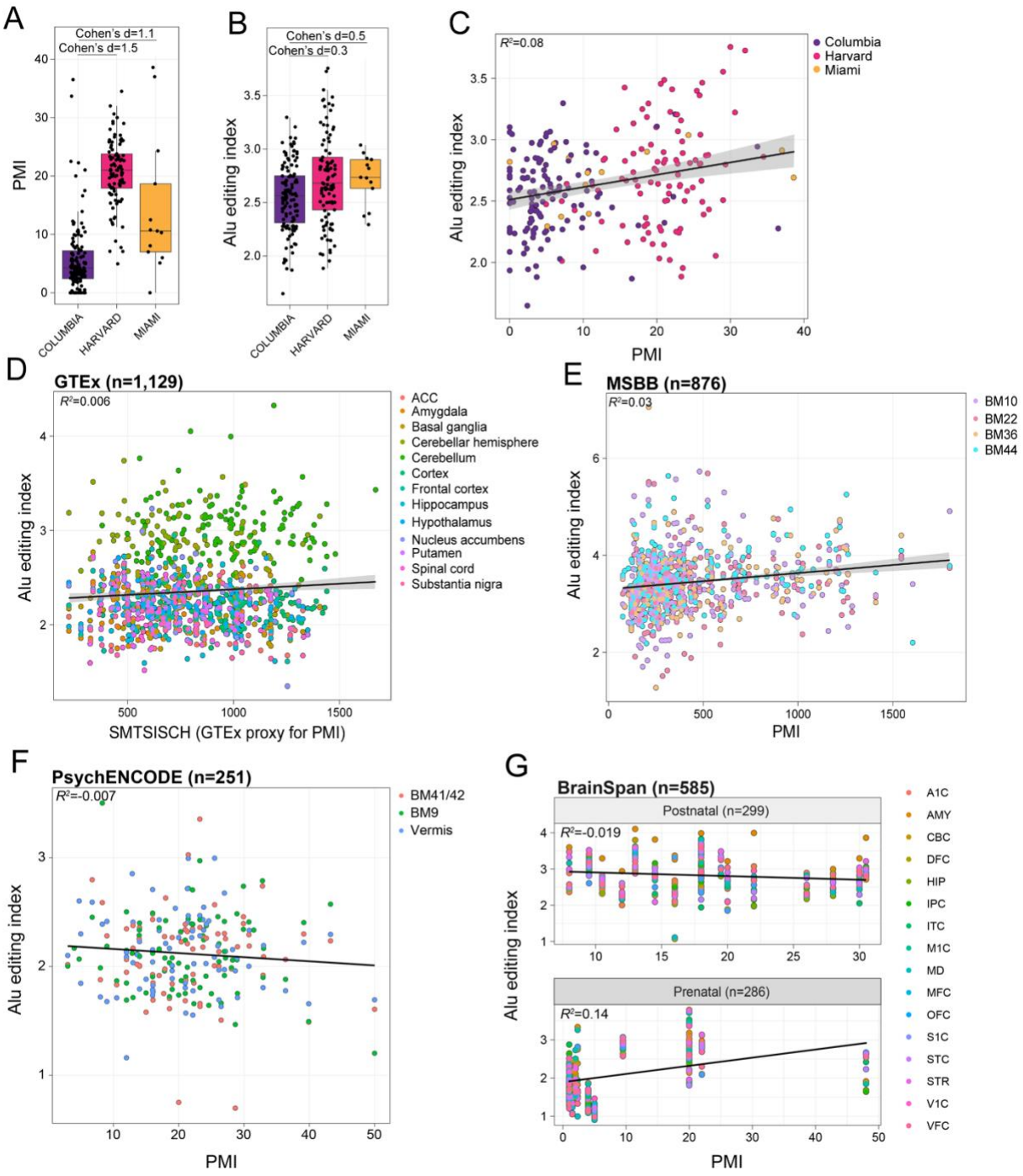
Supplemental Figures 1-16



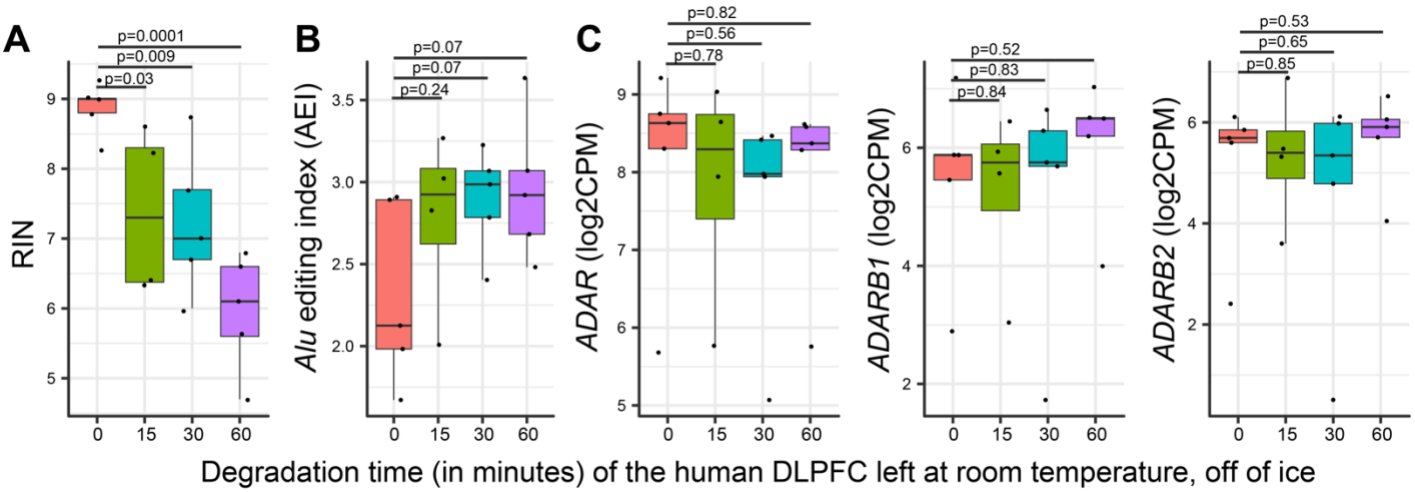
Supplemental Figure 1. Transcriptome-wide differential gene expression between postmortem and living DLPFC. (**A**) Volcano plot of differentially expressed genes between postmortem (PM) and living DLPFC compares $\log_2$ fold change (x-axis) with strength of significance ($-\log_{10}$ adjusted p-value, y-axis). Genes *ADAR*, *ADARB1* and *ADARB2* are colored black to visualize their effect size differences relative to the remaining transcriptome. (**B**) Pearson's correlation coefficients between the AEI and normalized gene expression (log2CPM) for *ADAR* (top), *ADARB1* (middle) and *ADARB2* (bottom).



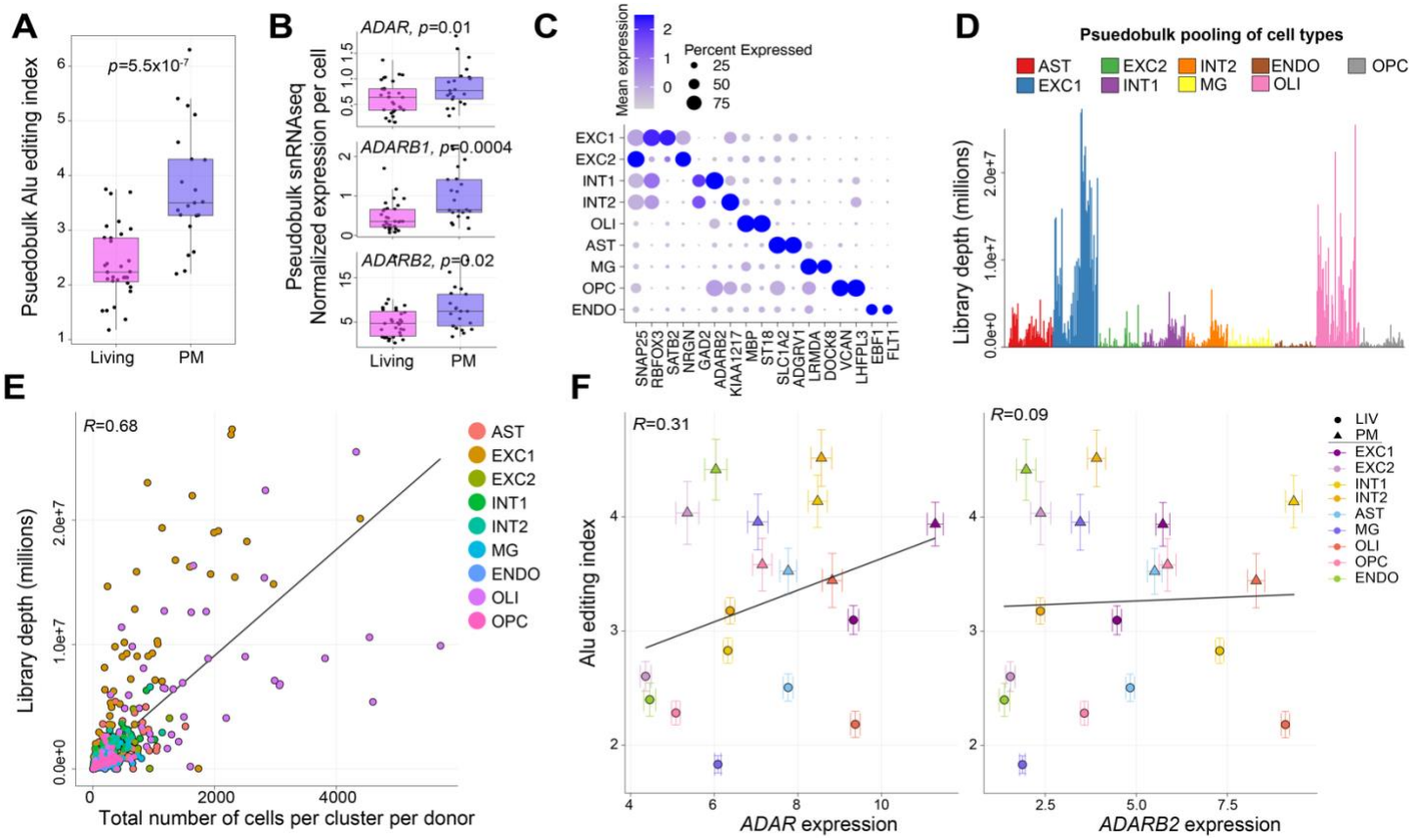
Supplemental Figure 2. The effect of medical diagnosis, biobank and neuronal content on *Alu* editing. (**A**) The *Alu* editing index (AEI; y-axis) between living and postmortem (PM) DLPFC parsed by individuals with Parkinson's disease (PD) and controls (x-axis). Two-sided linear regression was used to test for significance. No significant differences (n.s.) were observed. (**B**) The AEI (y-axis) between living and PM DLPFC parsed by different brain banks (x-axis). A *Kruskal–Wallis test* was used to test for significance among PM DLPFC samples. Further, differences between the left and the right hemisphere were tested within living DLPFC using a two-sided linear regression. No significant differences (n.s.) were observed. (**C**) dtangle estimated neuronal cell type proportions (y-axis) between living and PM DLPFC (x-axis). Two-sided linear regression was used to test for significance. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**D**) Estimated neuronal proportions (y-axis) relate to changes in the AEI within both living and PM DLPFC (x-axis). A Pearson's correlation coefficient was used to test each association.



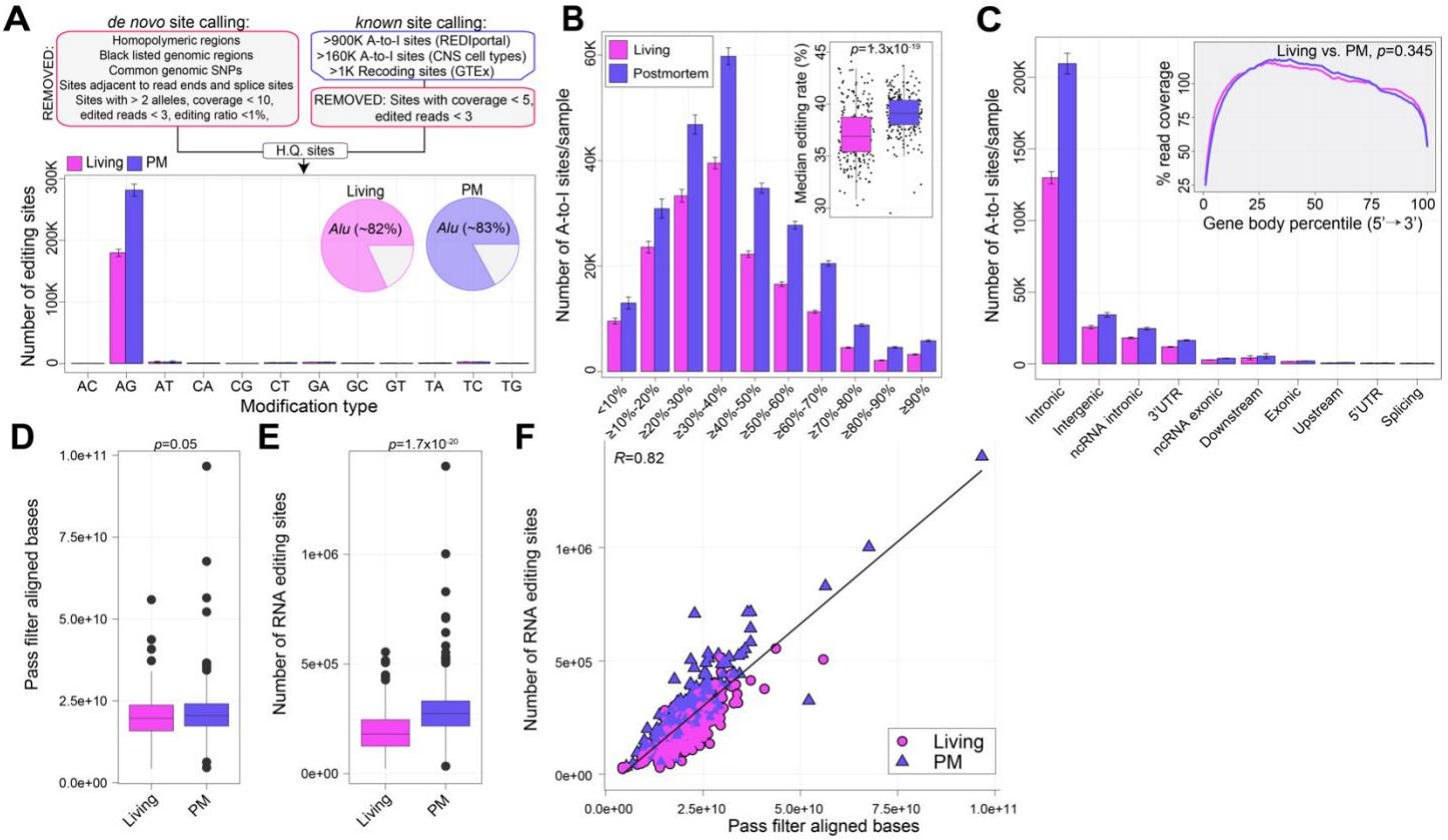
Supplemental Figure 3. The relationship between extended postmortem interval and *Alu* editing. (**A**) Differences in postmortem interval (PMI; y-axis) within three postmortem brain banks (x-axis) used in the Living Brain Project. (**B**) Differences in Alu editing index (AEI; y-axis) between the three postmortem brain banks (x-axis). Cohen's D was used to measure effect size differences between brain banks. All boxplots show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**C**) Pearson's correlation coefficient of the relationship between PMI (x-axis) and the AEI (y-axis) in the Living Brain Project. (**D**) The relationship between PMI (x-axis) and the AEI (y-axis) for 1,129 bulk tissue postmortem RNA-seq samples across 13 brain regions. (**E**) The relationship between PMI (x-axis) and the AEI (y-axis) for 876 postmortem bulk RNA-seq samples across four cortical areas from the Mount Sinai Brain Bank (MSBB). A Pearson's correlation coefficient was used to test each association.



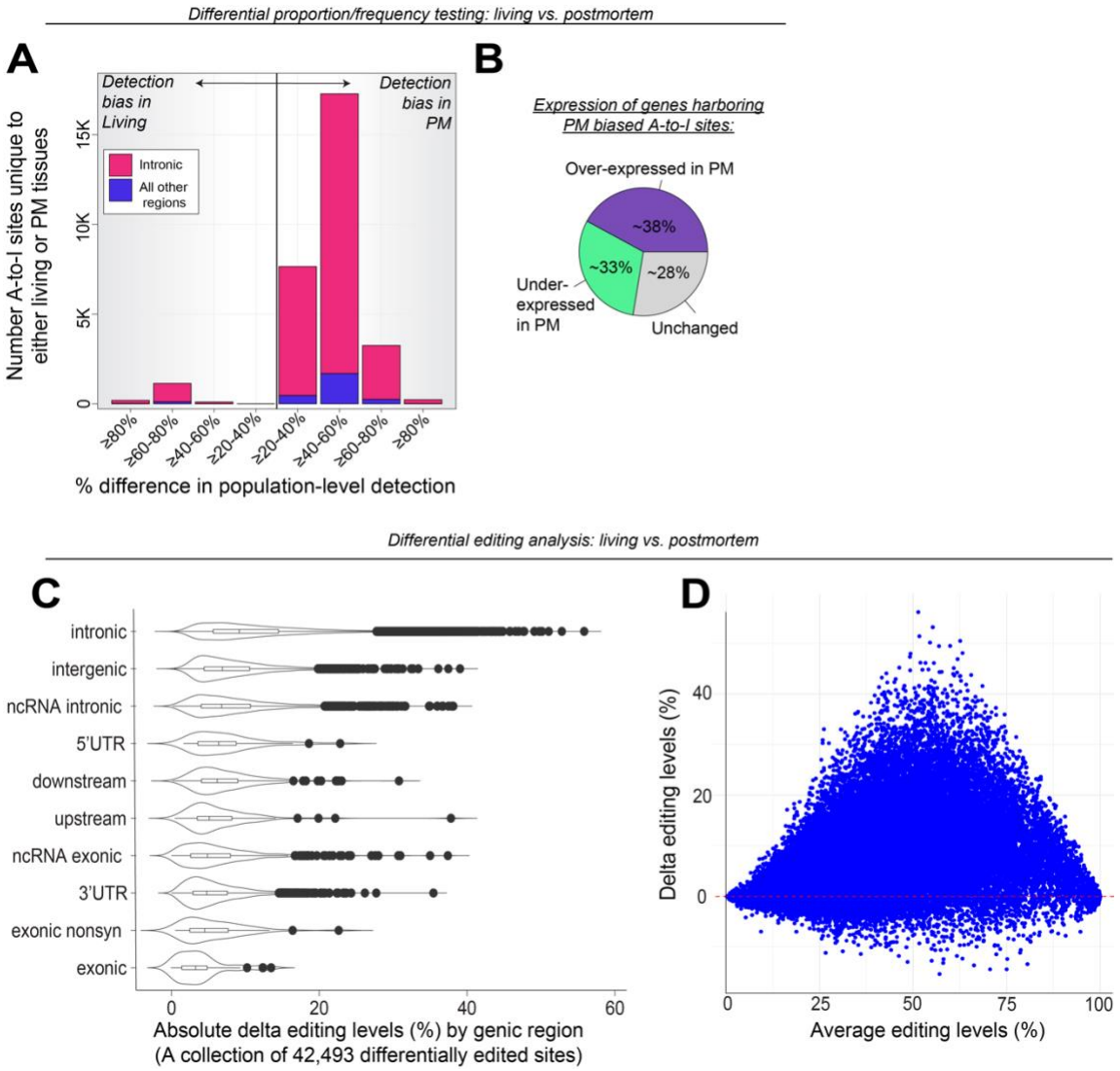
Supplemental Figure 4. **Quantification of RNA editing metrics throughout molecular degradation of the human DLPFC**. We downloaded existing RNA-sequencing data of a molecular degradation assay of the human DLPFC (PMID: 28634288). These data were prepared using RiboZero RNA-seq sequencing library, which align with the RNA-seq library preparation methods used in the current study. Quantification of (**A**) RNA integrity numbers (RINs) and (**B**) the AEI, as well as (**C**) *ADAR*, *ADARB1*, and *ADARB2* expression throughout advancing degradation of the DLPFC. Degradation of the DLPFC was measured in minutes left at room temperature (off of ice). For each measurement, a two-sided *t*-test compared 0 minutes (baseline) to each of the subsequent time-points without adjusting for multiple comparisons.



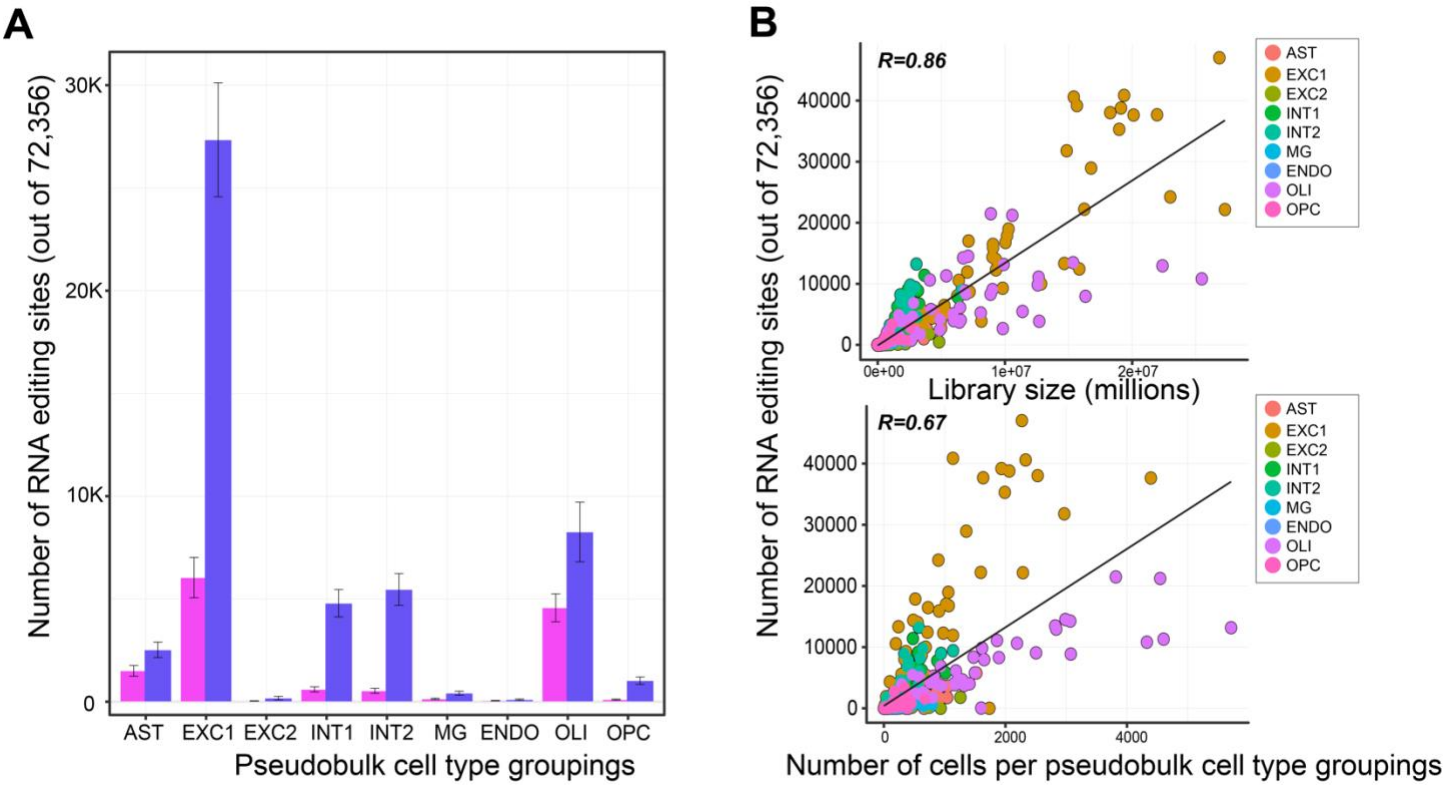
Supplemental Figure 5. Pseudo-bulk snRNA-seq analyses. (**A**) Leveraging snRNA-seq as pseudo-bulk tissue to measure the *Alu* editing index (AEI) between living and postmortem (PM) DLPFC. Two-sided linear regression was used to test for significance. (**B**) Raw counts of *ADAR*, *ADARB1* and *ADARB2* normalized to the total number of cells sequenced per donor, illustrating increased expression per cell across each gene in PM relative to living DLPFC. Mann-Whitney U test was used to test for significance. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**C**) Dot plot of 17 cell marker genes used to define cell identities for nine cell populations identified via UMAP from snRNA-seq data. (**D**) Cell-type specific bam files were constructed for nine different cell types for each donor. Large differences in final library depth (y-axis) were observed across each pseudo-bulk cellular pool (x-axis). (**E**) Library depth for each pseudo-bulk cellular pool (y-axis) is correlated with the total number of cells per pseudo-bulk cellular pool per donor (x-axis). (**F**) Mean AEI (y-axes) relative to mean expression of *ADAR* (left) and *ADARB2* (right) for each pseudo-bulk cellular pool within living and PM DLPFC (x-axes). A Pearson's correlation coefficient was used to test each association. Standard error bars capture group-wise variance within living and postmortem tissues, respectively.



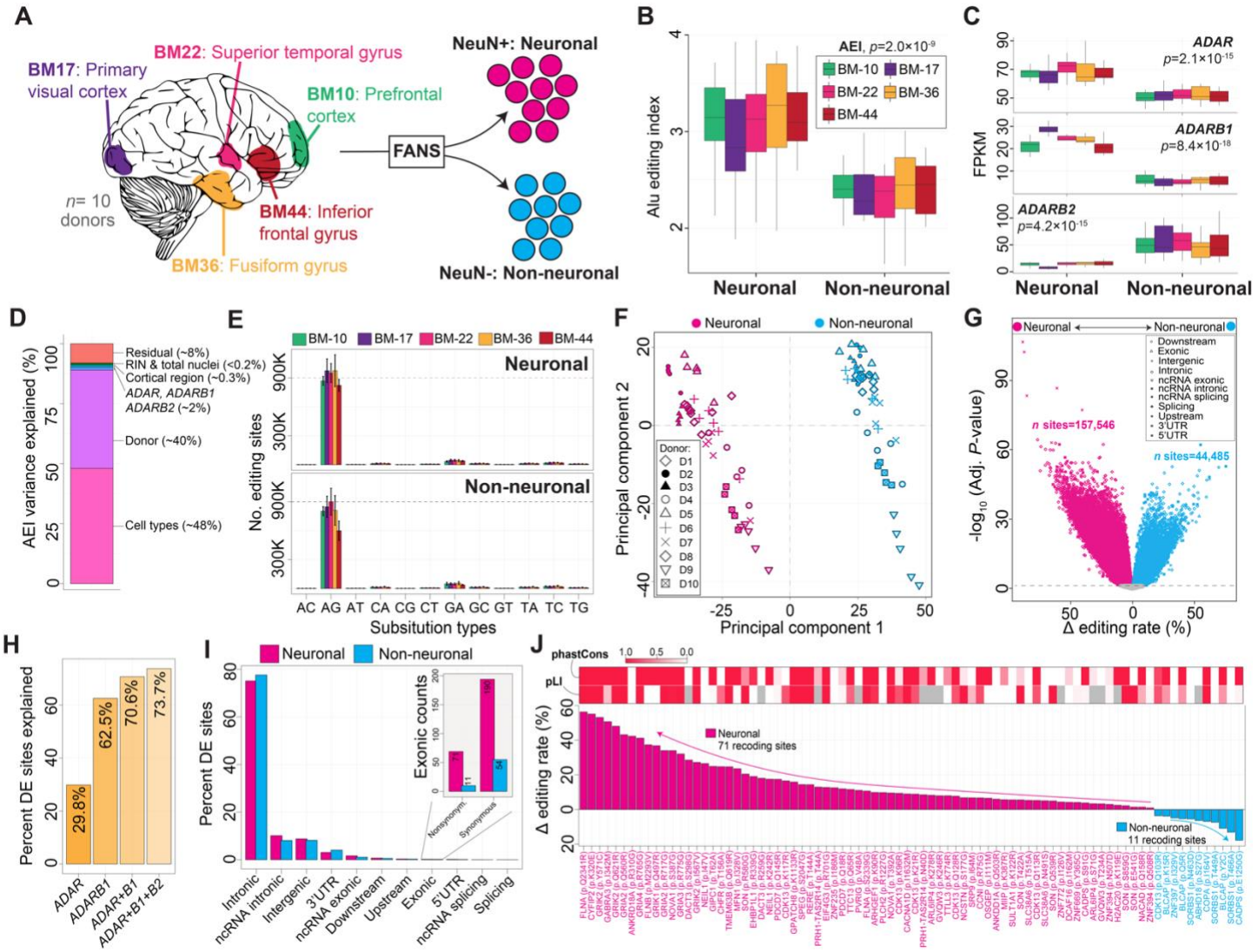
Supplemental Figure 6. RNA editing site detection from bulk RNA-sequencing data. (**A**) Uncovering high-quality (HQ) sites (top) using a combined de novo and supervised RNA editing site detection approach. Bar plots depict mean (with standard error) number of HQ sites for living (n = 164) and postmortem (PM) (n = 233) DLPFC samples based on substitution type and *Alu* repeat element (bottom). (**B**) The number of A-to-I sites (y-axis) binned by different editing levels in living and PM DLPFC (x-axis). Standard error bars show variance. Inset boxplots depict the median editing rates across all A-to-I sites per donor. Two-sided linear regression was used to test for significance. (**C**) Number of A-to-I sites detected per sample for living and postmortem tissues. Standard error bars show variance. Inset figure shows percentage of RNA-seq read coverage across each gene body (5' to 3'). Mann-Whitney U test was used to test for significance. (**D**) Median number of pass filter aligned bases (via Picard tools) (y-axis) and (**E**) median number of RNA editing sites per sample (y-axis) by living and PM DLPFC (x-axes). Two-sided linear regression was used to test for significance. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**F**) Association between the number of RNA editing sites detected (y-axis) and the number of pass filter aligned bases (x-axis). A Pearson's correlation coefficient was used to test the association.



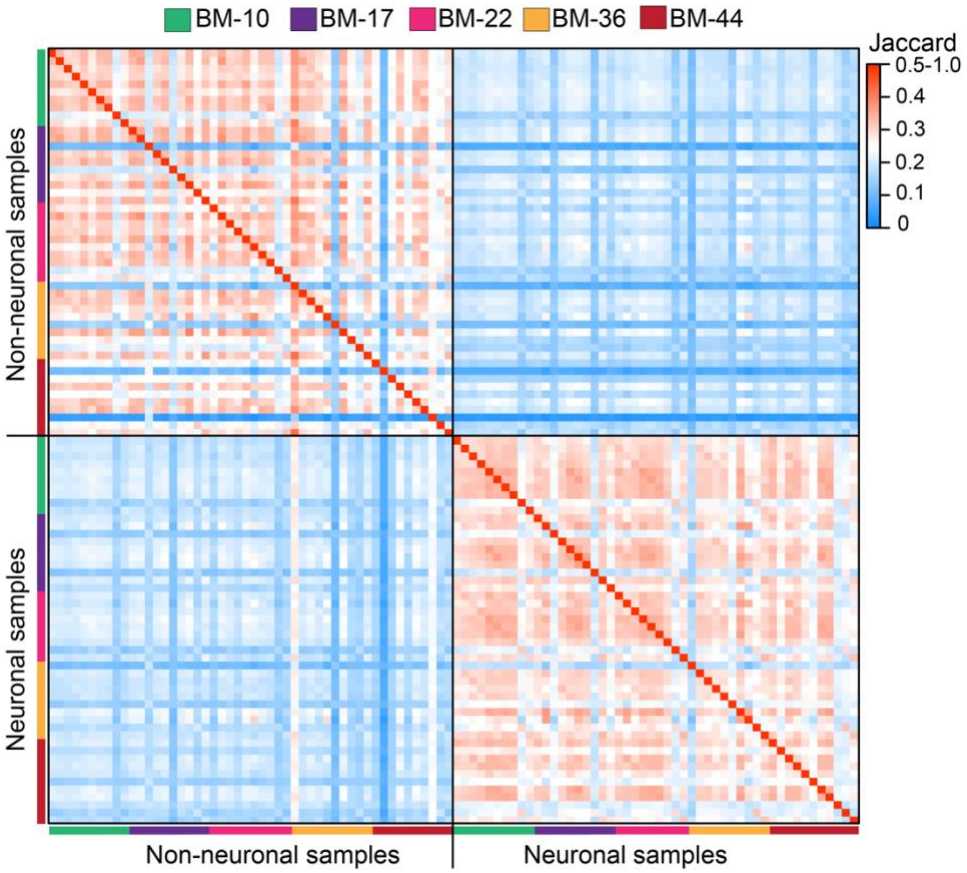
Supplemental Figure 7. Differential testing between living and postmortem tissues. Differential proportion testing: (**A**) The total number of A-to-I sites that are unique to either living or postmortem (PM) DLPFC (y-axis) according to their population-level detection (x-axis). For example, more than 16,824 A-to-I sites are detected in at least 40-60% of PM DLPFC and are not detected in living DLPFC. Most of these sites are intronic. (**B**) The expression patterns of genes harboring sites that are uniquely detected in PM DLPFC, indicate that overexpression of gene expression in PM tissues does not explain this enrichment. Differential editing analysis: (**C**) A collection of 42,493 differentially edited sites (FDR <5%) between living and postmortem DLPFC were used to plot the absolute delta values according to each genic region. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**D**) MA-plot displaying delta editing levels (y-axis) against average editing levels (%; x-axis) across each transcriptomic sample. The red dotted line indicates 0% change in editing levels between living and postmortem tissues.



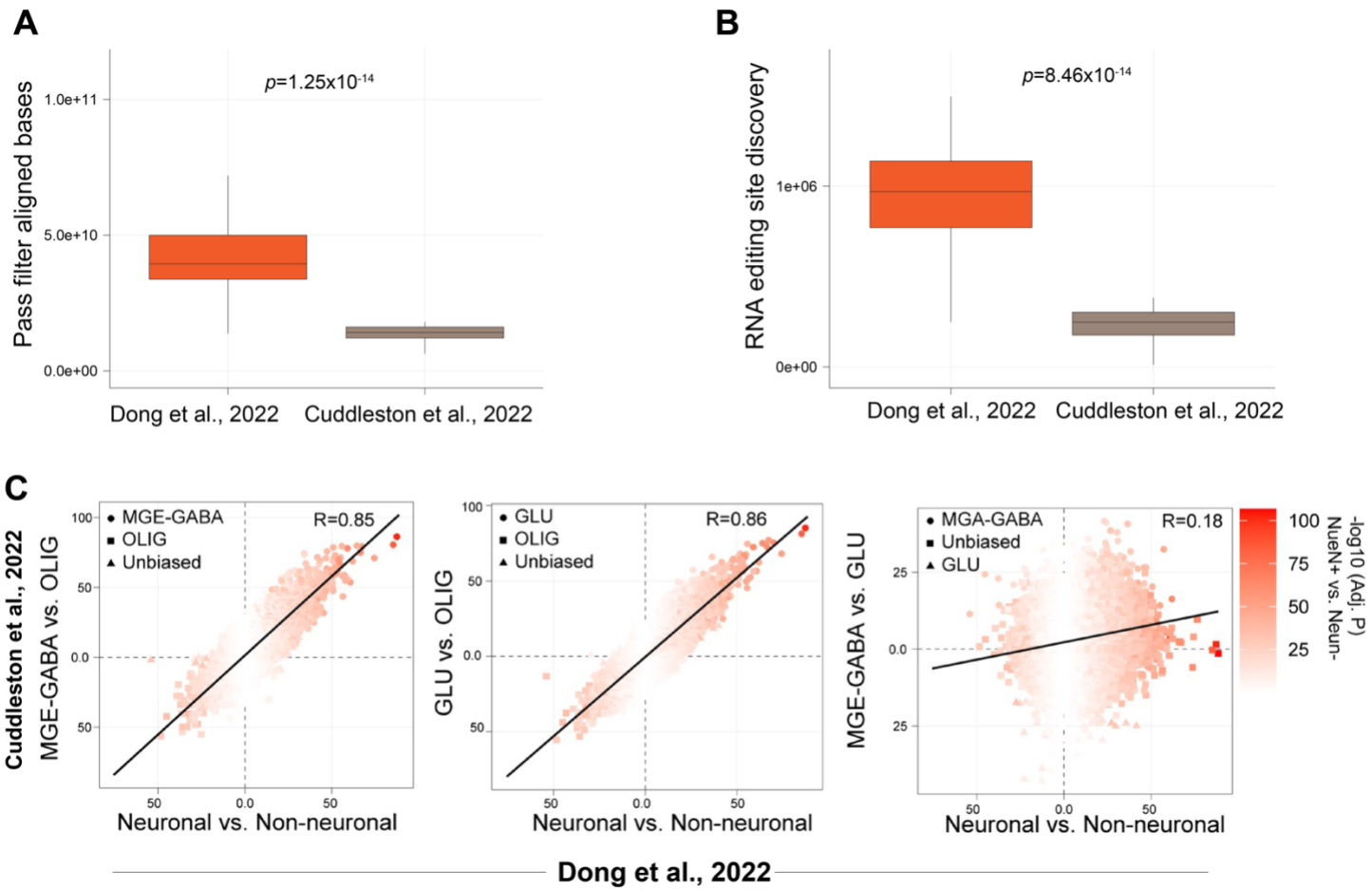
Supplemental Figure 8. Querying LIV-PM sites in snRNA-seq cellular pools. (**A**) We queried 72,356 A-to-I sites that were either differentially edited or differentially detected in either living or postmortem bulk RNA-seq DLPFC (y-axis) in snRNA-seq pseudo-bulk cellular pools (x-axis). The mean number of detected sites (with standard error bars) per pseudo-bulk cellular pool for each living and postmortem DLPFC are displayed. (**B**) The total number of A-to-I sites detected (out of 72,356) per pseudo-bulk cellular pool for sample (y-axes) as a function of library size (top, x-axis) and number of cells per pseudo-bulk cellular pool (x-axis, bottom). A Pearson's correlation coefficient was used to test the association.



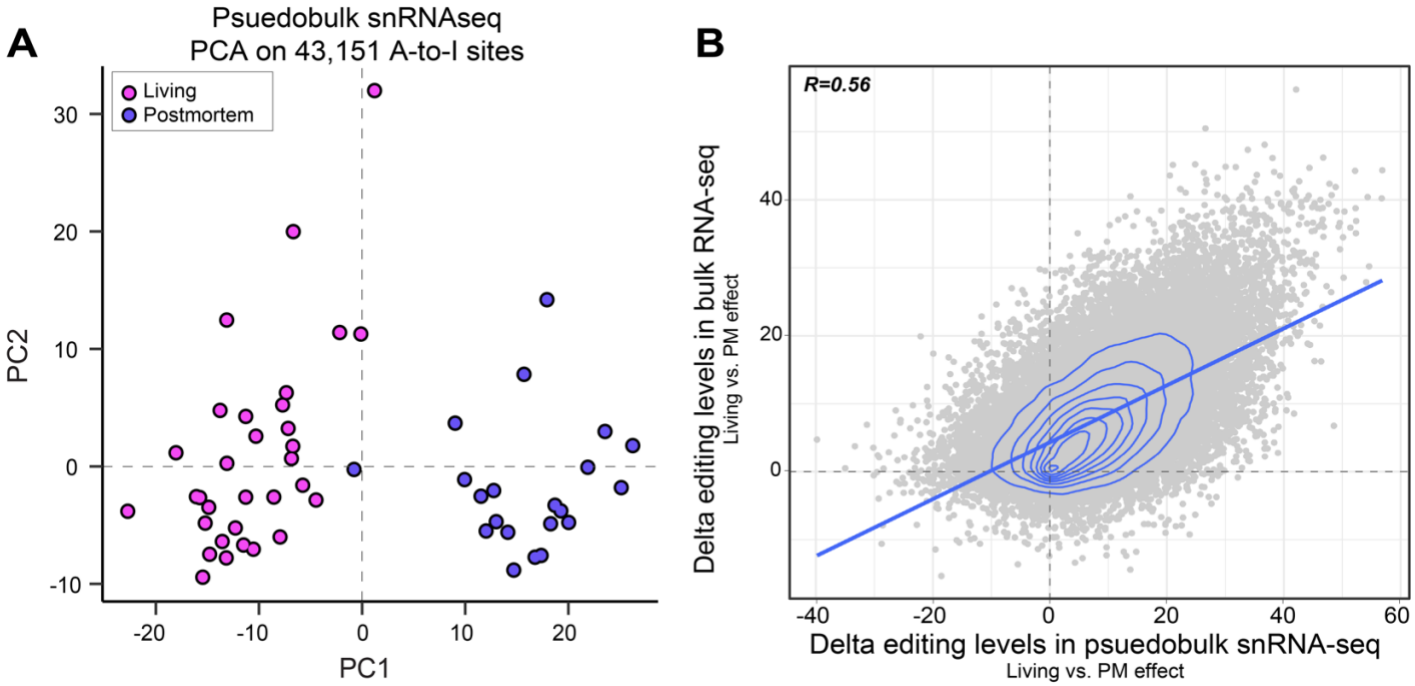
Supplemental Figure 9. Neuronal and non-neuronal A-to-I editing across five postmortem cortical areas. **(A)** Experimental design to generate deeply sequenced neuronal and non-neuronal nuclei across five cortical areas from ten biological replicates. (**B**) The *Alu* editing index (AEI) and (**C**) *ADAR*, *ADARB1* and *ADARB2* normalized expression within each cortical region between neuronal and non-neuronal nuclei. Two-sided linear regression was used to test for significance. Significance was tested between neuronal and non-neuronal cell types. No regional differences were observed. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers). (**D**) A linear mixed model quantified the fraction of AEI variance explained by eight known factors. (**E**) Mean number of RNA editing sites detected per sample according to different substitution types for neuronal (top) and non-neuronal nuclei (bottom). Standard error bars depict sample-level variation. (**F**) Principal component analysis of editing levels for 290,495 sites detected across neuronal and non-neuronal nuclei from all regions separates by cell type (PC1, x-axis) and donor as a repeated measure (PC2, y-axis). (**G**) Differential editing analysis compares the delta editing rates (%; x-axis) and the strength of significance (-log$_{10}$ adjusted p-value; y-axis) between neuronal and non-neuronal cell types. Neuronal biased sites are pink and non-neuronal sites are blue, and shapes represent unique genic regions. (**H**) Percentage of differential edited sites explained (y-axis) after covarying for different combination of ADAR enzymes as continuous measures (x-axis). (**I**) Percentage of differentially edited sites (y-axis) that map to different genic regions (x-axis). (**J**) Recoding sites that are either neuronal or non-neuronally biased ranked by their delta editing rate (%; y-axis) between cell populations. phastCons for each site and pLI for each gene were calculated and colored on a scale from 0-1.



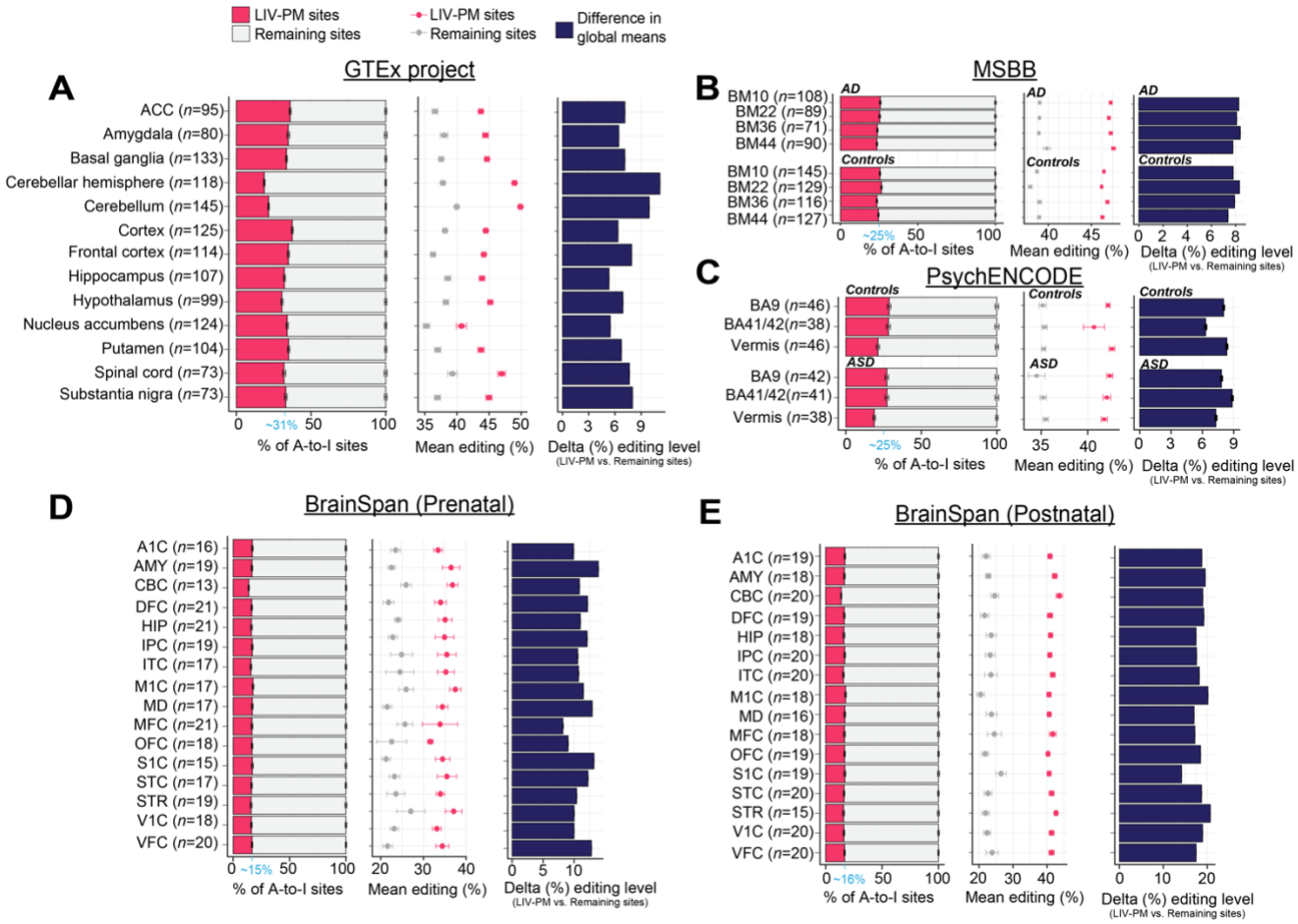
Supplemental Figure 10. Jaccard overlap matrix of A-to-I sites per donor. Pairwise overlaps for A-to-I sites detected for neuronal (NeuN+) and non-neuronal (NeuN-) cell populations across five cortical regions (n=10 biological replicates). Overlaps were measured using Jaccard Index and site detection converged mainly within either neuronal or non-neuronal nuclei, with little to no regional effect.



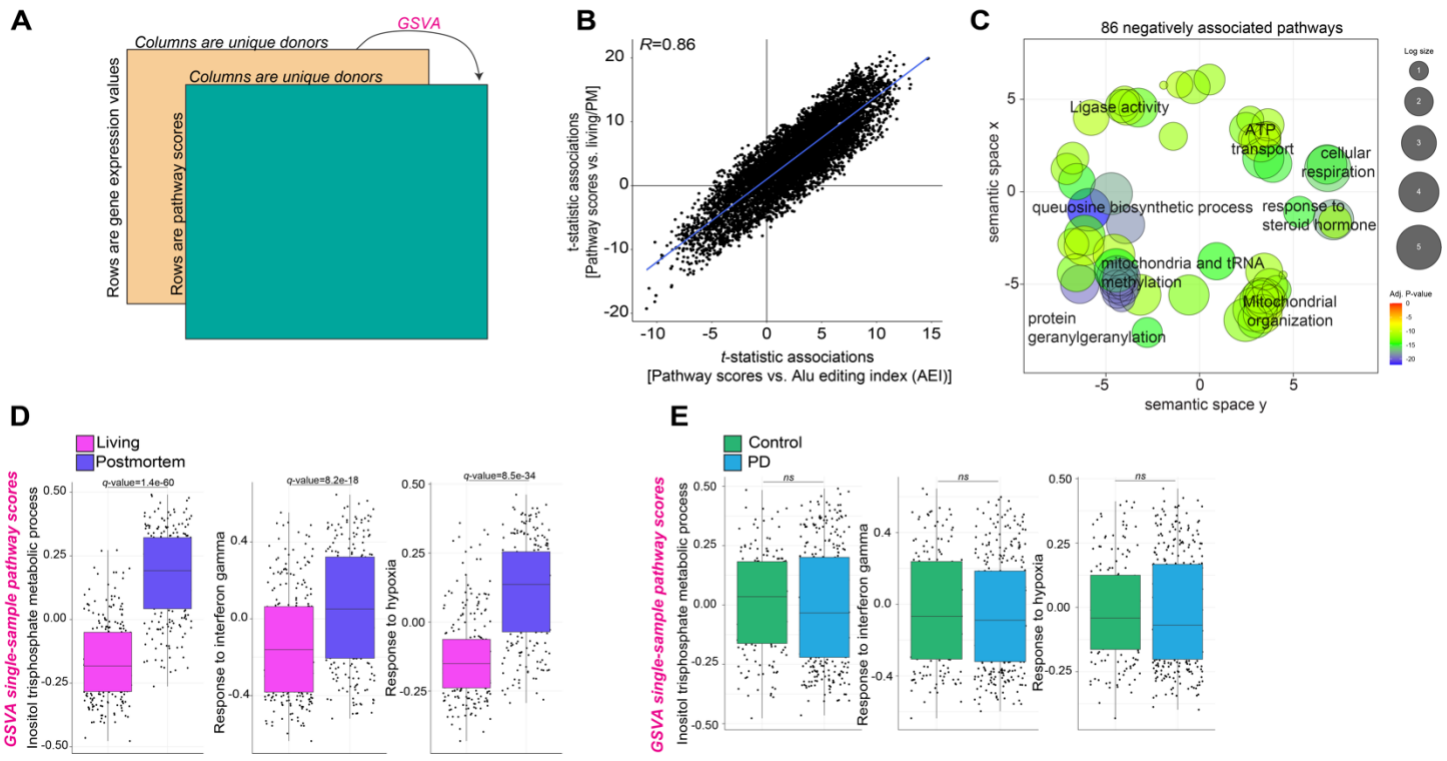
Supplemental Figure 11. Validating cell-specific A-to-I sites. To validate the specificity of our findings, we leveraged an independent RNA-seq study that performed fluorescence activated nuclei sorting (FANS) of MGE-GABAergic neurons (MGE-GABA), glutamatergic neurons (GLU) and oligodendrocytes (OLIG) from the human prefrontal cortex (*n*=9 biological replicates) (PMID: **35637184, Cuddleston et al., 2022**). The FANS data in the current data set exhibits (**A**) significantly deeper sequencing depth and (**B**) significantly more A-to-I site discovery. Significance was tested with a Mann Whitney U test. (**C**) A total of 48,921 cell-specific sites were commonly detected between our data set and this previous study. Each site illustrated strong hallmark signatures to be either neuronal or non-neuronal and validated across studies. The neuronal vs. non-neuronal effect was compared with the MGE-GABA vs. OLIG effect (left), the GLU vs. OLIG effect (middle) and the MGE-GABA vs. GLU effect (right). All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers).



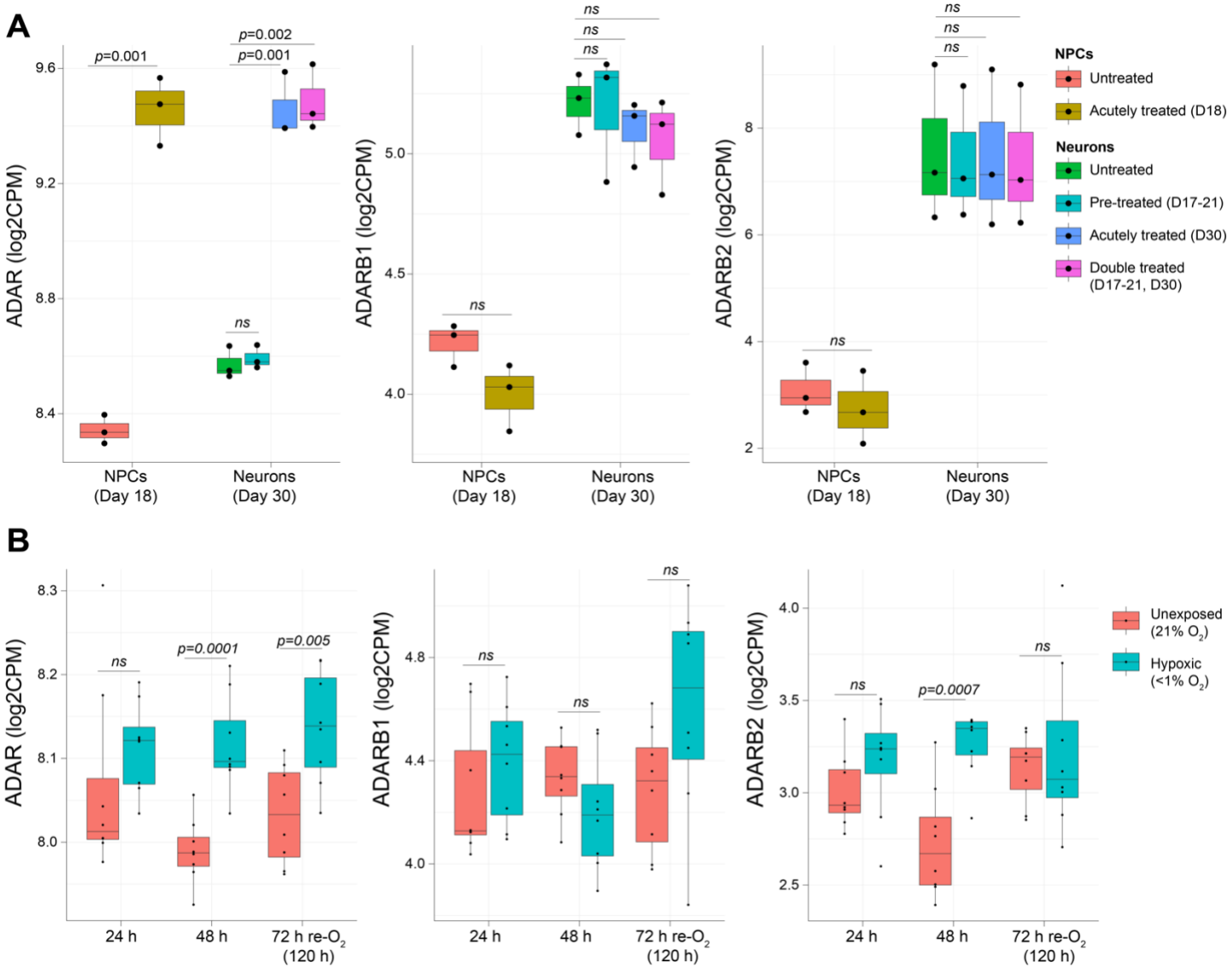
Supplemental Figure 12. Replicating differentially edited sites using pseudo-bulk snRNA-seq. (A) A total of 54,826 A-to-I sites detected across all bulk tissue samples were queried in pseudo-bulk snRNA-seq in an independent cohort of 21 living and 22 postmortem DLPFC samples. Of these, we detected 43,151 of these sites (), which were subjected to downstream analysis. (**A**) Principal component analysis (PCA) clearly separated living and postmortem DLPFC based on editing levels for these 43,151 sites along PC1. (**B**) Differential editing analysis was conducted (in an identical fashion as applied to bulk RNA-seq) to the pseudo-bulk profiles. To determine the degree of replication, the delta editing values obtained between living and postmortem DLPFC (x-axis) were regressed onto those derived from the bulk RNA-seq comparison (y-axis). This analysis demonstrated a positive concordance between the postmortem-induced effects of editing on these sites across the two data types and cohorts.



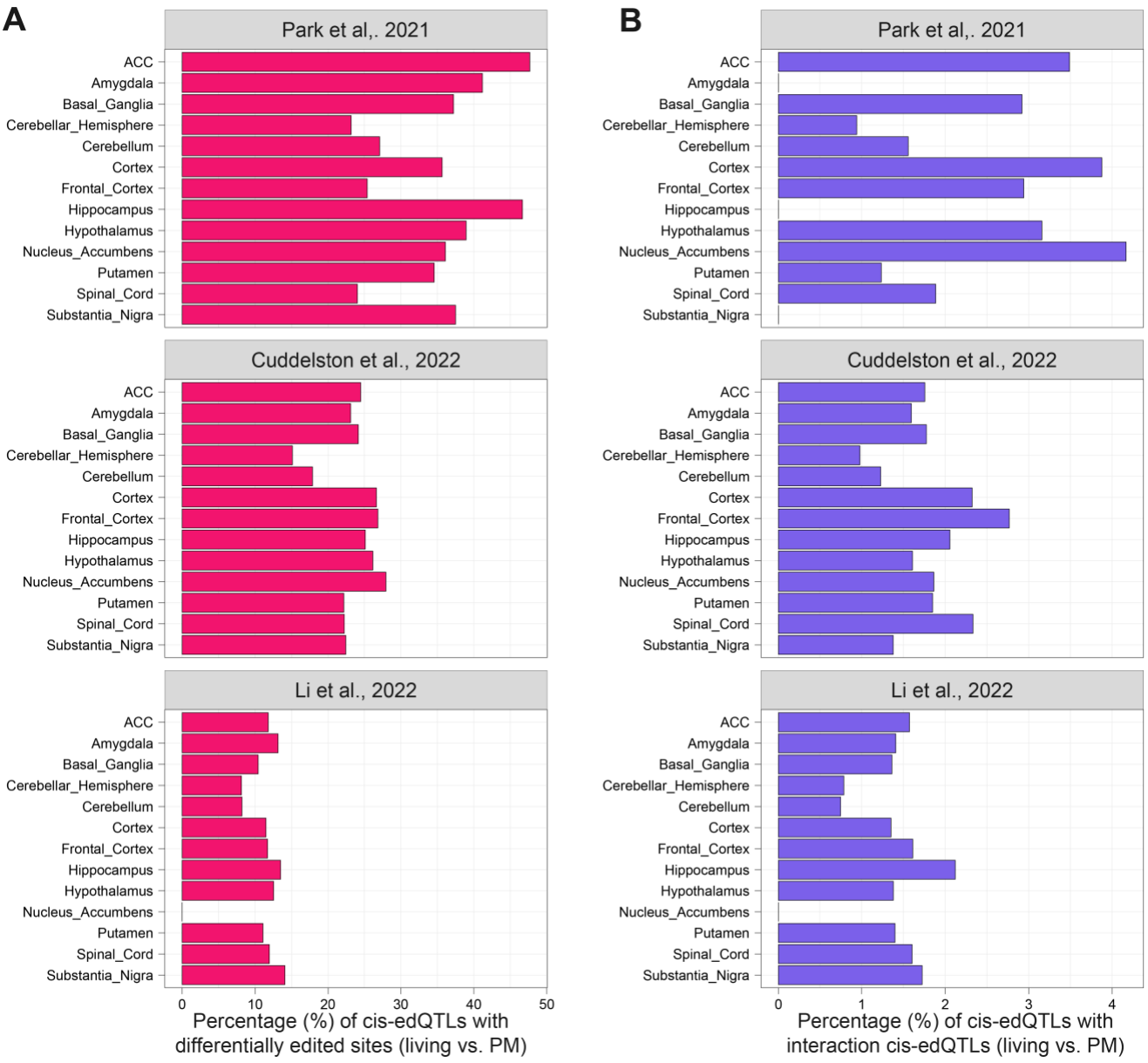
Supplemental Figure 13. Quantifying postmortem effects in independent postmortem brain transcriptome data sets. Four datasets were quality controlled and known A-to-I sites were quantified using a supervised approach requiring at least 5 reads and 2 edited reads (*see Materials and Methods*). For each dataset, we queried the fraction of sites catalogued as LIV-PM per sample (left), the global mean editing levels (%) for the LIV-PM sites relative to the remaining sites detected (middle) and the difference in global mean editing levels (LIV-PM site editing levels subtracted from the editing levels from the remaining sites (right)). These analyses were performed on **(A)** 13 anatomical regions from GTEx, (**B**) 4 anatomical regions from the Mount Sinai Brain Bank (MSBB), (**C**) 3 anatomical regions from PsychENCODE, and 16 anatomical regions from BrainSpan, covering (**D**) prenatal and (**E**) postnatal stages of development. MSBB samples included individuals with a clinical diagnosis of Alzheimer's disease (AD; CERAD scores indicative of definitely AD). PsychENCODE samples included individuals with a clinical diagnosis of autism spectrum disorder (ASD).



Supplemental Figure 14. Single-sample gene-set variation analysis. (A) The GSVA framework provides a summary measure of gene expression profiles underlying discrete biological processes by transforming an input gene-by-sample expression data matrix into a corresponding gene-set-by-sample expression data matrix. A total of 10,493 biological processes were generated for each living and postmortem bulk RNA-seq sample using the GSVA method. (**B**) Pathway activation scores were regressed onto the *Alu* editing index (AEI; x-axis) as well as differences between living and postmortem DLPFC (y-axis). (**C**) REVIGO semantic similarity analysis of the 86 pathways that negatively predict AEI. Examples of GSVA single-sample pathway scores (y-axes) that stratify (**D**) living versus postmortem tissues (x-axes) and (**E**) controls versus individuals with Parkinson's disease (PD). *Q*-values are multiple test corrected measures of significance (*see* ***Supplemental Data 2*** *for full summary statistics*).



Supplemental Figure 15. ADAR mRNA expression profiles. VOOM normalized expression data were used to plot normalized expression of *ADAR* (left), *ADARB1* (center) and *ADARB2* (right) across two in vitro models: **(A)** an in vitro model of hiPSC-derived neural progenitor cells (NPCs) and mature neurons treated with interferon-gamma; and (**B**) human cortical spheroids (hCS) exposed to hypoxic conditions (<1% O_2) and after 72 hours of reoxygenation versus unexposed (21% O_2). A two-sided student's *t*-test was used to test for significance. All boxplots in this figure show the medians (horizontal lines), upper and lower quartiles (inner box edges), and 1.5× the interquartile range (whiskers).



Supplemental Figure 16. Annotation of independent cis-edQTLs from GTEX. Three independent studies have quantified cis-edQTLs across 13 different GTEx brain regions. Here we show the percentage of each previously described set of eSites (*i.e.* A-to-I sites with at least one FDR significant cis-edQTL) that are either (**A**) differentially edited between living and postmortem DLPFC or (**B**) cis-edQTLs with a significant interaction effect between living and postmortem DLPFC. Ultimately, anywhere between 16-50% of previously described cis-edQTLs are associated with a differentially edited site between living and postmortem DLPFC and ~2% of all cis-edQTLs show interaction effects.