Lifespan analysis of repeat expression reveals age-dependent upregulation of HERV-K in the neurotypical human brain

- Taylor A. Evans ^{1,3}, Arthur S. Feltrin ¹, Kynon J. Benjamin ^{1,3,4}, Tarun Katipalli ¹, Thomas Hyde ¹⁻³, Joel E.
 Kleinman ^{1,4}, Daniel R. Weinberger ¹⁻⁵, Apua C. Paguola ¹⁻³, Jennifer A. Erwin ^{* 1-3}
- 5 1. Lieber Institute for Brain Development, Baltimore, MD, United States of America
- Johns Hopkins University School of Medicine, Department of Neuroscience, Baltimore, MD, United
 States of America
- Johns Hopkins University School of Medicine, Department of Neurology, Baltimore, MD, United
 States of America
- Johns Hopkins University School of Medicine, Department of Psychiatry & Behavioral Sciences,
 Baltimore, MD, United States of America
- 12 5. Johns Hopkins University School of Medicine, McKusick-Nathans Department of Genetic Medicine,
- 13 Baltimore, MD, United States of America
- 14 *Corresponding Author

15 Abstract

16 DNA repetitive sequences (or repeats) comprise over 50% of the human genome and have a 17 crucial regulatory role, specifically regulating transcription machinery. The human brain is the 18 tissue with the highest detectable repeat expression and dysregulations on the repeat activity are 19 related to several neurological and neurodegenerative disorders, as repeat-derived products can 20 stimulate a pro-inflammatory response. Even so, it is unclear how repeat expression acts on the 21 aging neurotypical brain. Here, we leverage a large postmortem transcriptome cohort spanning 22 the human lifespan to assess global repeat expression in the neurotypical brain. We identified 23 21,696 differentially expressed repeats (DERs) that varied across seven age bins (Prenatal; 0-24 15; 16-29; 30-39; 40-49; 50-59; 60+) across the caudate nucleus (n=271), dorsolateral prefrontal 25 cortex (n=304), and hippocampus (n=310). Interestingly, we found that long interspersed nuclear 26 elements and long terminal repeats (LTRs) DERs were the most abundant repeat families when 27 comparing infants to early adolescence (0-15) with older adults (60+). Of these differentially 28 regulated LTRs, we identified 17 shared across all brain regions, including increased expression 29 of HERV-K-int in older adult brains (60+). Co-expression analysis from each of the three brain 30 regions also showed repeats from the HERV subfamily were intramodular hubs in its 31 subnetworks. While we do not observe a strong global relationship between repeat expression 32 and age, we identified HERV-K as a repeat signature associated with the aging neurotypical brain. 33 Our study is the first global assessment of repeat expression in the neurotypical brain.

34

35 Keywords

36 HERV-K, repeat biology, retroelements, aging, neurodegeneration, RNA-seq, Co-Expression

37

38 Introduction

Present in multiple copies, repetitive sequences, herein referred to as repeats, are a broad category 39 40 of DNA sequences known to play crucial homeostatic roles within organisms including plants, 41 insects, and humans through evolution (Ding et al., 2017; Petersen et al., 2019). In humans, repeats 42 comprise over 50% of the human genome and have evolved to contribute to cell biology through 43 more than insertional mutagenesis. However, following their discovery and causative link to 44 hemophilia A, repeats have primarily been associated with their mutagenic capacity and disease 45 (Gorbunova et al., 2021). A subcategory of repeats, retroelements, can move through an RNA 46 intermediate, yet most of these elements are immobilized. Even when immobilized, repeats can 47 contribute to the transcriptome of a cell, tissue, and organ (Schrader & Schmitz, 2018; Yamamoto 48 et al., 2022).

49

50 Expression of repeat sequences poses a threat to genomic stability but is particularly relevant in 51 the healthy human brain as it is the tissue with the highest detectable repeat expression (Bogu et 52 al., 2019). Given the functional consequences of repeat expression in both healthy and diseased 53 tissue, repeat sequences are regulated at the genomic, transcriptional, and post-transcriptional 54 levels. Among the many mechanisms that human cells employ to silence repeat expression, 55 maintenance of epigenetic marks on DNA and histones is crucial to healthy brain development.

56

57 Dysregulation of repeat-derived RNAs and proteins has been reported in neurological diseases 58 including neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and 59 multiple sclerosis (MS) (Evans & Erwin, 2021). A series of studies investigating the role of repeat-

60 derived products in MS found that human endogenous retroviruses, also known as HERVs, are 61 increased in several patient samples and are associated with disease status (Macías-Redondo et 62 al., 2021).

Our current understanding of repeat expression and age relies on limited data from non-human model systems, peripheral tissues, and postmortem tissue from limited developmental timepoints (Bogu et al., 2019; de Cecco et al., 2013; He et al., 2021; Lee et al., 2012; Li et al., 2013; Schmitt et al., 2013). As a result, the field has struggled to contextualize changes to repeat expression in neurological disease states without a comprehensive understanding of repeat expression across age.

69

70 This same hypothesis has been applied to aging, termed sterile inflammation, and suggests repeat-71 derived RNAs contribute to a positive feedback loop of functional and biochemical decline with age 72 (Dumetier et al., 2022; López-Otín et al., 2013). Dysregulation of epigenetic machinery is a hallmark 73 of aging, and one theory suggests aging induces a global relaxation of heterochromatin. Thus, 74 remodeling of the epigenetic landscape confers transcriptomic changes and a global de-repression 75 of repeats (LaRocca et al., 2020). The link between repeat expression and aging is evidenced by 76 several observations. Patients with progeroid syndromes, genetic diseases that mimic physiological 77 aging, exhibit increased LINE-1 repeat expression (Della Valle et al., 2022). Furthermore, 78 prematurely aged mice, exhibiting hallmarks of aged chromatin, showed increased LINE-1 (long 79 interspersed nuclear elements) repeat expression (Simon et al., 2019). Upon treatment with reverse 80 transcriptase inhibitors, compounds that target LINE-1 RNA products from being reverse 81 transcribed, mice showed increased health and longevity. Together, these findings suggest repeat 82 derepression contributes to aging phenotypes.

83

These observations converge in Alzheimer's Disease, a spontaneous/idiopathic neurodegenerative disease where disease risk is associated with age (Guerreiro & Bras, 2015). Tau neurofibrillary

tangles, a neuropathological hallmark of AD, has been implicated as causative in AD though the
mechanism is unclear. Studies report that tau promotes neurodegeneration through chromatin
relaxation and thus, activates repeat expression in human postmortem brains (Frost et al., 2014;
Guo et al., 2018). It is still unclear how repeat expression contributes to disease etiology, but it is
possible repeat expression is a feature of aging that is exacerbated by disease.

91

92 Our study leverages a large postmortem transcriptome cohort spanning the human lifespan to 93 assess repeat expression in three neurotypical brain regions - caudate nucleus (n=271), 94 dorsolateral prefrontal cortex (DLPFC; n=304), and hippocampus (n=310). We pay particular 95 attention to the expression of repeats associated with human health and disease, specifically LINEs, 96 long terminal repeats (LTRs), short interspersed nuclear elements (SINEs), and satellite repeats. 97 By employing co-expression and differential expression analyses across the three brain regions as 98 a function of age, we glean valuable insights into the region-specific dynamics of repeat expression 99 during aging and its links to age-related neurological disorders.

- 100
- 101
- 102 Results
- 103

104 Selection of repeat quantification method

Here, we aimed to quantify repeat-derived RNAs from 885 postmortem brain samples (n=291 female, n=594 male; **Supplemental Figure 1A**). To build this resource of repeat-derived RNA expression across three brain regions (caudate nucleus [n=271], DLPFC [n=304], and hippocampus [n=310]) (Benjamin et al., 2022; Jaffe et al., 2018; Schubert et al., 2015), we first tested two distinct methods designed to quantify repeat expression, TEcount (Jin et al., 2015) and featureCounts (Liao et al., 2014). Both featureCounts and TEcount quantify genomic features (genes and repeats) and share 2,336,733 unique repeat sequences (**Supplemental Figure 1B**).

112 However, when considering all quantifiable repeats, we observe differences between the 113 methods' potential benefits as featureCounts can quantify genomic features on the negative 114 strand and when considering all repeats, we observed substantial differences between the two 115 methods with about 50% of the repeats unique to each method (Supplemental Figure 1C). 116 TEcount method is uniquely designed to quantify transposable elements, as reflected in the 117 composition of its annotation file (Supplemental Figure 1D). For featureCounts, we generated a 118 custom annotation file similar to TEcounts, using RepeatMasker but allowing for strand-specific 119 information and additional satellite repeat annotations (Supplemental Figure 1E).

120

Despite this, the methods show high Pearsons's correlation, thus perform similarly, when quantifying evolutionarily young repeats such as L1HS (R²=0.98) and SVA F (R²=0.89) (**Supplemental Figure 1F-G**). Due to its ability to quantify simple and satellite repeats, including those previously associated with Huntington's disease, amyotrophic lateral sclerosis, and frontotemporal dementia (Kmetzsch et al., 2020), featureCounts was selected for repeat quantification and downstream analyses. The authors welcome questions to help generate a custom annotation file.

128

129 Quantification of repeat expression across brain age

130 To generate repeat counts with featureCounts from 885 brain samples spanning prenatal - 90 131 years across the caudate nucleus (n=271), DLPFC (n=304), and hippocampus (n=310) (Figure 132 **1A**, **1B**), we used a custom annotation file (GTF) generated from University of California Santa 133 Cruz's RepeatMasker containing stranded information about genes and repeats. Using this 134 custom GTF, featureCounts randomly assigns multi-mapping reads from aligned .bam files to a 135 corresponding genomic feature (Figure 1C). Subsequently, quality control was performed, and an expression matrix was generated containing 268 caudate samples, 287 DLPFC, and 306 136 137 hippocampus samples (Supplemental Table S1).

1	3	8
---	---	---

This expression matrix contained quantification of repeat sequences across all samples and was the basis for downstream analyses including differential expression and Weighted Gene Correlation Network Analysis (WGCNA; **Figure 1D**) (Langfelder & Horvath, 2008). Given the high sequence similarity and multiple copies of repeats, along with the ambiguity of short reads, we acknowledge the likelihood for repetitive sequences to map to multiple locations in the genome. Therefore, downstream analyses investigate repeat expression at a superfamily or class level rather than the behavior of individual repeats at specific genomic loci/coordinates.

146

147 Figure 1. Characterizing repeat expression over lifespan of the neurotypical postmortem 148 brain. A) Neurotypical postmortem brain samples (n=855) included in study span from prenatal 149 to 91 years of age. B) Samples originate from caudate nucleus (n=271), DLPFC (n=304), and 150 hippocampus (n=310) and were binned according to age of death. C) Overview of repeat and 151 gene expression quantification utilizing featureCounts algorithm and custom annotated gene 152 transfer format (GTF) file with hg38 genome. D) Overview of downstream characterization for co-153 expression analysis (weighted gene correlation network analysis [WGCNA]) and differential 154 expression analysis.

155

156 Hippocampal repeat expression correlates with brain age

157

Age-related epigenetic modifications may lead to widespread activation of repetitive elements, with a positive correlation observed between total repeat expression and chronological age. To test this hypothesis, we correlated total repeat expression with ages (0-91 years) across brain regions. Here, we found a slight positive correlation between total repeat expression and age (Figure 2A; R=0.14, p=2.4e-5). Interestingly, we see a minor correlation between total repeat expression and age of death (**Figure 2A**; R=0.14, p=2.4e-5). When we stratify this correlation by

brain region, we observe the hippocampus has the highest correlation between total repeat expression and brain age (**Figure 2B**; R=0.17, p=0.0024; Spearman Rank Correlation). The caudate nucleus and DLPFC had negligible correlation. Taken together, this data suggests a relationship between age and repeat expression that is unique to neurotypical hippocampus.

168

169 Figure 2. Repeats cluster together by co-expression and genes associated with RNA 170 biology. A) Total repeat expression from uniquely mapped reads mildly correlates with age of 171 death (Spearman two-sided; rho=0.14, p=2.4e-5) B) When stratified by brain region, correlation 172 of total repeat expression with age of is driven by the hippocampus (Spearman two-sided; 173 rho=0.17, p=0.0024) C) Schematic representing how WGCNA utilizes expression matrix to place 174 repeats and genes into neighborhoods and generate co-expression networks. D) WGCNA module 175 composition across caudate nucleus, DLPFC, and hippocampus as a proportion of genes and 176 repeats. E) Heatmap of correlation between repeat-dense modules and clinical traits of caudate 177 nucleus, DLPFC, and hippocampus samples. Repeat-dense modules negatively correlate with 178 age of death across repeat-dense modules. Yet, repeat-dense modules positively correlate with 179 neuropathological scores in the caudate nucleus. F) Gene ontology enrichment for DLPFC Brown 180 module. G) Gene ontology enrichment for caudate nucleus, Light Cyan module.

181

182 **Repeats co-expression with RNA binding genes**

To understand the biological pathways associated with repeats and age across the neurotypical brain, we performed weighted gene co-expression network analysis (WGCNA) on both gene and repeat expression profiles derived from bulk RNA-seq data for each of the three brain regions. WGCNA utilizes the gene expression matrix generated from uniquely mapped reads to cluster expressed features (repeats or genes) into co-expression modules, referred to as modules in the remainder of the text, based on the Pearson correlation coefficient between a pair of features. Previously, this method has been applied to gene expression, miRNA expression and DNA

190 methylation (Euclydes et al., 2022; Langfelder & Horvath, 2008; Pascut et al., 2020), but here, we 191 apply WGCNA to a residualized expression matrix containing both repeat and gene expression. 192 To reduce artifacts introduced by evolutionarily young repeats, only uniquely mapped reads were 193 considered (Methods) when generating brain region networks using power $\beta = 14$ threshold 194 (Supplemental Table 2). Given the potentially region-specific relationship between repeat 195 expression and age, we applied WGCNA to organize 2336 unique repeats and 21986 unique 196 genes into modules by brain region to generate three regional co-expression networks (Figure 197 **2C**). The caudate nucleus, DLPFC, and hippocampus produced 16, 14, 13 modules, respectively. 198 Module size varies and the total number of features (repeats and genes) within a module is 199 indicated in parentheses and is detailed in (Supplemental Table 3).

200

To investigate the heterogeneity of each module, we visualized the proportion of repeats and genes. Modules composed of >50% of total features were classified as repeat-dense. The number of repeat-dense modules varied by brain region (**Figure 2D; Supplemental Table 3**). Interestingly, while it had the highest correlation between total repeat expression and age, the hippocampus had the fewest number (2) of repeat-dense modules compared to caudate nucleus (5) or DLPFC (3).

207

208 To evaluate if repeat-dense clusters correlate with age and others clinical features related with 209 senescence, we calculated the Pearson correlation coefficient between each WGCNA's module 210 eigengene (kME) and our clinical features. Sample characteristics selected include age, as well 211 as neuropathological neuritic plaque (CERAD) and neurofibrillary tangle (Braak) scores 212 (Supplemental Table 1). Both CERAD and Braak scales are neuropathological metrics of 213 neurodegeneration associated with Alzheimer's and Parkinson's Disease, respectively (Burke et 214 al., 2008; Fillenbaum et al., 2008). Repeat-dense modules had mild correlations of <[0.2] with age 215 of death, with DLPFC's Brown and Salmon modules having p-value=9e-04 and p-value=0.009.

respectively (**Figure 2E**). Both hippocampus' repeat-dense modules had a negative correlation with age of death at (R=-0.19, p=9e-04; Pearson correlation) and (R-0.15, p=0.009; Pearson correlation), respectively. In contrast, repeat-dense modules positively correlate (>0.2) with both CERAD and Braak scores in both the caudate nucleus and DLPFC (**Figure 2E**). In the caudate, the Light Cyan module correlates moderately with Braak score (R=0.28, p=3e-6; Pearson correlation) and is composed of 100% repeats. In the DLPFC, the Brown module correlates moderately with CERAD score (R=0.32, p=4e-8; Pearson correlation).

223

224 The DLPFC Brown module is composed of 72.71% repeats and 27.29% genes. We next 225 assessed the biological function of the genes contained in these repeat-dense modules. Thus, 226 we performed gene ontology (GO) enrichment analysis on all clusters to identify enrichment of 227 molecular function, biological processes, and cellular compartments with particular interest in GO 228 enrichment of repeat-dense clusters containing genes (Supplemental Table 4). GO analysis of 229 the DLPFC Brown module, caudate Light Cyan module, and caudate Yellow module revealed 230 enrichment of genes associated with RNA and protein binding, RNA processing, and molecular 231 functions associated with neurodegeneration including ubiquitin-protein transferase activity and 232 ubiquitin conjugating enzyme (Figure 2E, 2F; Supplemental Figure 3A). These results suggest 233 repeats correlate with clinical metrics of neurodegeneration.

234

Age-associated gene modules enriched for immune response and transcriptional regulation

Gene expression is known to change with age (de Magalhães et al., 2009), therefore, we also investigated GO enrichment of gene-dense modules that correlate with age of death. There were several gene-dense clusters that have a strong, positively correlated relationship with age of death including DLPFC's blue and Hippocampus' yellow module **(Supplemental Table 3, 4**). The DLPFC's Blue module (n= 3329 genes and 20 repeats) correlated negatively with age of death

(R=-0.40, p=2e-12) was enriched for genes associated with transcriptional regulation by RNA Pol
II (Supplemental Figure 3B). The hippocampus' Yellow module (n=1925 genes and 11 repeats)
correlated positively with age of death (R=0.55, p=2e-25) and was enriched for genes associated
with immune response, inflammatory response, and defense response to virus (Supplemental
Figure 3C). Together, these results support that gene-dense modules hold meaningful
information about brain age.

248

Figure 3. Repeats are differentially expressed between human brains 0-15 vs. 60+ years.
A) Total number of DERs across differential expression comparisons with brains >60 years. B)
Volcano plot visualizing 0-15 vs. 60y+ DERs by fold change and FDR. C) Most significant 0-15
vs. 60+ DER in caudate nucleus, L1P3b (FDR=2.17e-21, log2(fold change)=1.18). D) Most
significant 0-15 vs. 60+ DER in DLPFC, GSAT (FDR=9.15e-15, log2(fold change)=2.66). E) Most
significant 0-15 vs. 60+ DER in hippocampus, TTCATn (FDR=9.55e-7, log2(fold change)=1.42).
F) R-RHO plot depicting concordance between 0-15 vs. 60y+ DERs between brain regions.

256

Identification of differentially expressed repeats across lifespan of neurotypical brain: brain regions show distinct repeat family association with aging

We hypothesized a differential expression (DE) analysis, comparing repeat expression between over age, could identify differentially expressed repeats (DER) that are potential biomarkers of the neurotypical aging brain. To perform this analysis, we utilized samples binned by age (Prenatal; 0-15; 16-29; 30-39; 40-49; 50-59; 60+) where each age bin contained an adequate sample size with a minimum of 25 samples. (**Figure 1B, Supplemental Table 5**). Prenatal samples were only available for hippocampus and DLPFC.

265

Using counts generated with featureCounts (Methods), residualized expression was generated
using covariates via voom linear model. In this analysis, we controlled for the effect of biological

sex, self-reported race, ancestry (SNP PCs 1-3), and RNA quality (RIN, mitochondria mapping
rate, gene assignment rate, genome mapping rate, and hidden effects using surrogate variable
analysis) (Methods; Model 1, Differential Expression Analyses, Eq.5), with age as our variable
of interest. Criteria for a differentially expressed repeat (DER) was a false discovery rate (FDR) <
0.05. Differential expression identified 21696 DERs across all three brain regions representing
16.58% of all repeats analyzed (Supplemental Figure 4A, Supplemental Table 6). For contrast,
differential expression analysis from TEcount resulted in 13489 DERs (Supplemental Table 7).

Given age is a primary risk factor for many neurodegenerative diseases, and prenatal samples were not available for caudate nucleus, we further investigated DERs between 0-15 vs. 60+ years of age, identifying 1,401 instances between ages 0-15 and 60+ across all three brain regions (**Figure 3A**). When stratifying DERs by significance (FDR<0.05) and magnitude of change, we observed the majority of young (0-15 years) versus older (60+ years) DERs have a relatively small fold change (log2(fold change) < 0.5) across all three brain regions (**Figure 3B**).

282 Across the caudate nucleus, DLPFC, and hippocampus, the most significant DERs between 283 young (0-15 years) and older (60+ years) individuals belonged to distinct repeat families. In the 284 caudate nucleus, the most upregulated DER was the LINE repeat L1P3b (FDR=2.17e-21). The 285 DLPFC showed the greatest increase in centromeric GSAT repeats (FDR=9.15e-15), while the 286 hippocampus exhibited the strongest upregulation of the satellite repeat TTCATn (FDR=9.55e-7) 287 (Figure 3C-E). To explore consistency in DER patterns across brain regions, we employed Rank-288 Rank Hypergeometric Overlap (RRHO) analysis (Plaisier et al., 2010). This analysis revealed the 289 strongest concordance in the direction of change (up or downregulation) between DERs in the 290 caudate nucleus and DLPFC (Figure 3F). Concordance was weaker between the DLPFC and 291 hippocampus, and the hippocampus and caudate nucleus. All three brain regions exhibited an 292 increase in expression from young (0-15 years) to older (60+ years) groups (Supplemental 293 Figure 4B).

294

295 Age-related repeat expression is brain region specific

296 Discussions of repeat expression are often centered around increases in repeat expression, as it 297 poses threat to genomic stability and cellular homeostasis. In our analysis, we unbiasedly 298 captured DERs with FDR<0.05, regardless of directionality of change. To test whether repeat 299 expression ubiquitously increases with age, we stratified 0-15 vs. 60+ DERs by the sign (positive 300 or negative) of log2(fold change) (Supplemental Table 6). We identified 549 DERs (403 301 upregulated, 146 downregulated) in caudate nucleus, 233 DERs (92 upregulated, 233 302 downregulated) in DLPFC, and 629 DERs (25 upregulated, 629 downregulated) in hippocampus 303 (Supplemental Figure 5A, 5B). While 73.40% of caudate nucleus DERs are upregulated in 304 brains >60 years, 96.02% hippocampus DERs are downregulated in brains >60 years.

305

306 To test if the observations were representative across multiple categories of repeats, we 307 investigated the behavior of LINEs, LTRs, SINEs, and satellite repeats present within 0-15 v. 60+ 308 DERs (Figure 4A-D). LINEs and LTRs were the most abundant categories present within 0-15 309 vs. 60y+ DERs and shared similar patterns of upregulation in caudate nucleus >60y and 310 downregulation in hippocampus >60 years (Figure 4A, 4B; Supplemental Table S6). 311 Interestingly, despite moderate correlation between total repeat expression and age in the 312 hippocampus (R=0.17; Figure 2B), hippocampus DERs are primarily downregulated in brains 313 between 0-15 and 60+ years. This suggests a small subset of non-differentially expressed repeats 314 may drive global correlation of repeat expression-age and are not representative of more nuanced 315 changes in repeat expression with age.

316

Figure 4. LTRs are over-represented in differential expression results. A-D) Total number of
0-15 vs. 60+ DERs downregulated (red) vs upregulated (blue), stratified by LINE, SINE, LTR, and
satellite repeats E-H) Over representation analysis of LINE, SINE, LTR, and satellite repeats

within DER results across all comparisons and brain regions (DER = differentially expressed
 repeat, LINE = Long Interspersed Element, LTR = Long Terminal Repeat, SINE=Short
 Interspersed Repeat)

323

324 <u>LTRs are over-represented within differential expression results</u>

325 Given the unique behavior of differentially expressed repeats across brain regions, we wanted to 326 see if any repeat classes were more likely to be differentially expressed. To test if LINEs, LTRs, 327 SINEs, or satellite repeats were over-represented within our differential expression results, we 328 performed the Hypergeometric distribution test, implemented by the SuperExactTest package. This 329 methodology calculates the significance of the intersection between two sets of elements (repeats), 330 considering a fixed background population (all unique annotated repeats, **Methods**). With this 331 approach, we can assess if a specific differentially expressed repeat is overrepresented within its 332 own repeat super family, compared to what would be expected by chance, considering all repeats 333 annotated in our custom .gtf file (Methods). Results from each individual comparison performed 334 across all three brain regions are shown in Figure 4E-H. We quickly identified a striking over-335 representation of LTR elements within our differential expression results including within 0-15 vs. 336 60y comparison across all three brain regions (Figure 4F).

337

Figure 5. HERV-K expression increases from 0-15 vs. 60+ years in all three brain regions. A) Venn diagram of 0-15 vs. 60+ DERs across brain regions B) Distribution of repeat classes within shared 0-15 vs. 60+ DERs across all three tissues C) List of 17 LTR 0-15 vs. 60+ DERs shared across all three tissues D) Correlation between total expression of HERV-K-int and age of death by brain region E) Expression of HERV-K-int across age bins, by brain region.

343

344 Age contributes to differential expression of HERV-K in brain

345 Given LTRs are abundant and overrepresented within 0-15 vs 60+ DERs across all three brain regions, we then wanted to evaluate if any differentially expressed LTRs were shared between the 346 347 caudate nucleus, DLPFC, and hippocampus. Out of the 1401 0-15 vs 60+ DERs identified, 36 DERs 348 were shared between all three tissues (Figure 5A). Of these, 17 were LTRs and 10 were satellite 349 repeats (Figure 5B). Further investigation into 17 shared LTRs yielded identification of human 350 endogenous retrovirus-K-int (HERV-K-internal sequences), a human-specific LTR, as a significant 351 0-15 vs 60y+ DER across caudate nucleus (FDR=0.011;), DLPFC (FDR=0.043), and hippocampus 352 (FDR=0.00026) (Figure 5C).

353 Previously, HERVs have been identified as being dysregulated in neurological disorders including 354 multiple sclerosis (Brudek et al., 2009; Laufer et al., 2009; Schmitt et al., 2013). Age is a risk factor 355 for multiple sclerosis, thus, we asked if total LTR, HERV, or HERV-K expression correlated with 356 brain age in our samples. We observed that total LTR expression has the highest Spearman Rank 357 Correlation (R=0.077, p=0.024; test) out of all repeat categories analyzed (Supplemental Figure 358 6A). When we look at total HERV-K expression, a subcategory of LTR elements, we see HERV 359 only mildly expression correlates with brain age (R=0.033, p=0.0024; test) and when plotted by brain region, is similar across caudate nucleus (R=0.068, p=0.00045), DLPFC (R=0.051, 360 361 p=0.0059), and hippocampus (R=0.044, p=0.014) Supplemental Figure 6B, 6C). Importantly, 362 when plotting total HERV-K expression, we observe a strong and significant positive correlation 363 with brain age across all three brain regions. Total expression of HERV-K has the highest 364 correlation with brain age in the DLPFC (Figure 5D, R=0.4, p=1.8e-12). Boxplots further confirm 365 HERV-K expression increases with age in the neurotypical caudate nucleus, DLPFC, and 366 hippocampus (Figure 5E).

367 Re-investigating the co-expression analysis, we observe HERV-K-int clusters into the DLPFC 368 Brown module containing 1718 repeats and 645 genes (Supplemental Table XX). Of the 1718 369 repeats, 61 (3.55%) are both human endogenous retrovirus and a highly connected feature (hub) 370 of the DLPFC Brown network. Of the 61 repeats that serve as hubs, 13 (21.3%) are from HERV-K subfamily, including HERV-K-int. All HERV-K hubs are all classified as intramodular hubs, hubs that 371 372 are highly connected within the DLPFC Brown module and drive architecture of network, as 373 indicated by a high kWithin and positive kDiff values (Supplemental Table S8, Bogenpohl et al., 374 2016). As previously mentioned, the DLPFC Brown module shows a significant positive correlation 375 with CERAD score (r=0.32, p<4e-08) (Figure 2E) indicating a potential relationship with HERV-K 376 expression and neuropathological protein aggregation.

377

GO analysis of genes in the DLPFC Brown module are enriched for molecular functions of protein and RNA binding (**Figure 2F**). Together, this data suggests expression of HERV-K-int correlates with brain age and shares co-expression patterns with several pathways critical to cellular homeostasis that have been previously implicated in the aging brain (Ham & Lee, 2020).

382

383 Discussion

In this study, we re-processed and re-quantified paired-end, stranded RNA-sequencing data from 885 neurotypical samples across the caudate nucleus, DLPFC, and hippocampus from 395 human, postmortem BrainSeq consortium donors (Benjamin et al., 2022; Collado-Torres et al., 2019) to build a repeat expression atlas of the aging human brain. Using co-expression networks, we placed repeat-derived RNAs within the brain's transcriptional network to discover biologically relevant relationships between repeats and genes.

390

391 The impact of genomic instability, epigenetic alterations, and altered cellular communication, all 392 hallmarks of aging, are not exclusive to genes and likely impact global expression (López-Otín et 393 al., 2013; Yamamoto et al., 2022). The interconnectedness between repeat and gene expression 394 has largely been understudied in the context of aging until recently. Through WGCNA, we 395 identified that repeats generally cluster together into repeat-dense modules, an expected result 396 given that repeats are the target of shared regulatory mechanisms at the transcriptional and post-397 transcriptional levels. Gene-dense modules appear to hold more information about brain age, an 398 expected result given previous studies on age-related expression changes in the brain. 399 Interestingly, repeat-dense modules correlate with CERAD and Braak scores, suggesting a 400 relationship between repeat expression and neuropathological hallmarks of disease, independent 401 of age (Yamamoto et al., 2022). For example, the repeats in Light Cyan module in caudate 402 consists exclusively of LINE-1 elements, sequences harboring conserved potential G-quadruplex 403 (G4) forming sequences in their 3' end which are associated with increased retrotransposition 404 (Sahakyan et al., 2017). Notably, LINE-1s have been associated with G4 formation in Alzheimer's 405 disease induced pluripotent stem cells derived neurons (Hanna et al., 2021), being an intragenic 406 feature reducing gene expression and potentially affecting the transcriptional programs.

407

We then went on to identify 21696 DERs (FDR<0.05) across the caudate nucleus, DLPFC, and hippocampus with the DLPFC containing the fewest DERs, mimicking differential expression results reported in Collado-Torres et al., 2019. We observed an overrepresentation of LTRs within our DER results and among our 0-15 vs. 60y+ DER results, we identified 17 LTRs that were shared across caudate nucleus, DLPFC, and hippocampus. Of these 17, we observed HERV-K increases with brain age and is upregulated in brains >60 years.

414

415 While we do not observe a strong global relationship between repeat expression and age, we 416 identified HERV-K as a repeat signature associated with the aging neurotypical brain. Not only is

417 HERV-K-int a shared DER across all three brain regions, but its total expression is moderately 418 correlated with brain age and stronger than the correlation of total expression HERV-K element 419 expression with brain age. A recent study supports the connection between endogenous 420 retrovirus expression and cellular senescence indicating HERV-derived proteins, including 421 HERV-K, can serve as a biomarker of tissue aging across lung, liver, and skin (36610399). Our 422 study expands upon this observation to confirm HERV-K RNA is a biomarker of aging across the 423 brain, broadening and strengthening HERV-K's position in the diagnostic and therapeutic 424 landscape of age-related neurodegeneration.

425

426 LTRs, more specifically HERVs, have also been associated with neurological disease (Dembny 427 et al., 2020; Macías-Redondo et al., 2021). HERV-derived RNA is capable of causing and 428 propagating neurodegeneration through Toll-like receptors (Dembny et al., 2020) and protein 429 aggregation (Liu et al., 2023). Beyond the production HERV RNA species, Turelli et al. observed 430 the regulatory impact of HERV-K also stems harboring transposable element-embedded 431 regulatory sequences (TEeRS) and subsequently altering KRAB-ZFP, a transcriptional repressor, 432 binding to neuronal genes (Turelli et al. 2020). We observe HERV-K-int, along with other HERV-433 K sequences, are intramodular hubs within our DLPFC and hippocampus co-expression networks 434 suggesting this LTR maintain and/or regulate biologically important relationships within a brain 435 region. Repeat-derived products, including repeat-derived RNAs, are not only capable of 436 propagating neurodegeneration through a pro-inflammatory response - thus contributing to 437 disease progression (Dembny et al., 2020); but also in some forms of cancer through the ubiquitin-438 proteasome pathway (Jin et al., 2019) and on aging phenotype (Gorbunova et al., 2021). Thus, 439 we also propose elevated HERV-K products in neurological disease may reflect a molecular 440 phenotype of accelerated aging that further drives transcriptional and proteomic hallmarks of 441 neurodegeneration.

442

443 Comprehensively, our work provides the largest global assessment of repeat expression across the aging neurotypical brain and refutes previous generalizations of repeat behavior. While 444 445 epigenetic alterations may change transcriptional landscape with age, we find repeat expression 446 shows high developmental and regional specificity making age only one important factor for 447 characterizing repeat behavior in a healthy aging tissue. We hope this global assessment will 448 serve as a resource to the greater scientific community.

449

450 Thus, as repeat expression becomes a popular target for biomarkers, diagnostics, and 451 therapeutics, our findings highlight the need to identify baseline expression dynamics of target 452 repeats in healthy tissues. As such, we anticipate this data will be used as a neurotypical baseline 453 for analyzing neurodevelopmental and neurodegenerative disease-related changes in repeat 454 expression.

- 455
- 456
- 457 Methods

458

459 Sample Selection

460 The LIBD BrainSeg Consortium consists of several brain regions and includes a wide range of 461 demographics and RNA-sequencing library preparation. We selected samples from the caudate 462 nucleus, DLPFC, and hippocampus based on three inclusion criteria: 1) Stranded RiboZero RNA-463 sequencing library preparation, 2) primary diagnosis of neurotypical control, and 3) self-reported 464 ancestry of either African American or European American. This resulted in a total of 395 unique 465 individuals for a total of 885 FASTQ files across the three brain regions.

466

467 **RNA-sequencing Data Processing**

468 We downloaded FASTQ files from the BrainSeq Consortium (Benjamin et al., 2022; Jaffe et al., 2018; Schubert et al., 2015). The reads were aligned to the hg38/GRCh38 human genome 469 470 (GENCODE release 26, GRCh38.p10) using HISAT2 (v2.1.0) (Kim et al., 2019). Following 471 genome alignment, we sorted and indexed the BAM files using SAMtools (v1.9) (Danecek et al., 472 2021) with HTSlib (v1.9) (Bonfield et al., 2021). We examined alignment and read quality with 473 RSeQC (v3.0.1) (L. Wang et al., 2012). We generated gene and repeat counts from both multi 474 mapping (expression and downstream analysis) and unique mapping (co-expression analysis) 475 using TEtranscripts (v2.2.1) (Jin et al., 2015) and featureCounts (v2.0.1) (Liao et al., 2014).

476

477 Generating Counts from Multi-Mapping Reads and Repeat Annotation

478 For TEtranscripts, we generated gene and repeat counts in one step using TEcount for paired 479 end, reversed stranded reads with default parameters and the GTF file of genes and repeats 480 provided by the Hammel Lab (http://hammelllab.labsites.cshl.edu/software/). Additionally, we 481 used featureCounts to generate gene and repeat counts in one step with a customized GTF file 482 of genes and repeats. For the GTF file generation, we combined the GENCODE release 26 with 483 repeat annotation obtained from downloading the repeat masker track for hg38 from the UCSC 484 Table Browser followed by annotating strand information with a python script. We generated 485 counts with featureCounts using the following parameters: 1) paired end, 2) reversed stranded 486 reads, 3) primary alignments only, 4) excluding chimeric reads, 5) allowing for multi-mapping 487 reads and, 6) one base as the minimum overlapping fraction in a read.

488

489 **Quality Control**

For quality control, we first aggregate results for RSeQC and HISAT2 with MultiQC (Ewels et al., 2016). To determine outliers, we first combined the read, alignment, and RNA quality (RIN: RNA Integrity Number and mitochondria mapping rate) for each tissue and scaled the data before applied dimensional reduction with PCA (principal component analysis) with the scikit-learn

494 package (Pedregosa et al., 2011). Following dimensional reduction, we calculated the distance
495 from the centroid for all samples (Equations 1 & 2) and excluded samples that were outside of
496 the 99 percentile (caudate and hippocampus) and 95 percentile (DLPFC). This resulted in a total
497 of 861 samples for caudate (n=268), DLPFC (n=287), and hippocampus (n=306).
498

centroid =
$$\frac{1}{n} \sum_{i=0}^{n} x_i$$
 (Equation 1)

500

499

501
$$distance from centroid = \sum_{j=0}^{k} (x_j - centroid)^2$$
 (Equation 2)

502

503 Low expression filtering and library normalization

504 To filter out low expression counts, we first constructed an edgeR object (McCarthy et al., 2012; 505 Robinson et al., 2009) of brain regions with sample information as well as raw counts. Following 506 this, we applied filterByExpr (Chen et al., 2016) from edgeR for genes and repeats together with 507 an interacting design matrix (Equation 3). This function keeps features (genes and repeats) that 508 have count-per-million (CPM) above a minimum count (10 CPM) in 70% of the smallest group 509 sample size. The smallest group sample size is determined by the design matrix. Furthermore, 510 each feature must have a minimum number of counts across all samples (15 CPM). After filtering, 511 we had a total of 28443, 28058, and 28740 genes and repeats for the caudate nucleus, DLPFC, 512 and hippocampus, respectively for the TEtranscripts method. For the featureCounts methods, this 513 resulted in 24861, 24545, and 25022 genes and repeats for the caudate nucleus, DLPFC, and 514 hippocampus, respectively. After filtering, we normalized the library size of genes and repeats 515 together using trimmed **M**-values (TMM). mean of 516 (Equation 3) 517 $E(Y) = \beta_0 + \beta_1 AgeGroup$

518

519 Repeat Expression Analyses

520

521 Expression residualization

522 For residualized expression, we regressed out covariates using limma-voom normalized 523 expression and null models created without the variable of interest (**Equation 5**) as previously 524 described in Benjamin et al. 2022. Following this, a z-score transformation was performed.

525

526
$$E(Y) = \beta_0 + \beta_1 Race + \beta_2 Sex + \beta_3 MitoRate + \beta_4 TotalAssignedGenes + \beta_5 OverallMappingRate + \beta_6 RIN + \sum_{i=1}^3 \eta_i snpPC_i + \sum_{j=1}^k \gamma_i SV_j$$
 (Equation 4)

528

529 <u>Differential expression analyses</u>

For differential expression analyses, we applied voom normalization law (Law et al., 2014) on the normalized filtered counts for genes and repeats together (Low expression filtering and library normalization), adjusted with the model covariates listed below (Leek et al., 2012) (Equation 4). Differentially expressed features were identified using the eBayes (Smyth; Hall, 2009) function from limma (Ritchie et al., 2015) for the age group fitted model. Features with a FDR < 0.05 were considered as differentially expressed.

536

537
$$E(Y) = \beta_0 + \beta_1 AgeGroup + \beta_2 Race + \beta_3 Sex + \beta_4 MitoRate + \beta_5 AlignmentRate + \beta_5 AlignmentAlignmentAlignmentAlignmentAlignmentAlignmentAlignmentAlignment$$

538
$$\beta_6 UnmappingRate + \beta_7 RIN + \sum_{i=1}^3 \eta_i snpPC_i + \sum_{j=1}^k \gamma_i SV_j$$
 (Equation 5)

539

540 Covariates included sex, self-reported race, ancestry (SNP PCs 1-3), and RNA quality (RIN, 541 mitochondria mapping rate, alignment rate, genome unmapping rate, and hidden variance using 542 surrogate variable analysis (SVA)).

543

544 Repeat Superfamily Hypergeometric Analysis

545 To evaluate the representation of repeat superfamilies in each differential expression result, we applied the supertest function from SuperExactTest (Wang et al., 2015). For superfamily analysis, 546 547 we utilized the total number of repeats (n = 30938) included in the custom GTF file as background 548 population. We then used the intersection between the set of repeats of a repeat superfamily (i.e. 549 all 342 LINEs repeats present in our GTF file) with the total number of differentially expressed 550 repeats within the same repeat family, obtained in each age comparison analysis. We performed 551 the Hypergeometric test in 16 repeat super families. The P-values were calculated by the same 552 function considering only the upper tail of the distribution of each intersection.

553

554 WGCNA analyses

555

556 Generating Counts from Uniquely Mapped Reads

557 Considering the uneven distribution of repeats across the genome and high sequence similarity, 558 we also generated counts from uniquely mapped reads to reduce artifacts introduced by 559 evolutionarily young repeats (Parsana et al., 2019). We generated counts with featureCounts 560 using the following parameters: 1) paired end, 2) reversed stranded reads, 3) primary alignments 561 only, 4) excluding chimeric reads, 5) excluding multi-mapping reads and, 6) one base as the 562 minimum overlapping fraction in a read.

563

564 <u>Co-expression Analysis</u>

We used the Weighted Correlation Network Analysis (WGCNA) to create co-expression networks for each brain region and identify co-expression modules (Langfelder & Horvath, 2008). As an input we used the residualized expression previously described (**Equation 5)**, with genes and repeats together, from featureCounts. We analyzed each brain region in separate, using all age groups from each tissue. We select a $\beta = 14$, (a value with all networks achieved a scale-free independence index of $\mathbb{R}^2 \ge 0.8$), using the following parameters: signed network, mergecutheight

571	= 0.25, deepsplit = 2, and minimum module size = 30 (Feltrin et al., 2019). Each individual module
572	eigengene value (kME) were correlated with the following co-variables: self-reported race, sex,
573	age of death, RIN, pH, PMI, MitoRate, AlignmentRate, CERAD/BRAAK scores and each one of
574	the 6 age groups. Modules with a Pearson correlation coefficient p-value < 0.05 were considered
575	as significant associations. For the identification of hub genes of each module, we selected all the
576	genes/repeats selected by the WGCNA function intramodularConnectivity().
577	
578	CERAD/BRAAK Scores
579	A subset of 108 samples representing 57 unique individuals were analyzed by a Lieber Institute
580	for Brain Development neuropathologist. All 108 samples were given both a CERAD scores (1-4)
581	and BRAAK score (1-4).
582	
583	Gene term enrichment analysis
584	For gene term enrichment analysis, we utilized GOATOOLS, a Python package using
585	hypergeometric tests with the Gene Ontology (GO) database (Klopfenstein et al., 2018). The GO
586	database included molecular functions (MF), cellular components (CC), and biological processes
587	(BP), however, MF and BP were primarily used for analyses.
588	
589	Code and data availability
590	Code will be available at https://github.com/orgs/paquolalab.
591	
592	Supplemental Figures
593	Figure S1. Selection of repeat quantification method. A) Breakdown of post-mortem
594	samples by sex. B) Overlap of quantifiable repeats on positive strand by featureCounts and
595	TEcounts. C) Overlap of all quantifiable repeats features by featureCounts and TEcounts. D)

596	Breakdown of categories of quantifiable repeats in TEcounts GTF file. E) Breakdown of
597	categories of quantifiable repeats in featureCounts GTF file. F) Correlation between
598	featureCounts and TEcounts, raw counts of L1HS. G) Correlation between featureCounts and
599	TEcounts, raw counts of SVA_F.
600	
601	Figure S2. Co-expression defines correlation of repeat-dense WGCNA modules with
602	clinical traits. A) WGCNA module-trait correlation heatmap across caudate nucleus samples.
603	B) WGCNA module-trait correlation heatmap across DLPFC samples. C) WGCNA module-trait
604	correlation heatmap across hippocampus samples.
605	
606	Figure S3. Gene ontology enrichment of gene- and repeat-dense WGCNA modules. A) Gene
607	ontology enrichment of gene-dense DLPFC blue module B) Gene ontology enrichment of gene-
608	dense hippocampus Yellow module C) Gene ontology enrichment of repeat-dense caudate
609	nucleus Yellow module D) Gene ontology enrichment of repeat-dense hippocampus Brown
610	module.
611	
612	Figure S4. Distribution of differentially expressed repeats (DERs) across lifespan of
613	neurotypical brain. A) Total differentially expressed repeats (DERs) across each age
614	comparison in caudate nucleus, DLPFC, hippocampus. B) Upset Plot with the distribution of
615	unique and shared 0-15 vs. 60y+ DERs by brain region.
616	
617	Figure S5. Directionality of DERs across lifespan of neurotypical brain. A) Up- and
618	downregulated 0-15 vs. 60y+ DERs across each brain region. B) Up- and downregulated DERs
619	across age comparisons by brain region.

620

Figure S6. LTR and HERV-K expression correlates with age of death. A) Correlation of total LTR expression with age of death across all brain regions. B) Correlation of total HERV expression with age of death across all brain regions. C) Correlation of total HERV expression with age of death across each individual brain region.

625

626 Supplemental Tables

627 **Table S1. Sample clinical data.** Information regarding all samples included in this project.

628

Table S2. Scale-free topology fit index for each tissue-specific network (WGCNA). Values

obtained from pickSoftThreshold() function, to calculate the appropriate Beta of each region's co-

631 expression network (Caudate, DLPFC and Hippocampus).

632

Table S3. WGCNA modules composition and features annotation for caudate nucleus, DLPFC, and hippocampus' co-expression networks. For the annotated genes, information regarding its chromosomal location is provided. For each repeats, information from its family and main class (obtained by RepeatMasker annotation) are also included. Features clustered in the 'grey' (null) module were excluded for further downstream analysis. Features without the module identification were absent for the WGCNA analysis of its respective co-expression network analysis.

640

Table S4. Gene Ontology enrichment results for each WGCNA co-expression module for
caudate nucleus, DLPFC and hippocampus co-expression networks. Results were obtained
with the GOATOOLS package. Only GO pathways with a FDR < 0.05 were considered as either
enriched ('e') or depleted ('p'). GO: Gene Ontology; NS: Gene Ontology Category; BP: Biological
Process; MF: Molecular Function; CC: Cellular Component.

646

647	Table S5. Power Analysis. Power was derived from the sample sizes of each differential
648	expression age group: Prenatal, A (0-15), B (16-29), C (30-39), D (40-49), E (50-59), F (60+),
649	using TTestIndPower and FTestPower functions from Python's statsmodels.stats.power.
650	
651	Table S6. Differential expression analysis applying limma-voom to features quantified by
652	featureCounts algorithm. Only features with a Benjamini-Hochberg false discovery rate (FDR)
653	< 0.05 were considered as differentially expressed and included. AveExpr: average expression
654	across all samples; logFC: estimate of the log2-fold-change corresponding to the effect; t:
655	moderated t-statistic; B: log-odds that the gene is differentially expressed.
656	
657	Table S7. Differential expression analysis applying limma-voom and features quantified by
658	TEcount algorithm. Only features with a Benjamini-Hochberg false discovery rate < 0.05 were
659	considered as differentially expressed and included. AveExpr: average expression across all
660	samples; logFC: estimate of the log2-fold-change corresponding to the effect; t: moderated t-
661	statistic; B: log-odds that the gene is differentially expressed.
662	
663	Table S8. Top intramodular hubs from each WGCNA module. List of features (genes and/or
664	repeats) that serve as intramodular hubs determined by the intramodularConnectivity.fromExpr()
665	function. kTotal = total connectivity; kWhitin = intramodular connectivity; kOut = extra-modular
666	connectivity; kDiff = the difference between the intra-modular and extra-modular connectivity.
667	
668	Data Availability
669	Raw data (.fastq files) are available under restricted access to protect research subjects. For total
670	RNA data from prefrontal cortex and hippocampus, researchers can access the files via the
671	Globus collections (jhpce#bsp2-dlpfc and jhpce#bsp2-hippo) at https://research.libd.org/globus/.

For caudate nucleus data, researchers can obtain access to FASTQ files via dbGaP accession
phs003495.v1.p1 at https://www.ncbi.nlm.nih.gov/gap/.

674

675 Acknowledgements

676 The authors gratefully acknowledge the families that have donated this tissue to the advancement 677 of science. The authors would like to extend their appreciation to the Offices of the Chief Medical 678 Examiner of Washington DC, Northern Virginia, Kalamazoo Michigan, Santa Clara County, 679 University of North Dakota, and Maryland for the provision of brain tissue used in this work. The 680 authors also extend their posthumous appreciation to to Dr. Llewellyn B. Bigelow and members 681 of the LIBD Neuropathology Section for their work in assembling and curating the clinical and 682 demographic information and organizing the Human Brain Tissue Repository of the Lieber 683 Institute. We thank Dr. Alan Lorenzetti for the thoughtful feedback on this manuscript. We thank 684 Johns Hopkins School of Medicine, Cellular and Molecular Medicine Graduate Program for 685 additional support.

686

687 Ethics Statement & Financial Disclosure

688 All sample identifiers used in this study are de-identified IDs and cannot reveal the identity of the 689 study subjects. Additionally, all available sample clinical information also cannot reveal the identity 690 of the study subjects. D.R.W. serves on the Scientific Advisory Boards of Sage Therapeutics and 691 Pasithea Therapeutics. J.E.K. has served as a consultant for Merck on an antipsychotic drug trial. 692 All other authors declare no competing interests. This work is supported by the Lieber Institute for 693 Brain Development. JAE is supported by a NARSAD Young Investigator Grant from the Brain and 694 Behavior Research Foundation and Collaborative Center For X-Linked Dystonia Parkinsonism 695 and the MGH Collaborative Center for X-Linked Dystonia-Parkinsonism. JAE and ASF are 696 supported by the Maryland Stem Cell Research Foundation. KJB is supported by T32 fellowship 697 (T32MH015330) and K99 award (K99MD016964).

698

699 Contributions

700

701 This work was collaborative, and many people contributed to many aspects of the project. 702 Previously published postmortem sample collection and RNA sequencing were directed by Tom 703 Hyde, Joel Kleinman and Daniel Weinberger at Lieber Institute for Brain Development. Repeat 704 quantification was performed by Kynon Benjamin, Taylor Evans, Arthur Feltrin, Tarun Katipalli 705 and Apua Paguola. WGCNA was largely conducted by Arthur Feltrin. Analysis and visualization 706 of WGCNA results were largely conducted by Taylor Evans and Arthur Feltrin. Differential 707 expression was performed by Kynon J Benjamin, Tarun Katipalli and Apua Paguola. Analysis and 708 visualization of differential expression results was largely conducted by Taylor Evans with 709 assistance from Kynon J Benjamin, Tarun Katipalli and Apua Paguola. The manuscript text and 710 figures were largely performed by Taylor Evans, with assistance from Arthur Feltrin and Jennifer 711 Erwin. Interpretation of findings was conducted by Taylor A. Evans with assistance from Arthur 712 Feltrin, Apua Paguola, and Jennifer Erwin. The study was conceptualized and designed by

- 713 Jennifer Erwin and Apua Paguola.
- 714

715 References

716

717 Benjamin, K.J.M., Chen, Q., Jaffe, A.E., Stolz, J.M., Collado-Torres, L., Huuki-Myers, L.A., Burke, E.E., Arora, R.,

718 Feltrin, A.S., Barbosa, A.R., Radulescu, E., Pergola, G., Shin, J.H., Ulrich, W.S., Deep-Soboslay, A., Tao, R.,

719 Matsumoto, M., Saito, T., Tajinda, K. and Hoeppner, D.J. (2022). Analysis of the caudate nucleus transcriptome in

720 individuals with schizophrenia highlights effects of antipsychotics and new risk genes. Nature Neuroscience, 25(11),

- 721 pp.1559-1568. doi:https://doi.org/10.1038/s41593-022-01182-7.
- 722
- 723 Bogenpohl, J.W., Mignogna, K.M., Smith, M.L. and Miles, M.F. (2016). Integrative Analysis of Genetic, Genomic, and
- 724 Phenotypic Data for Ethanol Behaviors: A Network-Based Pipeline for Identifying Mechanisms and Potential Drug
- 725 Targets. In: Methods in molecular biology. [online] Springer Science+Business Media, pp.531–549.
- 726 doi:https://doi.org/10.1007/978-1-4939-6427-7_26.

- 727 Bogu, G.K., Reverter, F., Marti-Renom, M.A., Snyder, M.P. and Guigó, R. (2019). Atlas of transcriptionally active
- transposable elements in human adult tissues. *bioRxiv (Cold Spring Harbor Laboratory)*.
- 729 doi:https://doi.org/10.1101/714212.
- 730 Bonfield, J.K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T. and Davies, R.M. (2021). HTSlib:
- 731 C library for reading/writing high-throughput sequencing data. *GigaScience*, 10(2).
- doi:https://doi.org/10.1093/gigascience/giab007.
- 733 Brudek, T., Christensen, T., Aagaard, L., Petersen, T., Hansen, H.J. and Møller-Larsen, A. (2009). B cells and
- 734 monocytes from patients with active multiple sclerosis exhibit increased surface expression of both HERV-H Env and
- HERV-W Env, accompanied by increased seroreactivity. *Retrovirology*, 6(1). doi:https://doi.org/10.1186/1742-4690-6104.
- 737 Chen, Y., Lun, A.T.L. and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of
- RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. [online] f1000research.com.
 Available at: https://f1000research.com/articles/5-1438.
- 740 Collado-Torres, L., Burke, E.E., Peterson, A.M., Shin, J.H., Straub, R.E., Anandita Rajpurohit, Semick, S.A., Ulrich,
- 741 W.S., Price, A.J., Valencia, C., Tao, R., Deep-Soboslay, A., Hyde, T.M., Kleinman, J.E., Weinberger, D.R. and Jaffe,
- A.E. (2019). Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and
- 743 Hippocampus across Development and Schizophrenia. *Neuron*, 103(2), pp.203-216.e8.
- 744 doi:https://doi.org/10.1016/j.neuron.2019.05.013.
- 745 Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy,
- 746 S.A., Davies, R.M. and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2).
- 747 doi:https://doi.org/10.1093/gigascience/giab008.
- 748 de Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., Peterson, A. L.,
- Kreiling, J. A., Neretti, N., & Sedivy, J. M. (2013). Genomes of replicatively senescent cells undergo global epigenetic
 changes leading to gene silencing and activation of transposable elements. *Aging Cell*, *12*(2), 247–256.
- 751
- de Magalhães, J.P., Curado, J. and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles
- identifies common signatures of aging. *Bioinformatics*, 25(7), pp.875–881.
- doi:https://doi.org/10.1093/bioinformatics/btp073.
- 755
- 756 Della Valle, F., Reddy, P., Yamamoto, M., Liu, P., Saera-Vila, A., Bensaddek, D., Zhang, H., Prieto Martinez, J.,
- 757 Abassi, L., Celii, M., Ocampo, A., Nuñez Delicado, E., Mangiavacchi, A., Aiese Cigliano, R., Rodriguez Esteban, C.,
- 758 Horvath, S., Izpisua Belmonte, J.C. and Orlando, V. (2022). LINE-1 RNA causes heterochromatin erosion and is a
- target for amelioration of senescent phenotypes in progeroid syndromes. Science Translational Medicine, 14(657).
- 760 doi:https://doi.org/10.1126/scitranslmed.abl6057.
- Dembny, P., Newman, A.G., Singh, M., Hinz, M., Szczepek, M., Krüger, C., Adalbert, R., Dzaye, O., Trimbuch, T.,
 Wallach, T., Kleinau, G., Derkow, K., Richard, B.C., Schipke, C., Scheidereit, C., Stachelscheid, H., Golenbock, D.,

- Peters, O., Coleman, M. and Heppner, F.L. (2020). Human endogenous retrovirus HERV-K(HML-2) RNA causes
 neurodegeneration through Toll-like receptors. *JCl Insight*, 5(7). doi:https://doi.org/10.1172/jci.insight.131093.
- 765 Ding, S., Wang, S., He, K., Jiang, M. and Li, F. (2017). Large-scale analysis reveals that the genome features of
- simple sequence repeats are generally conserved at the family level in insects. *BMC Genomics*, 18(1).
- 767 doi:https://doi.org/10.1186/s12864-017-4234-0.
- 768 Dumetier, B., Sauter, C., Hajmirza, A., Pernon, B., Aucagne, R., Fournier, C., Row, C., Guidez, F., Rossi, C., Lepage,
- 769 C., Delva, L. and Callanan, M.B. (2022). Repeat Element Activation-Driven Inflammation: Role of NFkB and
- 770 Implications in Normal Development and Cancer? *Biomedicines*, [online] 10(12), pp.3101–3101.
- doi:https://doi.org/10.3390/biomedicines10123101.
- Euclydes, V.L.V., Gastaldi, V.D., Feltrin, A.S., Hoffman, D.J., Gouveia, G., Cogo, H., Felipe-Silva, A., Vieira, R.P.,
- 773 Miguel, E.C., Polanczyk, G.V., Chiesa, A., Fracolli, L., Matijasevich, A., Ferraro, A., Argeu, A., Maschietto, M. and
- 8774 Brentani, H.P. (2022). DNA methylation mediates a randomized controlled trial home-visiting intervention during
- pregnancy and the Bayley infant's cognitive scores at 12 months of age. Journal of Developmental Origins of Health
- 776 *and Disease*, 13(5), pp.556–565. doi:https://doi.org/10.1017/S2040174421000738.
- 777 Evans, T.A. and Erwin, J.A. (2021). Retroelement-derived RNA and its role in the brain. Seminars in cell &
- 778 *developmental biology*, 114, pp.68–80. doi:https://doi.org/10.1016/j.semcdb.2020.11.001.
- Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools
- and samples in a single report. *Bioinformatics*, 32(19), pp.3047–3048.
- 781 doi:https://doi.org/10.1093/bioinformatics/btw354.
- 782 Feltrin, A.S., Tahira, A.C., Simões, S.N., Brentani, H. and Martins, D. (2019). Assessment of complementarity of
- 783 WGCNA and NERI results for identification of modules associated to schizophrenia spectrum disorders. *PLOS ONE*,
- 784 14(1), pp.e0210431–e0210431. doi:https://doi.org/10.1371/journal.pone.0210431.
- 785 Frost, B., Hemberg, M., Lewis, J. and Feany, M.B. (2014). Tau promotes neurodegeneration through global
- 786 chromatin relaxation. *Nature Neuroscience*, 17(3), pp.357–366. doi:https://doi.org/10.1038/nn.3639.
- 787 Gorbunova, V., Seluanov, A., Mita, P., McKerrow, W., Fenyö, D., Boeke, J. D., Linker, S. B., Gage, F. H., Kreiling, J.
- A., Petrashen, A. P., Woodham, T. A., Taylor, J. R., Helfand, S. L., & Sedivy, J. M. (2021). The role of
- retrotransposable elements in ageing and age-associated diseases. *Nature*, 596(7870), 43–53.
- Guerreiro, R. and Bras, J. (2015). The age factor in Alzheimer's disease. *Genome Medicine*, [online] 7(1), pp.1–3.
 doi:https://doi.org/10.1186/s13073-015-0232-5.
- Guo, C., Jeong, H.-H., Hsieh, Y.-C., Klein, H.-U., Bennett, D.A., De Jager, P.L., Liu, Z. and Shulman, J.M. (2018).
- 793 Tau Activates Transposable Elements in Alzheimer's Disease. Cell Reports, 23(10), pp.2874–2880.
- 794 doi:https://doi.org/10.1016/j.celrep.2018.05.004.

- Ham, S. and Lee, S.-J.V. (2020). Advances in transcriptome analysis of human brain aging. *Experimental & Molecular Medicine*, [online] 52(11), pp.1787–1797. doi:https://doi.org/10.1038/s12276-020-00522-6.
- Hanna, R., Flamier, A., Barabino, A., & Bernier, G. (2021). G-quadruplexes originating from evolutionary conserved
 L1 elements interfere with neuronal gene expression in Alzheimer's disease. *Nature Communications*, *12*(1).
- Jaffe, A.E., Straub, R.E., Shin, J.H., Tao, R., Gao, Y., Collado-Torres, L., Kam-Thong, T., Xi, H.S., Quan, J., Chen,
- 801 Q., Colantuoni, C., Ulrich, W.S., Maher, B.J., Deep-Soboslay, A., BrainSeq Consortium, Cross, A.J., Brandon, N.J.,
- 802 Leek, J.T., Hyde, T.M. and Kleinman, J.E. (2018). Developmental and genetic regulation of the human cortex
- transcriptome illuminate schizophrenia pathogenesis. *Nature Neuroscience*, [online] 21(8), pp.1117–1125.
 doi:https://doi.org/10.1038/s41593-018-0197-y.
- 805

806 Jin, X., Xu, X.-E., Jiang, Y.-Z., Liu, Y.-R., Sun, W., Guo, Y.-J., Ren, Y.-X., Zuo, W.-J., Hu, X., Huang, S.-L., Shen, H.-

- 807 J., Lan, F., He, Y.-F., Hu, G.-H., Di, G.-H., He, X.-H., Li, D.-Q., Liu, S., Yu, K.-D., & Shao, Z.-M. (2019). The
- endogenous retrovirus-derived long noncoding RNA TROJAN promotes triple-negative breast cancer progression via
 ZMYND8 degradation. *Science Advances*, *5*(3), doi:10.1126/sciadv.aat9820
- 810
- Jin, Y., Tam, O.H., Paniagua, E. and Hammell, M. (2015). TEtranscripts: a package for including transposable
- 812 elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, 31(22), pp.3593–3599.
 813 doi:https://doi.org/10.1093/bioinformatics/btv422.
- Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019). Graph-based genome alignment and
- genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), pp.907–915.
- 816 doi:https://doi.org/10.1038/s41587-019-0201-4.
- 817

818 Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes,

- J.M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P. and Tang, H. (2018). GOATOOLS: A Python
- 820 library for Gene Ontology analyses. *Scientific Reports*, 8(1). doi:https://doi.org/10.1038/s41598-018-28948-z.
- 821 Kmetzsch, V., Anquetil, V., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., Jornéa, L., Forlani, S., Couratier, P.,

Wallon, D., Pasquier, F., Robil, N., De La Grange, P., Moszer, I., Le Ber, I., Colliot, O., & Becker, E. (2020). Plasma

823 microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia

- and amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 92(5), 485–493.
- 825
- Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, [online] 9(1). doi:https://doi.org/10.1186/1471-2105-9-559.
- 828 LaRocca, T.J., Cavalier, A.N. and Wahl, D. (2020). Repetitive elements as a transcriptomic marker of aging:
- 829 Evidence in multiple datasets and models. *Aging Cell*, [online] 19(7), p.e13167.
- 830 doi:https://doi.org/10.1111/acel.13167.

- Laufer, G., Mayer, J., Mueller, B.F., Mueller-Lantzsch, N. and Ruprecht, K. (2009). Analysis of transcribed human
- 832 endogenous retrovirus W env loci clarifies the origin of multiple sclerosis-associated retrovirus env sequences.
- 833 *Retrovirology*, 6(1), p.37. doi:https://doi.org/10.1186/1742-4690-6-37.
- Law, C.W., Chen, Y., Shi, W. and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for
 RNA-seq read counts. *Genome Biology*, 15(2), p.R29. doi:https://doi.org/10.1186/gb-2014-15-2-r29.
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., Lohr, J. G., Harris, C. C., Ding, L., Wilson,
- R. K., Wheeler, D. A., Gibbs, R. A., Kucherlapati, R., Lee, C., Kharchenko, P. V., & Park, P. J. (2012). Landscape of
 Somatic Retrotransposition in Human Cancers. *Science*, *337*(6097), 967–971.
- 839
- Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. (2012). The sva package for removing batch
 effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, [online] 28(6), pp.882–883.
 doi:https://doi.org/10.1093/bioinformatics/bts034.
- Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., & Dubnau, J. (2013). Activation of
 transposable elements during aging and neuronal decline in Drosophila. *Nature Neuroscience*, *16*(5), 529–531.
- Liao, Y., Smyth, G.K. and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning
 sequence reads to genomic features. *Bioinformatics*, 30(7), pp.923–930.
- doi:https://doi.org/10.1093/bioinformatics/btt656.
- Liu, S., Heumüller, S.-E., Hossinger, A., Müller, S.A., Buravlova, O., Lichtenthaler, S.F., Denner, P. and Vorberg, I.M.
- 850 (2023). Reactivated endogenous retroviruses promote protein aggregate spreading. *Nature Communications*, [online]
- 851 14(1), p.5034. doi:https://doi.org/10.1038/s41467-023-40632-z.
- López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. and Kroemer, G. (2013). The Hallmarks of Aging. *Cell*,
 153(6), pp.1194–1217. doi:https://doi.org/10.1016/j.cell.2013.05.039.
- Macías-Redondo, S., Strunk, M., Cebollada-Solanas, A., Ara, J.-R., Jesús Martín, J. and Schoorlemmer, J. (2021).
 Upregulation of selected HERVW loci in multiple sclerosis. *Mobile DNA*, 12(1). doi:https://doi.org/10.1186/s13100021-00243-1.
- 857 McCarthy, D.J., Chen, Y. and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq
- experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), pp.4288–4297.
- doi:https://doi.org/10.1093/nar/gks042.
- 860 Parsana, P., Ruberman, C., Jaffe, A.E., Schatz, M.C., Battle, A. and Leek, J.T. (2019). Addressing confounding
- artifacts in reconstruction of gene co-expression networks. *Genome Biology*, [online] 20(1).
- doi:https://doi.org/10.1186/s13059-019-1700-9.

- 863 Pascut, D., Pratama, M.Y., Gilardi, F., Giuffrè, M., Crocè, L.S. and Tiribelli, C. (2020). Weighted miRNA co-
- 864 expression networks analysis identifies circulating miRNA predicting overall survival in hepatocellular carcinoma
- 865 patients. *Scientific Reports*, 10(1). doi:https://doi.org/10.1038/s41598-020-75945-2.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer,
- 867 P., Weiss, R., Weiss, R.J., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.
- 868 (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res., [online] 12.
- 869 doi:https://doi.org/10.5555/1953048.2078195.
- Petersen, M., Armisén, D., Gibbs, R.A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis, O. and Misof, B.
- 871 (2019). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to
- 872 insects. *BMC Ecology and Evolution*, 19(1). doi:https://doi.org/10.1186/s12862-018-1324-9.
- 873 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2019). limma powers differential
- 874 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), pp.e47–e47.
- 875 doi:https://doi.org/10.1093/nar/gkv007.
- 876 Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2009). edgeR: a Bioconductor package for differential expression
- analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.
- doi:https://doi.org/10.1093/bioinformatics/btp616.
- Sahakyan, A. B., Murat, P., Mayer, C., & Balasubramanian, S. (2017). G-quadruplex structures within the 3' UTR of
 LINE-1 elements stimulate retrotransposition. *Nature Structural & Molecular Biology*, 24(3), 243–247.
- 881
- 882 Schmitt, K., Richter, C., Backes, C., Meese, E., Ruprecht, K. and Mayer, J. (2013). Comprehensive Analysis of
- 883 Human Endogenous Retrovirus Group HERV-W Locus Transcription in Multiple Sclerosis Brain Lesions by High-
- Throughput Amplicon Sequencing. *Journal of Virology*, 87(24), pp.13837–13852.
- 885 doi:https://doi.org/10.1128/jvi.02388-13.
- Schrader, L., & Schmitz, J. (2018). The impact of transposable elements in adaptive evolution. *Molecular Ecology*,
 28(6), 1537–1549.
- 888
- Schubert, C., O'Donnell, P., Quan, J., Wendland, J.R., Xi, H.S., Winslow, A.R., Domenici, E., Essioux, L., Kam-
- Thong, T., Airey, D., Calley, J.N., Collier, D.A., Wang, H., Eastwood, B.J., Ebert, P.A., Liu, Y., Nisenbaum, L., Ruble,
- 891 C.L., Scherschel, J.E. and Smith, R. (2015). BrainSeq: Neurogenomics to Drive Novel Target Discovery for
- 892 Neuropsychiatric Disorders. *Neuron*, 88(6), pp.1078–1083. doi:https://doi.org/10.1016/j.neuron.2015.10.047.
- Simon, M., Van Meter, M., Ablaeva, J., Ke, Z., Gonzalez, R.S., Taguchi, T., De Cecco, M., Leonova, K.I., Kogan, V.,
- Helfand, S.L., Neretti, N., Roichman, A., Cohen, H.Y., Meer, M.V., Gladyshev, V.N., Antoch, M.P., Gudkov, A.V.,
- 895 Sedivy, J.M., Seluanov, A. and Gorbunova, V. (2019). LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient
- Mice Drives Inflammation. *Cell metabolism*, [online] 29(4), pp.871-885.e5.
- 897 doi:https://doi.org/10.1016/j.cmet.2019.02.014.

- 898 Smyth, G. and Hall, E. (2009). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in
- 899 Microarray Experiments. Statistical Applications in Genetics and Molecular Biology, [online] 3(1). Turelli, P., Playfoot,
- 900 C., Grun, D., Raclot, C., Pontis, J., Coudray, A., Thorball, C., Duc, J., Pankevich, E.V., Deplancke, B., Busskamp, V.
- 901 and Trono, D. (2020). Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene
- 902 expression in human neurons. *Science Advances*, 6(35), p.eaba3200. doi:https://doi.org/10.1126/sciadv.aba3200.
- 903 Wang, L., Wang, S. and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, 28(16),
- 904 pp.2184–2185. doi:https://doi.org/10.1093/bioinformatics/bts356.
- 905 Wang, M., Zhao, Y. and Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. *Scientific*
- 906 *Reports*, [online] 5(1). doi:https://doi.org/10.1038/srep16923.
- 907 Yamamoto, R., Chung, R., Vazquez, J.M., Sheng, H., Steinberg, P.L., Ioannidis, N.M. and Sudmant, P.H. (2022).
- 908 Tissue-specific impacts of aging and genetics on gene expression patterns in humans. *Nature Communications*,
- 909 [online] 13(1), p.5803. doi:https://doi.org/10.1038/s41467-022-33509-0.

910









Hippocampus



Caudate



DLPFC



E. LINE





Pre











Caudate nucleus







16 30 50 40 ns • • . • 0-15 • • • • 16-29 • ns ns 30-39 ns ns

40-49

DLPFC

60+

•

•

ns

ns

ns

ns

ns

50-59



Pre

0

0-15

•

16-29

Pre



Caudate nucleus



Caudate nucleus

H. Satellite











0

Pre





50-59



0



Observed Overlap



Β.









Ε.

Popost	Strand	Repeat	FDR,	FDR,	FDR,	
<u>repear</u>	Stranu	<u>Class</u>	<u>Caudate</u>	DLPFC	<u>Hippo</u>	
MER4D0	-	LTR	0.042	0.02	0.012	
MLT1G-int	+	LTR	1.24E-06	4.80E-05	0.0001	
MER84-int	-	LTR	7.11E-08	0.00024	8.70E-05	
MLT1F1-int	+	LTR	1.95E-06	3.20E-05	0.0024	
MamGypLTR2c	+	LTR	2.29E-10	0.0002	0.00076	
HERVK-int	+	LTR	0.011	0.043	0.00026	
MLT1E3-int	-	LTR	0.0059	0.029	0.0029	
LTR57-int	-	LTR	0.03	1.37E-06	0.0035	
MER66-int	-	LTR	0.015	0.032	0.0025	
MLT1F2-int	-	LTR	0.00035	0.022	0.0051	
LTR12B	+	LTR	0.00034	0.00057	0.00067	
MER101	+	LTR	0.0015	0.0075	0.0037	
LTR45	-	LTR	0.0028	0.00055	0.0028	
MER57C1	+	LTR	0.0027	0.0094	0.036	
LTR2752	-	LTR	0.0014	0.00066	0.00014	
MER61F	+	LTR	0.0011	0.00057	0.0019	
HERV9NC-int	+	LTR	0.0011	0.0044	0.01	









D. TEcounts



E. featureCounts









A. Certrateprint clerct doi.org/10.1101/2024.05.17.24307184; this version posted May 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	,	death	tegitty		د	çe	entrat	ancest	ananc		~ ~ ~	\ \L
	Pop	PHA	i d	PMI	Mitor	Alight	Atica	in Entope	Male Male	Femal	CERAN	Braat
MEblack	-0.018	-0.0024	0.13	0.054	0.0019	0.007	0.00063	-0.00063	-0.0026	0.0026	0.1	-0.17
	(0.8)	(1)	(0.03)	(0.4)	(1)	(0.9)	(1)	(1)	(1)	(1)	(0.09)	(0.004)
MEred	-0.0099	-0.023	0.11	0.045	0.0013	-0.021	-0.002	0.002	-8.3e-05	8.3e-05	0.13	-0.21
	(0.9)	(0.7)	(0.06)	(0.5)	(1)	(0.7)	(1)	(1)	(1)	(1)	(0.03)	(5e-04)
MEyellow	0.057	-0.012	-0.034	-0.0022	-0.00038	-0.0031	-0.0063	0.0063	0.0032	-0.0032	0.14	0.057
	(0.4)	(0.8)	(0.6)	(1)	(1)	(1)	(0.9)	(0.9)	(1)	(1)	(0.03)	(0.3)
//Elightcyan	-0.084	-0.001	0.087	0.028	0.00016	0.0036	0.00062	-0.00062	-0.00056	0.00056	0.046	0.28
	(0.2)	(1)	(0.2)	(0.7)	(1)	(1)	(1)	(1)	(1)	(1)	(0.5)	(3e–06)
MEsalmon	0.032	-8e-04	0.083	0.0013	-5.4e-06	0.0058	7e-04	-7e-04	-0.0015	0.0015	0.15	0.2
	(0.6)	(1)	(0.2)	(1)	(1)	(0.9)	(1)	(1)	(1)	(1)	(0.01)	(9e–04)



C. Hippocampus





C. Caudate nucleus, yellow





A. Total, Differentially Expressed Repeats

B. DERs shared by tissue, 0-15 v. 60y+









B. Directionality of DERs across comparisons, brain region

A. Total repeat category expression, all regions





C. HERVs, total expression, by brain region

