Nanopore-based DNA long-read sequencing analysis of the aged human brain

Paulino Ramirez, PhD[1,2,3], Wenyan Sun, PhD[1,2,3,6], Shiva Kazempour Dehkordi [2,3], Habil Zare, PhD[2,3], Bernard Fongang, PhD[2,4], Kevin F. Bieniek, PhD[2,5], Bess Frost, PhD[1,2,3]

[1]Barshop Institute for Longevity and Aging Studies, [2]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, [3]Department of Cell Systems and Anatomy, University of Texas Health San Antonio, San Antonio, Texas, [4]Department of Biochemistry & Structural Biology, University of Texas Health San Antonio, San Antonio, Texas, [5]Department of Pathology, University of Texas Health San Antonio, San Antonio, Texas, [6]School of Pharmacy, University of Missouri-Kansas City, Kansas City, Missouri

*Corresponding Author and Lead Contact: Bess Frost, Ph.D.
Associate Professor
4939 Charles Katz Dr.
Barshop Institute, Rm 1041
University of Texas Health San Antonio
San Antonio, TX 78229
Phone: 210-562-5037
bfrost@uthscsa.edu

ABSTRACT

Aging disrupts cellular processes such as DNA repair and epigenetic control, leading to a gradual buildup of genomic alterations that can have detrimental effects in post-mitotic cells. Genomic alterations in regions of the genome that are rich in repetitive sequences, often termed "dark loci," are difficult to resolve using traditional sequencing approaches. New long-read technologies offer promising avenues for exploration of previously inaccessible regions of the genome. Using nanopore-based long-read whole-genome sequencing of DNA extracted from aged 18 human brains, we identify previously unreported structural variants and methylation patterns within repetitive DNA, focusing on transposable elements ("jumping genes") as crucial sources of variation, particularly in dark loci. Our analyses reveal potential somatic insertion variants and provides DNA methylation frequencies for many retrotransposon families. We further demonstrate the utility of this technology for the study of these challenging genomic regions in brains affected by Alzheimer's disease and identify significant differences in DNA methylation in pathologically normal brains versus those affected by Alzheimer's disease. Highlighting the power of this approach, we discover specific polymorphic retrotransposons with altered DNA methylation patterns. These retrotransposon loci have the potential to contribute to pathology, warranting further investigation in Alzheimer's disease research. Taken together, our study provides the first long-read DNA sequencing-based analysis of retrotransposon sequences, structural variants, and DNA methylation in the aging brain affected with Alzheimer's disease neuropathology.

INTRODUCTION

The brain accumulates structural variants that result in genomic mosaicism between cells over the course of human aging [1]. Structural variants can occur in progenitor cells during development, proliferating glial and epithelial cells, and in post-mitotic neurons. Such genetic variants can arise from natural biological processes such as DNA repair, recombination, replication and retrotransposition. Studies of the human brain report changes in DNA content, DNA copy number variation, somatic gene recombination, and tandem repeat expansions, pointing to accumulation of genomic rearrangements in the brain [2-4]. The functional impacts of various structural genetic variants depend on the cell type and region of the genome affected.

Class II transposable elements, also known as retrotransposons, make up 35% of the human genome [5] and are a potential source of genomic variation. Active retrotransposons follow a life cycle in which mRNA is reverse transcribed into complementary DNA that can then be integrated back into the host genome. Some retrotransposons encode protein products such as reverse transcriptase and endonuclease that facilitate retrotransposition [6]. In the human genome, specific subfamilies of long interspersed elements (LINEs), short interspersed elements (SINEs), and SINE-VNTR-Alu (SVA) retrotransposons are capable of autonomous retrotransposition and trans-retrotransposition, respectively. Retrotransposition of the active human-specific LINE-1 element (L1Hs) has been documented in neurons and glia, and is thought to be a natural occurrence in neurodevelopment, with estimated cellular insertion rates ranging from 0.04 to 13.7 per neuron and 0.58 to 6.5 per glial cell [7-9]. The estimated frequency of L1Hs transposition varies among regions of brain analyzed and the methodology used (i.e. single cell vs. whole brain lysates).

Several technologies have been employed to investigate retrotransposon mobilization, including oligonucleotide arrays [10], probe-based detection [11], qPCR (quantitative polymerase chain reaction) [12], and DNA sequencing [13]. While low-cost, high-throughput paired-end sequencing has revolutionized the detection of small genetic variants, the analysis of larger complex structural variants (>50 bp) such as retrotransposon insertions, duplications, translocations, and tandem repeat expansions in low-complexity regions is not possible using traditional sequencing approaches. Specialized tools are needed for short-read analysis of these challenging variants, leading to conflicting quantification and identification [14, 15]. These hurdles extend to detecting somatic variants, typically requiring single-cell whole-genome sequencing. This approach is, however, limited to small cell numbers and relies on PCR amplification, which can introduce artifacts [16]. Clonal expansion of isolated proliferating cells in culture provides another option [17], however this approach cannot be applied to post-mitotic cells such as neurons and does not capture the complex biology of the adult human brain.

The advancement of long-read sequencing now allows resolution of "dark loci," highly repetitive regions of the genome that were previously difficult to sequence. These regions include low-complexity regions consisting of simple and tandem repeats, as well as repetitive regions rich in transposable elements. Traditionally, dark loci have been difficult to characterize due to either lack of sequencing depth or mapping quality, with many of these loci originating from genomic duplication [18]. Long-read sequencing has addressed this problem by generating reads with lengths of tens of kilobases without the need for PCR amplification, compared to the 150-300 bp lengths of short-read technologies. Long-read sequencing can generate reads spanning the entire structural variant and flanking regions, allowing for more confident identification of variants present in repetitive regions of the genome [18-20]. Long-read sequencing also enables detection of germline-derived and somatic retrotransposon insertions, including those present in only a single cell, as demonstrated in *Drosophila melanogaster* head and midgut using singleton reads (somatic insertions supported by a single read) [21]. Importantly, many dark loci are present in genes previously implicated in human disease, but our inability to detect them with affordable high-throughput technologies has led to their exclusion from analysis in most studies.

In addition to producing long reads, Oxford Nanopore DNA sequencing allows for simultaneous calling of DNA modifications such as 5-methylcytosine (5mC). The combination of long reads and accompanying methylation data allows

for methylation calls within low complexity and repetitive regions [22] that are generally ignored in whole genome bisulfite sequencing and are generally missing from array-based methods [23]. The 5mC DNA modification regulates many cellular processes such as transcription, retrotransposon silencing, genome stability and cellular proliferation [24, 25]. Methylation patterns are unique to cell type, brain region, age, and sex. In the brain, DNA methylation plays a role in learning, memory, and synaptic plasticity [26, 27]. In general, CpG methylation decreases across the genome as organisms age, although specific sites are reported to undergo hypermethylation [28]. 5mC hypomethylation has also been implicated in many human age-related disorders, including cancer [29] and Alzheimer's disease [30-35]. Most studies to date utilize probe-based methylation arrays to analyze a specific subset of CpG sites in the genome or apply whole genome bisulfite sequencing (WGBS) to analyze the entire CpG methylome with short read sequencing. WGBS suffers from the inability to distinguish between 5mC and 5-hydroxymethylcytosine modifications, leading to possibly inflated estimations of 5mC frequency.

With the advent of long-read sequencing, the potential to answer difficult questions regarding problematic regions of the genome and their association with disease has increased. In the current study, we utilize Oxford Nanopore DNA sequencing to analyze retrotransposition, structural variants, and DNA methylation in 18 frontal cortex brain samples from individuals aged 67 to 92 years of age. Neuropathological changes in the aged human brain are common; most individuals over the age of 60 have some degree of deposition of pathological forms of tau [36]. Tau protein accumulates in a diverse group of neurodegenerative disorders, including Alzheimer's disease, that are collectively referred to as "tauopathies." Alzheimer's disease is the most common neurodegenerative disorder, with an incidence of 5% of individuals aged 65-74, 13.1% of individuals aged 75-84, and 33.2% of individuals aged 85+ [37]. Tau pathological distribution in Alzheimer's disease is defined by Braak neurofibrillary tangle staging, compared to a brain void of tau pathology (Stage 0), tau pathology progresses from brainstem structures (Stages a-c) to transentorhinal/entorhinal regions (Stage I), to the primary hippocampus (Stage III) and ultimately, association and primary neocortical regions of the brain (Stages V-VI) [38]. To capture an aspect of the neuropathological diversity of the aged human brain, we included equal numbers of control (Braak 0), mid-stage (Braak III) and late-stage (Braak V/VI) Alzheimer's disease cases among the 18 brains analyzed.

We find that Oxford Nanopore is well-suited for identifying germline and potentially somatic retrotransposon insertions and structural variants in dark loci of the aged human brain. We then describe and contrast 5mC patterns within the aged human brain, focusing on repetitive regions such as satellites, dark regions, and transposable elements. We discover that many structural variants, including retrotransposons, are in close proximity to single nucleotide polymorphisms (SNPs) that have been previously identified as risk variants for neurodegeneration. We further identify significant differences in DNA methylation within specific retrotransposon, repetitive and dark loci in the aged human brain according to Braak neurofibrillary tangle stage.

RESULTS

Nanopore long-read sequencing of DNA extracted from aged human brain

The 18 human frontal cortex samples utilized in this study were age- and sex-matched with an average age of 76.6. DNA was extracted from isolated nuclei of postmortem frontal cortex of six individuals with no clinical or pathological diagnosis of neurodegeneration (Braak 0), six individuals at Braak stage III and six at Braak stage V/VI with a clinical diagnosis of Alzheimer's disease (**Supplemental Table 1**). Sequencing was performed using the Oxford Nanopore Promethion sequencing platform. The mean read quality score was 11, with a n50 of 23.2 kbp and rather shallow genome coverage (average of 7.38X) (**Table 1, Supplemental Table 2**).

	Braak 0	Braak III	Braak V/VI	All	Range
Total # sequenced	6	6	6	18	
Age	76	82.3	71.3	76.6	67-92
% Female	50%	50%	33%	44%	
Postmortem interval	10.33 h	14.67 h	8.75 h	11.89 h	5-24

	Mean	Range
Total bases sequenced	25.5 gbp	14.2-41.6 gbp
Read quality	11.7	11.5-11.9
Read length (kbp)	23.2 kbp	9.66-31.58 kbp
Genome coverage	7.4X	4.1X-12.1X

Table 1 | Brain demographics and sequencing metrics
Sample characteristics and associated sequencing statistics.

Novel retrotransposon insertions in the aged human brain

We first analyzed retrotransposon sequences across all 18 human brain samples using TLDR [39], a tool designed to detect retrotransposon insertion using long-read data, with a focus on elements that are not included in the current human reference genome assembly (hg38). These non-reference retrotransposon insertions would consist of retrotransposons in dark loci of the human genome, retrotransposons that are polymorphic within the human genome that the individuals used to generate hg38 happened to lack, and retrotransposition events that occurred in a given individual in utero or during life. Among the 18 human brains analyzed, we detect 4,645 non-reference retrotransposon insertions, 97% of which contained a tandem site duplication (TSD) (**Supplemental File 1**). TSDs are a signature of retrotransposon insertion that consist of a duplication of the target site of insertion on either side of the inserted element. After further refining this list of potentially active elements to remove elements detected in other human sequencing databases, we detect 1,044 non-reference "novel" retrotransposons present among the 18 brains analyzed, 634 of which were unique to a single individual. Among retrotransposon families, we find that the Alu family of SINE elements composed the most abundant class of retrotransposon insertions not documented in TLDR-provided human databases, with AluYa5 as the most active member. This is consistent with previous work indicating that AluYa5 is the most active retrotransposon subfamily in humans [40] (**Fig. 1A**). Among non-reference retrotransposons detected, most were present in introns, potential enhancer regions or intergenic regions (**Fig. 1B**).

The human endogenous retrovirus, or "HERV" elements, are thought to be transposition-incompetent. HERV-K is the "youngest" HERV, meaning that it integrated into the human genome more recently than any other HERV element [41]. There are many documented HERV-K elements that are variable within in the human population (polymorphic), highlighting the heterogeneity of HERV-K among individuals. We detected five HERV-K elements: HERV-K103 (10p12.1), HERV-K113 (19p12), HERV-K (12q13.2), HERV-K111 (GL000219.1) and HERV-K (4p16.3) that have been previously documented [42] but are absent in the hg38 reference genome. In addition, we discover that HERV elements previously reported as polymorphic at 1p31.1 [43] and 19p12 [44] are HERV-K (**Supplemental File 1**). Interestingly, these elements are in the same cytological band as other HERV-K polymorphs but are composed of different HERV-K elements, highlighting the genetic complexity of these regions between individuals.

LINE and SINE retrotransposons are reported to be transpositionally active in some somatic tissues, including postmitotic neurons [10, 45-47]. We sought to identify potential retrotransposition events in postmitotic cells of the aged human brain following methods similar to Siudeja and colleagues [21]. As transposition is stochastic and the location of a *de novo* insertion in a postmitotic cell is thus unique to that particular cell, we identified retrotransposition events that were A) unique to an individual brain, B) present as a single read, and C) absent in the TLDR reference data set. To decrease false positives, only insertion sites with a read coverage of 10 or greater were included (**Supplemental Fig. 1A, B, C**). 163 insertions in potentially postmitotic cells were detected across 18 brains analyzed, most of which were AluYa5 (**Fig. 1C,**

Supplemental File 1). While most Alu insertions appear to be intact (**Supplemental Fig. 1D**), LINE-1 and SVA insertions in putatively postmitotic cells were short, indicating severe truncation of the inserted element (**Supplemental Fig. 1E, 1F**). Most putatively postmitotic Alu, LINE-1 and SVA insertions were detected in introns (**Fig. 1D**).

Retrotransposon transcripts, particularly those of the ERV class, are significantly elevated in *Drosophila* and mouse models of tauopathy, as well as in human Alzheimer's disease and progressive supranuclear palsy [11, 48, 49], a "primary" tauopathy. While evidence in *Drosophila* suggest that pathogenic forms of tau cause retrotransposition [48], it is currently unknown if retrotransposons mobilize to a greater degree in human brain affected by Alzheimer's disease. We thus compared the extent of retrotransposition in putatively postmitotic cells in brains lacking tau pathology (Braak 0, control) and at early (Braak III) and late (Braak V/VI) stages of tau deposition. While we do not detect statistically different levels of somatic retrotransposition among Braak stages, we note that individuals with the greatest number of Alu (**Fig. 1E**) or L1 (**Fig. 1F**) retrotransposition events in putatively postmitotic cells are at the later stages of Alzheimer's disease. Taken together, our findings from this analysis clearly indicate that key Alu and LINE-1 family members are mobile in the human aged brain; the number of Alu and LINE-1 insertions detected per sample is likely an undercount that results from the relatively low coverage of our nanopore sequencing.

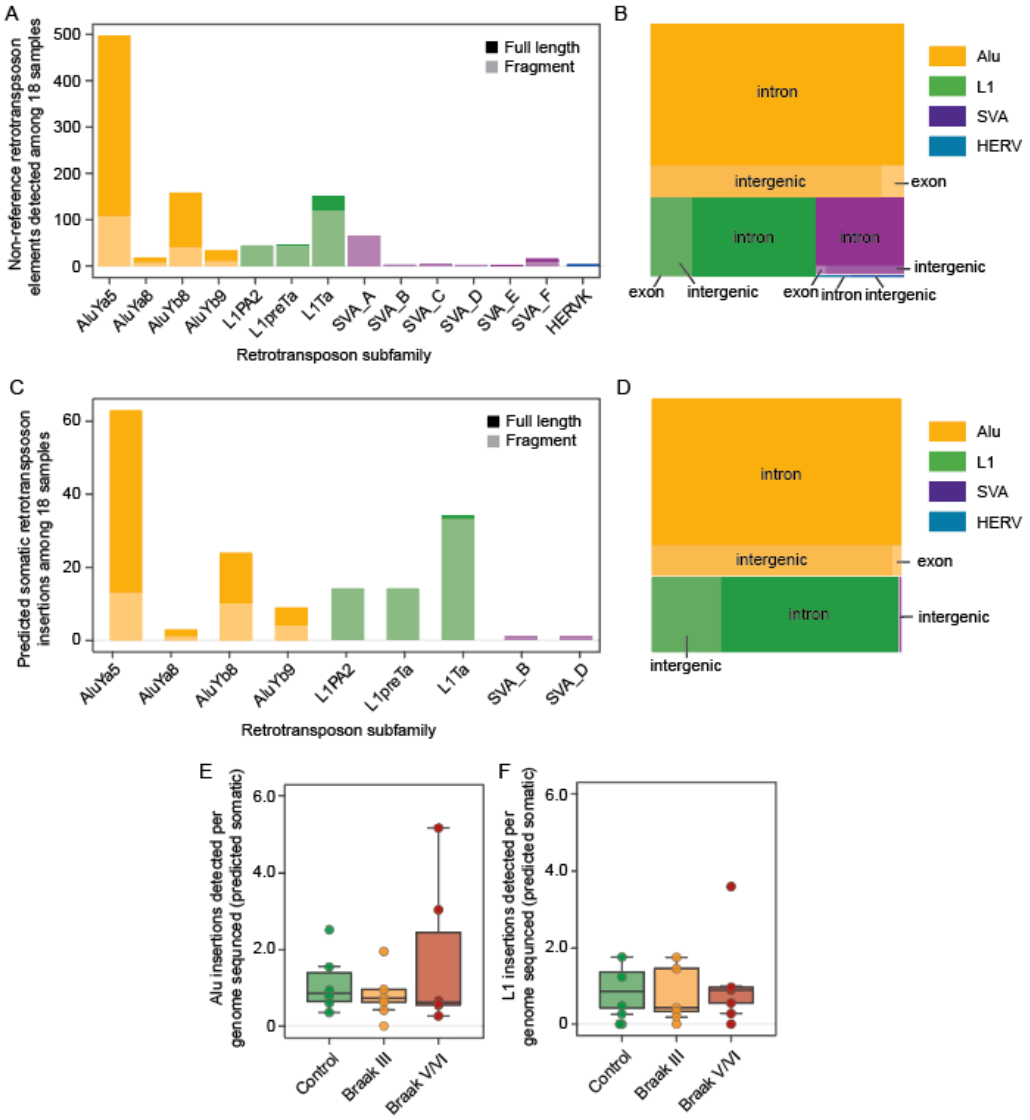


Figure 1 | Non-reference genome retrotransposon insertions in the aged human brain

A) Total counts of all non-reference retrotransposon insertions by subfamily and B) their respective insertion location in terms of gene structure. C) Total counts of all predicted somatic non-reference retrotransposon insertions by subfamily and D) their respective insertion location. Dark shading represents full length insertions and light shading represents fragmented insertions compared to the reference sequence. Boxplots of predicted somatic non-reference retrotransposon insertion counts normalized to sequencing depth for E) Alu and F) L1Hs families.

Structural variants in the aged human brain

We next leveraged our dataset to identify novel structural variants in the aged human brain. SNIFFELS2 [50] was used to detect a total of 47,315 structural variants among the 18 samples (81.4% with a QUAL value of 60). Insertions were the most abundant type of structural variant detected (**Fig. 2A**). Using BLASTN, NCBI and RefSeq and requiring at least 50% alignment, we find that a fifth of non-reference genome insertions are processed pseudogenes, copies of mRNA that have been inserted into the genome. Of these insertions, 37 fully align to full-length processed or fragmented transcripts. For example, we detect a full-length insertion of *HSPE1* (heat shock protein family E (Hsp10) member 1) cDNA into an intron of *PTN* (pleiotrophin). Similarly, a full-length insertion of a *PABPC1* (poly(A) binding protein cytoplasmic 1) cDNA was detected in an intergenic region of chromosome 3. Most samples contain an inserted pseudogene of *RBMX* (RNA binding motif protein X-linked), which is reported to suppress splicing of exon ten of *MAPT* (microtubule-associated protein tau), the gene encoding tau [51], into an intron of *KDM4C* (lysine demethylase 4C) (**Supplemental File 2**).

Structural variants were most commonly detected in introns and exons (**Fig. 2B)**. Length distribution of structural variants is highly variable, with insertions having a larger median length compared to other variants (**Fig. 2C**). We next utilized SNPeff [52] to determine how these variants may affect gene function. Most structural variants were classified as modifier (variants in non-coding genes or intergenic regions) or moderate (non-disruptive variants) impact variants (**Fig. 2D**). We then identified structural variants that are unique to control, Braak III or Braak V/VI samples. To ensure each detected variant was unique to a particular group, we required that the given locus was genotyped in all 18 samples, resulting in 35,231 variants to be analyzed (mean normalized count of 5,835 structural variants per sample (**Supplemental Fig. 2A)**). All samples were normalized for read depth. The number of structural variants detected per sample did not significantly differ between control, Braak III, and Braak V/VI brains (**Fig. 2E**), nor did structural variant length (**Supplemental Fig. 2B**). These results also hold true when Braak III and V/VI were combined into one group and compared to control. The number of insertions and deletions were highly correlated within each sample. In addition to novel variants, our shallow long-read sequencing approach detected structural variants that are present in other databases (**Supplemental File 2**); we detect a variant consisting of an intronic 69 bp deletion (NCBI variation ID: rs2044409223) and 541 bp inversion (gnomAD: INV_CHR21_5C7AFB8C) in *APP,* the gene encoding the amyloid precursor protein, a 10,306 bp tandem duplication (gnomAD: DUP_CHR6_2E62888F) in the intron of spinocerebellar ataxia-1 implicated gene *Ataxin 1* [53], a 70 bp deletion in a 651 bp TA simple repeat region in the first intron of *MAPT* (gnomAD: DEL_CHR17_14E37638, DEL_CHR17_ABF8637E), as well as other variable nucleotide tandem-repeat expansions known to be associated with neurological disorders, clearly demonstrating the utility of low-coverage nanopore long-read sequencing for repeat expansion analyses.

We next identified non-reference structural variants that fall within the 6,054 dark loci of the human hg38 reference genome [18]. Among the 18 brains analyzed, we detect 1,123 insertions in the dark regions of 787 genes, with 60 of those insertions occurring in dark exons. 1,781 deletions associated with dark loci were detected among the 18 brains analyzed. Interestingly, 23 Alzheimer's disease risk genes contain dark regions with previously-undocumented structural variants, some of which are in close proximity to SNPs defined as risk variants for neurodegenerative diseases based on ADSP (Alzheimer's Disease Sequencing Project) [54] or the European Bioinformatics Institute genome wide association study (EFO ID: EFO_0005772). For example, three dark region-associated insertions and a retrotransposon insertion were detected near or overlapping Alzheimer's disease-associated SNPs within *ABCA7* (ATP binding cassette subfamily A member 7) (**Fig. 2F**), which encodes an ABC transporter that regulates lipid metabolism [55, 56], amyloid

processing and clearance [57-59]. A deletion and an insertion were detected within a dark region of an intronic SVA retrotransposon within *VIPR2* (vasoactive intestinal peptide receptor 2), near an Alzheimer's disease-associated SNP that overlaps a deletion present in 10 out of 18 brains analyzed (**Fig. 2G**). Similarly, most brains analyzed have a deletion overlapping a dark intronic AluYa5 element within *CHD2* (chromodomain helicase DNA binding protein 2) in close proximity to an Alzheimer's disease-associated SNP (**Fig. 2H**) (**Supplemental Table 3**). Dark region insertions were also detected near SNPs associated with amyotrophic lateral sclerosis, multiple systems atrophy, multiple sclerosis, and spinocerebellar ataxia 31. Many non-reference insertions and deletions within dark loci were shared among all samples, likely reflecting common circulating variants, while others were only detected in single samples, reflecting less common variants, variants that occurred in the germline, or variants of somatic origin. The overlap of structural variants within dark loci of genes associated with human disease highlights the value of approaches that can resolve complex regions of the genome.

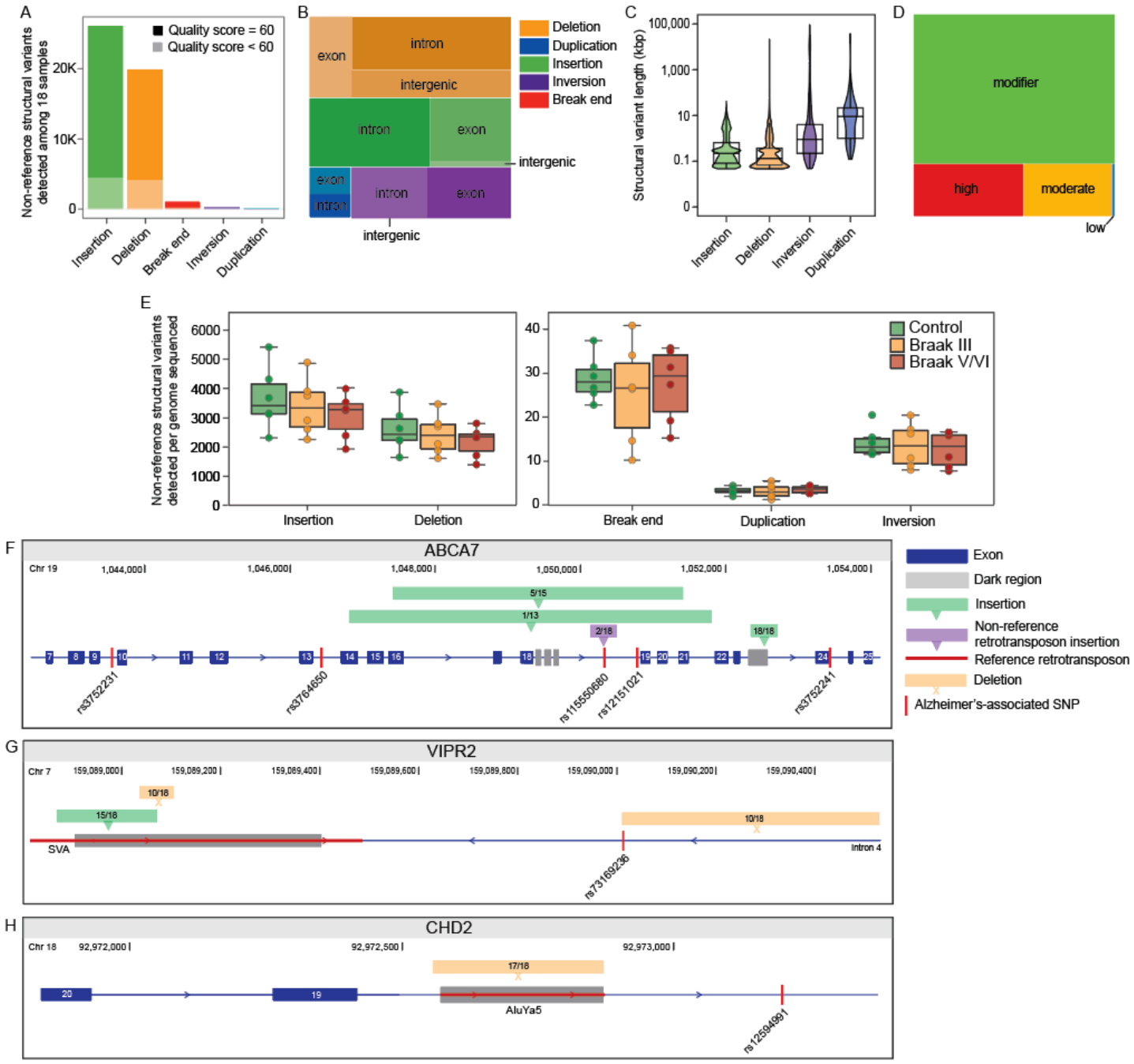


Figure 2 | Non-reference structural variants in the aged human brain

Total counts of all non-reference structural variants. Dark shading represent variants with the highest quality score (60) while lighter shading represents lower scores. B) Treemap depicting respective location of structural variant according to gene structure. C) Violin plot of structural variant lengths between 500 bp and 100 mbp. D) Treemap of predicted effects of structural variants on gene function. E) Total structural variants per brain sample detected by Braak stage, normalized by sequencing depth. F, G, H) Novel insertions and deletions detected in close proximity to Alzheimer's disease associated SNPs. Ratios indicate the proportion of human brains carrying the given variant.

DNA methylation within genes of the aged human brain

We next analyzed DNA methylation patterns (5mC) among our 18 human brain samples. We first utilized Nanopolish to determine the fraction of the 29 million 5mC modified CpG sites documented in the hg38 human genome reference that were detected with confidence via nanopore sequencing. A sample from an individual with a previous cancer diagnosis was an outlier in terms of 5mC methylation and was thus removed from the analysis (**Supplemental Fig. 3A**). We focused on CpG sites with a 5mC call in at least three sample, resulting in 23 million CpG sites for analysis. This is less than the previously estimated number of CpG sites in the hg38 human genome reference [60], likely reflecting the relatively low coverage of our nanopore sequencing.

To determine if nanopore-based methylation analysis aligns with previous work reporting the density of 5mC at CpG sites within the human genome, we identified 5mC patterns of all called CpG dinucleotides and analyzed their localization to CpG islands, CpG shores, CpG shelves or inter-CpG regions (**Fig. 3A**). Compared to the levels of CpG methylation genome-wide, we find that CpG islands and, to a lesser extent, shores have low methylation frequencies compared to the rest of the genome (**Fig. 3B**), in line with previous findings in humans, non-human primates and mice [61].

We next analyzed 5mC changes within promoters of control, Braak stage III and Braak stage V/VI brains. We detect differential methylation of 20,352 unique promoters of 7,852 genes in Braak stage III compared to control, while 19,638 promoters of 5,765 genes were found to be differentially methylated at Braak V/VI (**Supplemental File 3**). We observe a subtle change in the bimodal distribution of 5mC modifications in differentially methylated promoters, in line with the promoter element MDS (multidimensional scaling) plot [62]. Differentially methylated regions in brains at Braak III have higher densities at 0.25 and 0.75 compared to control, while brains at Braak V/VI shift towards no methylation (0) or complete methylation (1) (**Fig. 3C)**. It is important to note that changes in methylation could represent changes in cell composition between Braak stages rather than (or in addition to) changes in transcriptional control. Indeed, cell type proportion changes in in Alzheimer's disease brain tissues are well-documented [63, 64]. While promoter methylation differs among control, Braak III, and Braak V/VI brains, we identify over 7,072 hypomethylated promoters that are unique to brains at Braak III related to pathways regulating phospholipase D, FC gamma receptor mediated phagocytosis, gonadotropin-releasing hormone (GnRH) signaling pathway, Rap1 signaling pathway and actin cytoskeleton regulation (**Fig. 3D**). Almost 3,000 hypermethylated promoters are shared between Braak III and Braak V/VI brains, related to pathways involving calcium and oxytocin signaling, focal adhesions, phospholipase D and glutamatergic synapses (**Fig. 3E**).

We then focused on DNA methylation within "dark" regions of the genome to analyze changes in loci that are not easily accessible with short-read sequencing and probe-based methods. We detect 1,150 and 1,333 differentially methylated regions in Braak stage III versus control and in Braak stage V/VI versus control, respectively. We identified 29 differentially methylated 5' untranslated regions (UTRs) in Braak stage III and 37 differentially methylated 5' untranslated regions in Braak stage V/VI (**Supplemental File 3**). For example, we found hypomethylation of dark loci in the 5' UTR of Alzheimer's disease risk genes *AMY1A* and *FMR1*. Additionally, we identify four differentially methylated regions in Braak stage III versus control within loci previously identified as Alzheimer's disease risk genes (*ABACA7*, *AMY1A*, *CHRFAM7A*, *CR1*). Changes in methylation of dark loci in *AMY1A*, *CHRFAM7A*, and *CR1* were also observed at Braak stage V/VI

(**Supplemental File 3**). The detection of a large number of differentially methylated dark loci, including transcription-associated 5' untranslated regions, highlights the importance of dark loci analyses in epigenetic studies.

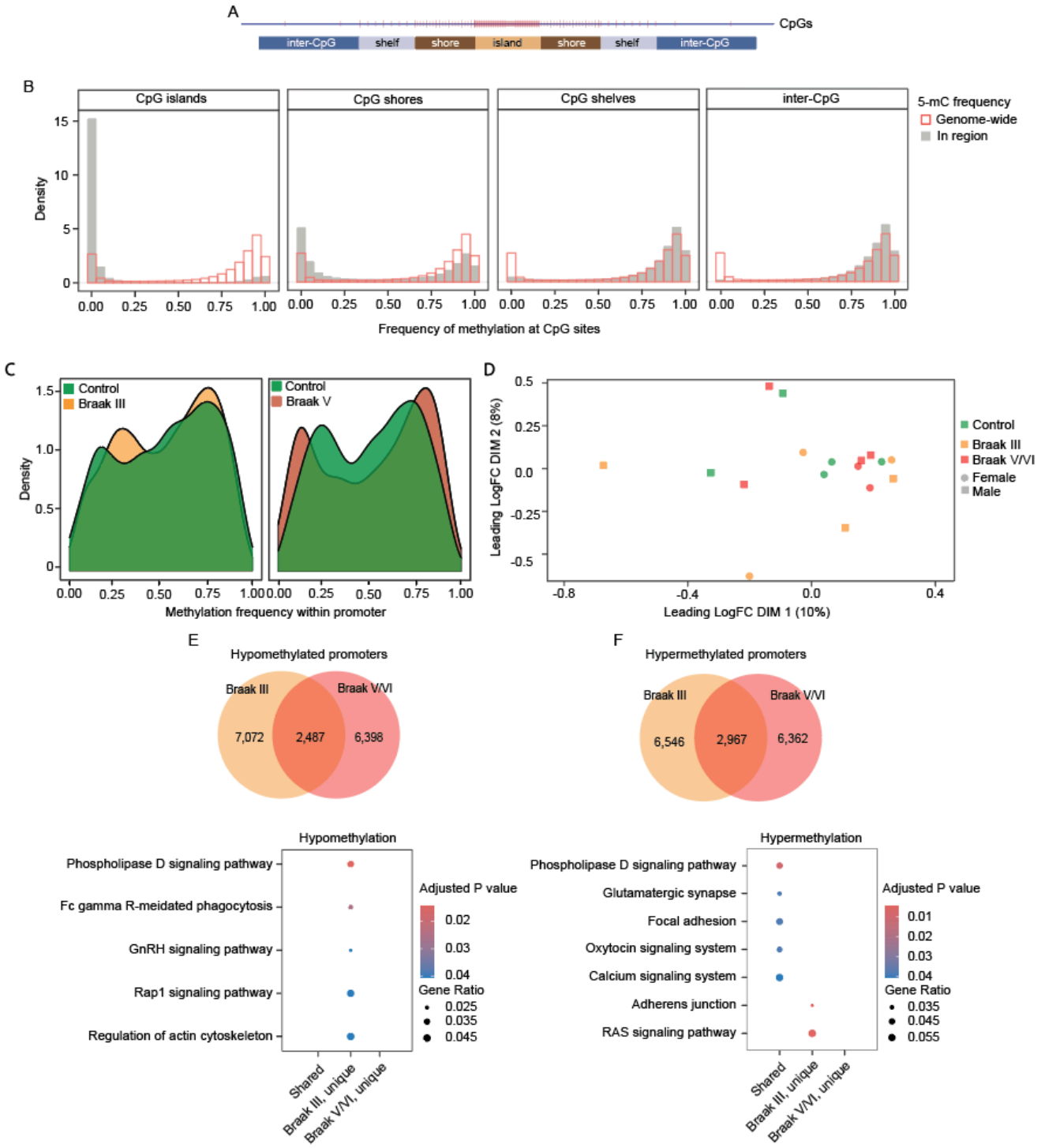


Figure 3 | Nanopore-based methylation analyses in the aged human brain

A) Visual representation of CpG islands. B) Histograms of mean methylation density (Y-axis) at different methylation frequencies (X-axis) in CpG islands, shores, shelves, and intergenic regions (gray bars) vs. distribution over all regions plotted (red). C) Density of differentially methylated regions within promoters (Y-axis) at mean methylation frequency in Braak III vs. control and Braak V/VI vs. control (X-axis). D) MDS plot of likelihood methylation ratios of promotor region CpG sites. E, F) Venn diagram of promoter methylation and respective comparison of functional KEGG profiles of promoters that are shared or unique to Braak III or Braak V/VI.

DNA methylation across chromosomes and repetitive regions in the aged human brain

We next leveraged our long reads to analyze 5mC across entire chromosomes. As repetitive regions such as satellites account for the majority of the genome, we determined the methylation frequency mean across repetitive satellite regions of all samples combined. Most regions showed highly variable amounts of methylation (**Supplemental Fig. 4A**), potentially reflecting locus-specific functions of repeats. We next plotted the rolling average of change in methylation of statistically significant differentially 5mC methylated regions across a meta-chromosome (not including sex chromosomes) across Braak stages. We find that DNA of brains at Braak III have generally less methylation (**Fig. 4A**), while DNA from brains at Braak V/VI is generally hypermethylated compared to control (**Fig. 4B**) across the p-arm, q-arm, and centromeric regions. Findings at Braak III are in line with previous observations of decreased DNA methylation in cortical tissue of Alzheimer's disease samples measured by 5mC immunoreactivity [30, 31], although these studies utilized brains at later Braak stage for analysis. Our findings at Braak V/IV are in line with studies leveraged an array-based approach that report DNA hypermethylation in Alzheimer's disease [33-35].

Based on what appeared to be a conserved directional change of differential DNA methylation at the centromere at both Braak III and Braak V/VI stages compared to control, we performed a permutation test to determine if differential methylation in the centromere is higher than expected by chance. Indeed, we find that centromeric 5mC is particularly affected at Braak III and Braak V/I compared to control (**Fig. 4C, D**). Among repetitive elements, we find that alpha satellites are significantly depleted in 5mC at Braak III compared to control (**Fig. 4E, Supplemental Fig. 4B**), in line with the enrichment of differentially methylated regions at the centromere.

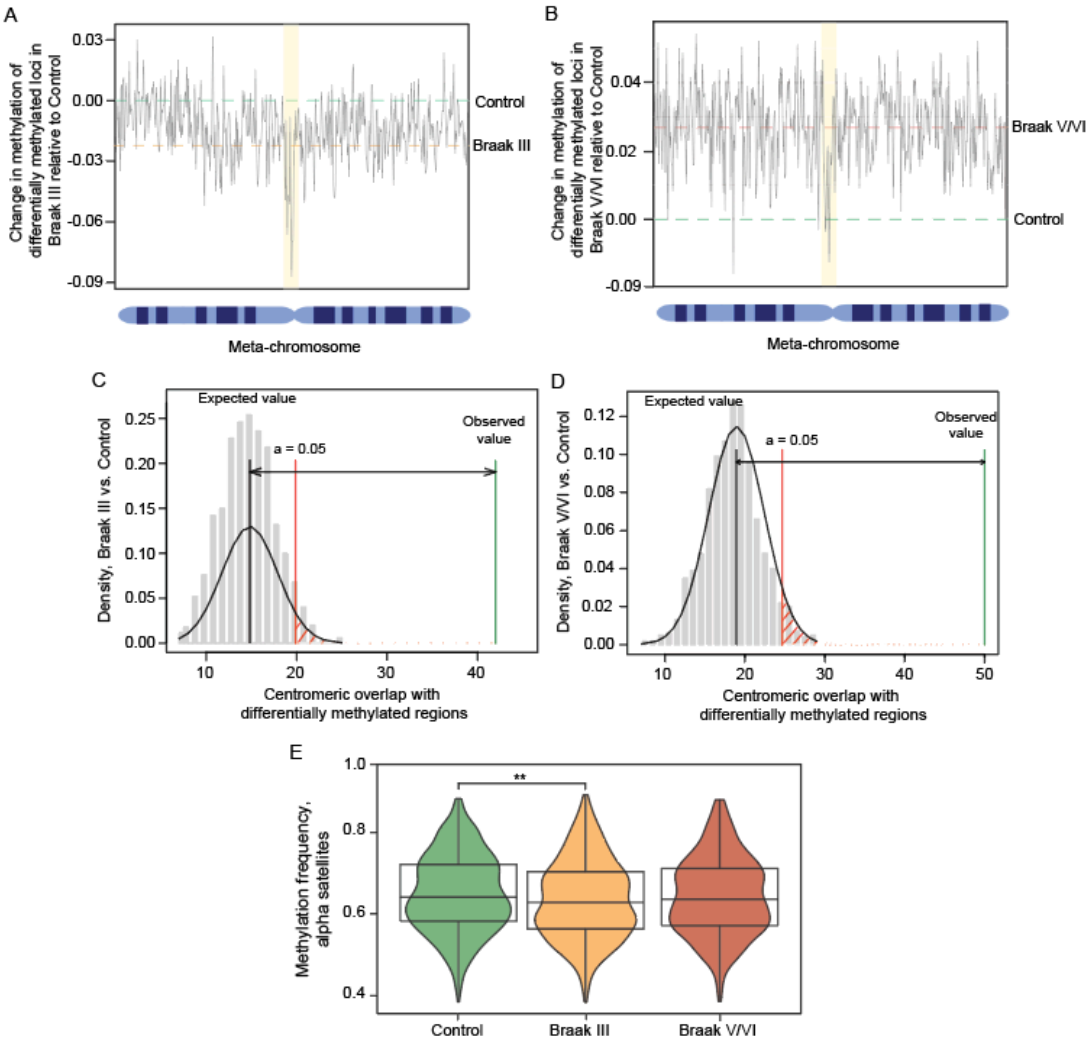


Figure 4 | DNA methylation patterns across chromosomes in Alzheimer's disease

Rolling mean of the difference in methylation values of differentially methylated regions across all autosomes and normalized for chromosome length for A) Braak III vs. control and B) Braak V/VI vs. control. Orange horizontal dashed represent the average change across the whole meta-chromosome for control and respective Braak stage. Permutation test of differentially methylated regions localized to the centromere in C) Braak stage III vs. control and D) Braak stage V/VI vs. control (P=0.0001). E) Mean methylation frequencies for alpha satellite loci with at least 10 CpG calls. Dunn's nonparametric pairwise multiple-comparison test. **Holmes P adjusted value = 0.0001.

DNA methylation within transposable element sequences of the aged human brain

While previous studies report widespread changes in 5mC at different stages of Alzheimer's disease in the frontal cortex, no work to date has focused on 5mC CpG methylation within repetitive regions of the genome in brain of aged individuals or those affected by Alzheimer's disease. We find that retrotransposons are heavily methylated in the aged human brain, with a mean methylation value exceeding 0.75, with the exception of HERV-K elements (**Fig. 5A, Supplemental Fig. 5A-F**). Young Alu and LINE retrotransposon family members have generally increased levels of 5mC compared to older Alu and LINE family members, perhaps reflective of the need to repress their transpositional capacities. LTR and HERV families, however, feature less methylation of younger elements. We detect promoter hypomethylation of an intact L1-Ta element in chromosome 13 (13q.12.3) that was previously reported to be a source of somatic retrotransposition in the developing hippocampus [7, 65], suggesting that this element may also be a source of active LINE transcripts and retrotransposition in the aged human brain (**Fig. 5B).**

An MDS plot of CpG methylation frequency in autosomal repetitive elements revealed no clustering based on sex or Braak stage (**Supplemental Fig. 6A**). At Braak III versus control, a total of 102,726 differentially methylated repetitive regions were detected, 7.6% of which have a change in methylation frequency of 0.2 or more. 98,175 differentially methylated repetitive regions were detected in Braak V/VI versus control, 6.67% of which have a difference in methylation frequency of 0.2 or more (**Supplemental File 4**). The distribution of 5mC modification in differentially methylated repetitive elements shifted in opposite directions according to Braak stage, with a shift towards less methylation at Braak III and a shift towards increased methylation at Braak stage V/VI (**Fig. 5C,5D**).

While we do not detect differences in 5mC methylation of retrotransposon subfamilies between Braak stages when retrotransposons are grouped *en masse*, similar to previous studies (**Supplemental Fig. 6B-E**), we readily detect statistically significant hypomethylation in the promoter region of individual L1Hs loci such as an intergenic L1-Ta-1 (transcriptionally active) (7q34) and an intronic L1-Ta in *SPOCK3* (4q32.3) at Braak stage III compared to control (**Fig. 5E, 5F, Supplemental File 4**). Our approach also reveals incidents of LINE hypermethylation at Braak III such as an older full length non-intact L1Hs (**Supplemental Fig. 6F**).

We identify several differentially methylated LTR5Hs loci in both provirus and solo form at Braak III and V/VI. Interestingly, we detect seven examples of 5' LTR5Hs hypomethylation adjacent to proviral HERV-K elements (many of which encode intact pro, pol and/or gag ORFs) (**Fig. 5G, 5H**). Several of these HERV-K elements (ex: HERV-K107 (7p22.1), HERV-K108 (6q14.1), HERV-K109 (6q14.1)) have been previously implicated in the pathogenesis of other diseases [66, 67]. We also detect differential 5mC methylation of several solo LTR5Hs and Alu elements that are predicted GeneHancer regions (**Supplemental Table 3**). This analysis highlights the importance of analyzing retrotransposon methylation on a single loci level rather than *en masse*.

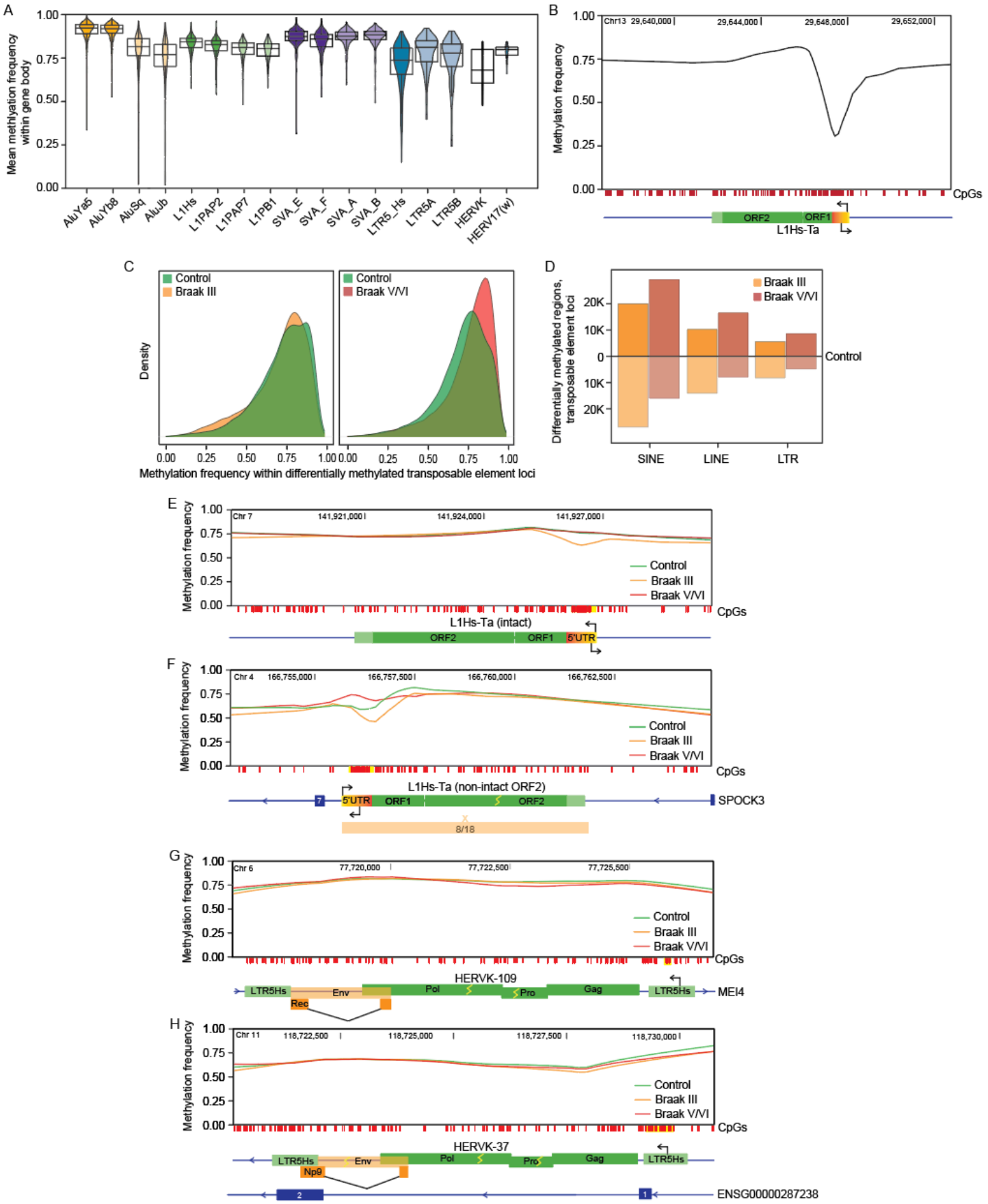


Figure 5 | Retrotransposon DNA methylation in the aged human brain

A) Mean methylation frequencies for retrotransposon loci with at least 10 CpG calls with a minimum length of 290 bp in Alu, 5.9 kbp in L1, 1 kbp in SVA, 900 bp in LTR and 6 kbp in HERV-K internal regions. Dark shading represents differential methylation in younger

elements, light shading represent differential methylation in older elements. B) Methylation profile of the body of a L1-Ta (13q12.3) retrotransposon with promoter hypomethylation. C) Density of differentially methylated regions within repetitive elements (Y-axis) at mean methylation frequency in Braak III vs. control and Braak V/VI vs. control (X-axis). D) Bar plot of hypermethylated and hypomethylated repetitive regions in Braak III vs. control and Braak V/VI vs. control. E, F, G, H) Methylation frequency of CpG sites across L1-Ta and HERV-K-proviral loci. Statically significant changes in CpG methylation detected in promoter regions are highlighted in yellow; yellow lightning bolts denote truncated ORFs.

METHODS

Sample collection

18 postmortem frontal cortex samples were selected to capture equal numbers of Braak stages 0, III and V/VI with an average age of 76.6 years (68 – 92). Samples were Caucasian and were relatively balanced for sex and age with respect to Braak group. Samples were provided by Dr. Dennis Dickson of the Mayo Clinic Brain Bank or acquired from the NIH Neurobiobank.

DNA processing and read alignment

Approximately 80 mg of tissue was homogenized with a Dounce and then centrifuged through a sucrose gradient to isolate nuclei. High molecular weight DNA was extracted from nuclear isolates using a standard phenol chloroform protocol. Library preparation was performed by the University of California Davis genomics core using the SQK-LSK 110 kit (ONT) with shearing to obtain 10 kbp length reads. Libraries were then multiplexed for three samples per PromethION flow cell and sequenced. Nanopore-derived reads were aligned to the hg38 genome (UCSC) using MiniMap2 [68] (nanopore sequencing specific parameters) to obtain bam files. Mosdepth [69] was used to determine average coverage per sample.

Non-reference retrotransposon insertion calling

All bam files were analyzed simultaneously with TLDR (default parameters) to reduce false negative calls in any single sample. This resulted in a retrotransposon file with calls across all 18 samples. Insertions called by TLDR were then filtered to retain insertions with a tandem site duplication and a PASS quality score to eliminate low confidence/quality insertions. We define a putatively somatic insertion in a post-mitotic cell as a call that is only supported by a single insertion-spanning read (singleton) in a single individual who has at least 10 additional reads lacking the insertion.

Structural variant calling

Individual bam files were subjected to structural variant calling using SNIFFLES2 (default parameters) and subsequently combined into a multi-sample variant call file with the SNIFFLES2 multi-sample calling module. The variant call file was then filtered to retain variants 50 bp or larger. The locations and potential effects of the structural variants were determined using snpEFF. Insertion nucleotide sequences were extracted from the variant file and searched against the NCBI nucleotide and RefSeq [70] databases using BLASTn [71] (default parameters) within the Galaxy platform [72]. Hits were retained if the query insertion was covered by 50% or more of a subject sequence in the database. To compare structural variants across Braak groups, we required that the region of a given structural variant was sequenced across all 18 samples. A variant that was only found in samples belonging to a single Braak group was considered unique to that group.

Normalization and statistical comparison of non-reference retrotransposon insertions and structural variants

Insertions or variants unique to a Braak group were compared by ANOVA. To account for differences in sequencing coverage, insertion counts were normalized by dividing the raw count value by the total number of diploid human genomes sequenced (gbp sequenced divided by 6.4 gbp) (**Supplementary Table 2**).

Characterization of retrotransposon insertions and structural variants

Coordinates from the TLDR or SNIFFLES2 output were processed with the R GenomicRanges package [73]. Overlap of insertions and variants with regions (genic/non-genic, dark regions, tandem repeat regions) were determined with AnnotatR [74]. Bedtools [75] was used to identify variants near single nucleotide polymorphisms. Gene graphs were first generated by IGV [76] and then modified for clarity. L1Hs subfamilies were identified by aligning L1Hs loci to the L1.2 (active LINE-1: GenBank accession number M80343) and identifying the L1-Ta diagnostic ACA and G nucleotides in the 3' untranslated regions. HERV-K ORFs were determined by aligning HERV-Ks of interest to the consensus HERV-K sequence (DFAM ID: DF0000188). Potential HERV-K non-reference insertions were manually analyzed for coding potential and completeness compared to the full length HERV-K consensus sequence. 24 (attributable to five insertions based on clustering of insertion calls) HERV-K calls containing non-hg38 reference genome insertions of 5,000 bp or more were detected. To identify other HERV insertions, SNIFFLES2 identified insertions were searched against a custom HERV internal and LTR5 database with sequences from RepBase [77] using BLASTn [71] (default parameters) within the Galaxy platform.

CpG 5mC methylation analysis

5mC modifications were called using the Nanopolish [78] call-methylation pipeline using single bam files described above. Alignments with a mapping quality score of 20 or more were included to calculate log likelihood methylation ratios. Nanopolish-generated log-likelihood ratio values were then converted to methylation frequencies with accompanying Nanopolish scripts. Samples were then analyzed with a MDS plot to visualize sample clustering. Sample CTRL-1 who also had a previous cancer diagnosis, was identified as an outlier and was thus removed from further analysis. The average methylation frequencies across all samples were annotated as within genes, CpG islands or repeat regions (defined by UCSC RepeatMasker) using annotatR. Methylation graphs of genomic loci were created with methylkit [79] and NanomethViz [80].

Differential methylation analysis

Differentially methylated CpG sites were identified using DSS [81], a tool that models methylation distributions with a beta-binomial model to compute changes in methylation at the loci and regional level, taking sequencing depth into account. Differentially methylated CpG sites were considered significant if they reached a False Discovery Rate (FDR) of 0.05 or less. DSS was then used to identify and aggregate neighboring CpG sites into regions with differential methylation (DMRs) [82]. Comparisons were made for Braak stage III vs. control **(Supplemental File 7)** and Braak stage V/VI vs. control **(Supplemental File 8)**. Alu elements (AluYa8b, AluYa5, etc.) and LTR5Hs with the largest changes in methylation between groups were then manually inspected using the GeneHancer Regulatory Elements and the Gene Interaction table [83] within the UCSC genome browser [84] to determine if any overlap occurred. For inclusion in methylation average comparisons of retrotransposons *en masse*, individual loci were required to have 10 or more called CpG sites. To exclude highly fragmented retrotransposon loci, a length threshold depending on the type of retrotransposon was implemented: HERV-K (> 6,000 bp), LTR5_Hs/A/B (> 900 bp), LINE1-Hs (> 5,900 bp), AluYa5/Yb8 (> 280 bp) and SVA-E/F (> 1,000 bp) [39]. Other repetitive elements such as centromeric satellite regions were required to have 10 or more CpG sites called within a given region. For analysis of DMRs in promoter regions, a promoter region was only counted once in each direction even if there are multiple differentially methylated regions within the region.

Enrichment analysis and statistical tests

A permutation test (regioneR function permTest [85]) was used to determine if there are more differentially methylated regions within centromeres (defined by RepeatMasker) that what would be expected by chance. The mean change in methylation in differentially methylated regions across a meta-chromosome was calculated following previously described

methods [86, 87]. Gene lists were analyzed for KEGG pathway [88] enrichment using ClusterProfiler [89]. All graphs were generated with R packages ggplot2 [90], TreeMap or adapted from IGV [76].

DISCUSSION

Retrotransposon insertions, complex structural variants and DNA modifications within repetitive regions have long been understudied due to the technical difficulty of their detection with traditional sequencing methods. The advent of long-read sequencing technologies such as Oxford Nanopore and associated bioinformatic tools allow us to overcome previous limitations. Here we demonstrate the feasibility of Oxford Nanopore sequencing to detect potentially pathogenic genomic elements in problematic regions of the human genome using shallow-coverage sequencing and a relatively small sample size.

While some polymorphic retrotransposon insertions are included in databases such as gnomAD [91] and have been identified in genome wide association studies [92], widespread identification of such elements is limited by their repetitive nature. Specialized tools are thus necessary to accurately detect novel transposition events, limiting their discovery largely to studies specifically focused on retrotransposons [93, 94]. We find that non-reference retrotransposon insertions can be readily called and annotated using shallow coverage long-read sequencing. For example, a specialized retrotransposon insertion caller (TLDR) and general variant caller (SNIFFLES2) can identify multiple polymorphic HERV-K elements and their nucleotide sequences within our dataset. We find that most non-reference and potentially somatic retrotransposon insertions in DNA isolated from the human brain are Alu elements and occur in intronic regions.

We extended our analyses to determine if the number of retrotransposon insertions differs based on Braak stage. The upregulation of retrotransposon transcripts in Alzheimer's disease and progressive supranuclear palsy, along with increased transposition in a *Drosophila* model of tauopathy suggest that transposable elements may mobilize in human Alzheimer’s disease [11, 48, 49, 95-98]. While our finding that potentially somatic insertion rates were not significantly different between human brains at Braak 0, Braak III, and Braak V/VI, we do find that two individuals at Braak V/VI have elevated insertions compared to other groups. While all data presented was normalized to sequencing coverage, we note that a higher number of insertions were identified with increased sequencing coverage. While we thus recommend deeper sequencing coverage and a greater number of biological replicates in future studies, we recognize that this approach is costly. We nevertheless speculate that complete retrotransposition is not a driving factor of neuronal death in the adult human brain affected by Alzheimer’s disease, and that the negative effects of retrotransposon activation in Alzheimer’s disease and related tauopathies are in greater part due to the production of neuroinflammatory double-stranded RNA (dsRNA) [99], extrachromosomal circular DNA, retrotransposon-encoded viral like proteins, and/or DNA double strand breaks that result from attempted insertions. It will also be of interest to apply long-read sequencing-based approaches to determine the degree of HERV-K polymorphism [41, 42, 100-104] among different cohorts. Future technological improvements in base calling accuracy will further allow investigators to determine whether specific HERV-K loci vary in their potential to produce functional proteins among different individual, ethnic and disease backgrounds [105].

While previous analyses of structural variants in brain cells, often relying on analysis of DNA copy number and DNA content, have revealed alterations that accumulate with age, these approaches can miss complex variants entirely, hindering a comprehensive understanding [106]. This has resulted in a lack of information regarding structural variants and/or pathogenic somatic retrotransposon insertions that are in linkage disequilibrium with SNPs. Among structural variants identified in our samples, insertions were most common and were most often present within introns. Interestingly, a large proportion of these variants were predicted to have a high impact on gene function. This analysis also enabled the analysis of tandem repeat lengths in genes implicated in familial neurodegenerative disorders such as Huntington's disease and spinocerebellar ataxia type 1 with high confidence without the need for complementary analyses. Given the potential impact of structural variants in Alzheimer's disease, we analyzed their changes across Braak stages. No statistically significant differences were observed among various types of variants. This finding could be due to

our depth of nanopore sequencing, as we only detected a median of ~5,000 structural variants per genome whereas others report a median of ~7,000 structural variants per genome [86].

We next turned our attention to the analysis of structural variants within dark regions of genes that have been previously associated with Alzheimer's disease. We identified several insertions (ranging in length from 300 bp to 7 kbp) in dark loci in close proximity to Alzheimer's disease-associated SNPs in the *ABCA7* gene. *ABCA7* encodes a transporter that regulates lipid release, lipid trafficking, and phagocytosis, and is expressed among different cell types of the brain [107]. Interestingly, all of these insertions overlap with a previously-described variable nucleotide repeat expansion that is highly correlated with the presence of a common high-penetrant risk SNP (rs3764650) for Alzheimer's disease [108]. We also detected overlap between an Alzheimer's disease risk SNP (rs115550680) and a polymorphic AluYa8 element, and identified several challenging-to-detect variants within other known risk genes. While the *ABCA7* gene, particularly in African Americans [109], harbors rare and common variants linked to either increased or decreased risk of Alzheimer's disease, various studies are contradictory on which specific variants confer risk vs. protection [107]. Multiple studies have investigated the potential mechanism of *ABCA7* involvement in Alzheimer's disease, and find that knockout of *ABCA7* results in accumulation of amyloid beta peptides and abnormal function of microglia in multiple *APP*-based mouse models of Alzheimer's disease [58, 59, 110]. These findings highlight the comparatively limited sensitivity of traditional sequencing techniques for resolving complex variants such as variable nucleotide tandem repeats. Future studies leveraging targeted sequencing methods such as Cas9, PCR, and hybridization capture focusing on these loci in a larger set of individuals would enable comprehensive population-level variant analysis [111].

While the role of DNA methylation in brain plasticity and cognition is well-established, its impact on repetitive regions that are enriched for dark loci remains largely unexplored. In the current study, we analyze average 5mC modifications within dark loci of the healthy and Alzheimer's disease-affected aged human brain. Several studies have investigated 5mC modification in Alzheimer's disease brain samples across various brain regions, reporting consistent hypermethylation of specific genes. Global 5mC modification patterns, however, are inconsistent across studies, highlighting the influence of factors such as tissue quality, cellular heterogeneity, disease stage and technical limitations [30, 33-35]. Our nanopore-based methylation calling approach generally aligns with previous bulk DNA methylation studies of DNA isolated from frontal cortex of Alzheimer's disease brains [33, 112-114]. We identified differentially methylated promoters unique to Braak stage, including some within dark loci that are missed using other approaches. We find that differentially methylated promoters among Braak stages are associated with pathways such as RAS signaling, phospholipase D signaling and MAPK signaling.

To study the methylation landscape in Alzheimer's disease at a chromosomal scale, we visualized statistically significant methylation differences between Braak stages and control on meta-chromosomes to observe regional variations. Intriguingly, we detect methylation dips near centromeres accompanied by enrichment of differentially methylated regions in these loci. Moreover, significant alterations in bulk 5mC modifications of alpha satellites, repeats that are enriched within centromeres, were identified. Methylation changes within repetitive elements carry potential for issues such as inflammation and genomic instability that can result from the formation of dsRNA that activate the innate immune system, as well as DNA damage through retrotransposition-induced double stranded DNA breaks [115-118].

Venturing beyond easily sequenced regions of the human genome, we next investigated methylation patterns in retrotransposons within the aged human brain. Consistent with prior observations, we find that methylation levels correlate with the evolutionary age of retrotransposons, with younger elements exhibiting higher methylation frequencies than older elements [119-121]. We further delved into the L1Hs family, known for its ability to autonomously retrotranspose in humans. Notably, we identified an L1-Ta element with a hypomethylated promoter, suggestive of activity, echoing a previous finding in a hippocampal sample from a 51-year-old woman [39]. Promoter hypomethylation of an intact L1-Ta element across our cohort of brains suggests that this element may be involved in normal aging and/or basic cellular processes [122].

Our analysis revealed intriguing Braak stage-specific methylation patterns: brains at Braak III exhibited hypomethylation of retrotransposons, while brains at Braak V/VI featured an unexpected shift towards hypermethylation of repetitive regions. This observed discrepancy may be a consequence of differences in cell type proportions over the course of Alzheimer's disease progression [123, 124]. Consistent with previous analyses in blood samples from patients with Alzheimer's disease [125, 126], we did not detect significant differences in global methylation patterns of retrotransposon subfamilies when analyzed *en masse*. Analysis of specific methylation changes within L1-Ta, Alu, and LTR5Hs elements (previously implicated as alternative promoters [127-130]), however, revealed significant differences among Braak stages. Our shallow nanopore sequencing proved sufficient to accurately capture polymorphic retrotransposon element methylation frequencies at the single-loci level, highlighting the potential loss of valuable information when analyzing data at the grouped subfamily level.

Despite documented changes in retrotransposon transcription in Alzheimer's disease, pinpointing source loci for transcriptional changes is not currently possible with traditional RNA sequencing approaches. Employing nanopore technology, we identified individual retrotransposon promoter loci harboring altered 5mC modification, suggesting that they may serve as loci for elevated transcription and/or retrotransposition in disease cases. Notably, brains at Braak stage V/VI exhibited hypomethylation within the 5' LTRs of HERV-K107, -108, and -109, previously linked to cancer development [131-133]. Some HERV-encoded proteins such as the HERV-K110-derived envelope protein have been implicated in other diseases, including rheumatoid arthritis [134]. HERV-K101 expression has been observed in teratocarcinoma and linked to schizophrenia enhancer function [135]. Intriguingly, we also identify several HERV-K loci with intact envelope protein ORFs that have previously been implicated in ALS [136]. We also identified hypomethylated L1-Ta loci that have the potential to encode reverse transcriptase. The detection of both hypermethylated and hypomethylated retrotransposons suggests that transcripts originating from specific loci, rather than subfamilies of transcripts, may contribute to Alzheimer's disease pathology through altered expression patterns. This hypothesis is supported by evidence that envelope proteins from different HERV-K loci have distinct immunological effects, as exemplified by HIV infection [137]. Future studies leveraging long-read RNA sequencing would facilitate the identification of individual HERV, L1, and Alu loci involved in the pathogenesis of Alzheimer's disease.

While our study revealed many new insights into retrotransposon biology, polymorphic structural variants and methylation within dark loci of the human genome, our approach is not without limitation. Our shallow sequencing coverage and small sample size limits our ability to detect rare somatic variants. While we fell short of the 30x coverage standard [138], our analyses demonstrate the potential of low-coverage nanopore sequencing to identify candidate variants for targeted testing in larger cohorts. Our strategy is also cost- and computational demand-effective, making it feasible for smaller studies. The incomplete nature of the hg38 reference genome, which does not capture all sequences of all chromosomes, is a further limitation. Many centromeric and telomeric repeats, for example, are predicted models rather than complete sequences and result in the misalignment of reads and missed regions [139-141]. The recent publication of the Telomere-To-Telomere Consortium genome, with increased development of associated annotations and new tools, should alleviate these issues. Our use of bulk tissue sequencing is an additional limitation, as it does not resolve whether changes occur within specific cell types. While single-cell long-read sequencing would address this issue, this approach utilizes an amplification step that prevents DNA modification calling [142]. Adapting existing cell composition correction methods used in methylation array projects could potentially hold promise for nanopore-derived methylation data. These methods can be time-consuming without cloud or cluster computing resources. Small studies often lack power to detect changes in diverse groups, raising concerns about feasibility in basic science and clinical research. Advancements in sequencing accuracy and computational efficiency (speed and storage) should overcome these challenges, as will large-scale collaborations such as a recent study employing a novel pipeline for structural variant detection in multiple frontal cortex brains [143]. Despite the limitations of long-read sequencing with shallow coverage and small sample size, our study clearly highlights the significant promise of long-read sequencing and its ability to unravel complexities within dark loci and repetitive genomic regions. Moreover, our identification of methylation changes in

specific retrotransposon loci offers the Alzheimer's disease research community novel avenues for exploration. Ultimately, advancements and widespread adoption of these sequencing techniques will continue to reveal medically-relevant genomic and epigenetic variation in neurodegenerative diseases.

DATA AVAILABILITY

All sequencing data will be made publicly available in the NCBI SRA database upon acceptance of the manuscript.

CODE AVAILABLITY

Example code used to generate data will be deposited in the Zenodo repository.

ACKNOWLEDGEMENTS

The Texas Advanced Computing Center (TACC) at The University of Texas at Austin provided HPC resources that contributed to the analysis and research results reported within this paper. URL: http://www.tacc.utexas.edu. DNA sequencing and library preparation were carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Human post-mortem frontal cortex tissues were acquired from the NIH Neurobiobank or Mayo Clinic Brain Bank.

FUNDING

Funding was provided by NINDS RF1 NS112391 (BF), the Rainwater Foundation (BF), NIGMS R25 (PR) and BrightFocus Foundation (WS).

ETHICS DECLARATIONS

BF serves on the Scientific Advisory Board of Transposon Therapeutics.

REFERENCES

1. Costantino, I., J. Nicodemus, and J. Chun, *Genomic mosaicism formed by somatic variation in the aging and diseased brain.* Genes, 2021. **12**(7): p. 1071.
2. Westra, J.W., et al., *Neuronal DNA content variation (DCV) with regional and individual differences in the human brain.* Journal of Comparative Neurology, 2010. **518**(19): p. 3981-4000.
3. Knouse, K.A., J. Wu, and A. Amon, *Assessment of megabase-scale somatic copy number variation using single-cell sequencing.* Genome Research, 2016. **26**(3): p. 376-384.
4. Chronister, W.D., et al., *Neurons with complex karyotypes are rare in aged human neocortex.* Cell reports, 2019. **26**(4): p. 825-835. e7.
5. Gorbunova, V., et al., *The role of retrotransposable elements in ageing and age-associated diseases.* Nature, 2021. **596**(7870): p. 43-53.
6. Huang, C.R.L., K.H. Burns, and J.D. Boeke, *Active transposition in genomes.* Annual Review of Genetics, 2012. **46**(1): p. 651-675.
7. Evrony, G.D., et al., *Cell lineage analysis in human brain using endogenous retroelements.* Neuron, 2015. **85**(1): p. 49-59.
8. Erwin, J.A., et al., *L1-associated genomic regions are deleted in somatic cells of the healthy human brain.* Nature Neuroscience, 2016. **19**(12): p. 1583-1591.
9. Upton, R., Kyle, et al., *Ubiquitous L1 mosaicism in hippocampal neurons.* Cell, 2015. **161**(2): p. 228-239.
10. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain.* Nature, 2011. **479**(7374): p. 534-537.
11. Ramirez, P., et al., *Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system.* Progress in Neurobiology, 2022. **208**: p. 102181.
12. Coufal, N.G., et al., *L1 retrotransposition in human neural progenitor cells.* Nature, 2009. **460**(7259): p. 1127-1131.
13. Evrony, G.D., et al., *Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain.* Cell, 2012. **151**(3): p. 483-496.
14. Gong, T., V.M. Hayes, and E.K.F. Chan, *Detection of somatic structural variants from short-read next-generation sequencing data.* Briefings in Bioinformatics, 2021. **22**(3).
15. Lanciano, S. and G. Cristofari, *Measuring and interpreting transposable element expression.* Nature Reviews Genetics, 2020. **21**(12): p. 721-736.
16. Huang, L., et al., *Single-cell whole-genome amplification and dequencing: methodology and applications.* Annual Review of Genomics and Human Genetics, 2015. **16**(1): p. 79-102.
17. Nam, C.H., et al., *Widespread somatic L1 retrotransposition in normal colorectal epithelium.* Nature, 2023. **617**(7961): p. 540-547.
18. Ebbert, M.T., et al., *Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight.* Genome biology, 2019. **20**: p. 1-23.
19. Ryan, N.M. and A. Corvin, *Investigating the dark-side of the genome: a barrier to human disease variant discovery?* Biological Research, 2023. **56**(1): p. 42.
20. Ahsan, M.U., et al., *A survey of algorithms for the detection of genomic structural variants from long-read sequencing data.* Nature Methods, 2023. **20**(8): p. 1143-1158.
21. Siudeja, K., et al., *Unraveling the features of somatic transposition in the Drosophila intestine.* The EMBO Journal, 2021. **40**(9).
22. Gershman, A., et al., *Epigenetic patterns in a complete human genome.* Science, 2022. **376**(6588): p. eabj5089.
23. Harris, R.A., et al., *Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications.* Nature Biotechnology, 2010. **28**(10): p. 1097-1105.
24. Moore, L.D., T. Le, and G. Fan, *DNA methylation and Its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
25. Pappalardo, X.G. and V. Barra, *Losing DNA methylation at repetitive elements and breaking bad.* Epigenetics & Chromatin, 2021. **14**(1).

26. Miller, C.A. and J.D. Sweatt, *Covalent modification of DNA regulates memory formation.* Neuron, 2007. **53**(6): p. 857-869.
27. Levenson, J.M., et al., *Evidence that DNA (Cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus.* Journal of Biological Chemistry, 2006. **281**(23): p. 15763-15773.
28. Jones, M.J., S.J. Goodman, and M.S. Kobor, *DNA methylation and healthy human aging.* Aging Cell, 2015. **14**(6): p. 924-932.
29. Ehrlich, M., *DNA methylation in cancer: too much, but also too little.* Oncogene, 2002. **21**(35): p. 5400-5413.
30. Mastroeni, D., et al., *Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation.* Neurobiology of Aging, 2010. **31**(12): p. 2025-2037.
31. Chouliaras, L., et al., *Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients.* Neurobiology of aging, 2013. **34**(9): p. 2091-2099.
32. Qazi, T.J., et al., *Epigenetics in Alzheimer's Disease: Perspective of DNA methylation.* Molecular Neurobiology, 2018. **55**(2): p. 1026-1044.
33. Zhang, L., et al., *Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease.* Nature Communications, 2020. **11**(1).
34. Bakulski, K.M., et al., *Genome-Wide DNA methylation differences between Late-Onset Alzheimer's Disease and cognitively normal controls in human frontal cortex.* Journal of Alzheimer's Disease, 2012. **29**(3): p. 571-588.
35. Watson, C.T., et al., *Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease.* Genome Medicine, 2016. **8**(1).
36. Braak, H., et al., *Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years.* Journal of Neuropathology & Experimental Neurology, 2011. **70**(11): p. 960-969.
37. Assoication, A.s. *2023 Alzheimer's disease facts and figures*. 2023.
38. Braak, H. and E. Braak, *Neuropathological stageing of Alzheimer-related changes.* Acta neuropathologica, 1991. **82**(4): p. 239-259.
39. Ewing, A.D., et al., *Nanopore sequencing enables comprehensive transposable element epigenomic profiling.* Molecular Cell, 2020. **80**(5): p. 915-928.e5.
40. Konkel, M.K., et al., *Sequence analysis and characterization of active human Alu subfamilies based on the 1000 Genomes Pilot Project.* Genome Biology and Evolution, 2015: p. evv167.
41. Subramanian, R.P., et al., *Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses.* Retrovirology, 2011. **8**(1): p. 90.
42. Thomas, J., H. Perron, and C. Feschotte, *Variation in proviral content among human genomes mediated by LTR recombination.* Mobile DNA, 2018. **9**(1).
43. Hughes, J.F. and J.M. Coffin, *Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution.* Nature genetics, 2001. **29**(4): p. 487-489.
44. Turner, G., et al., *Insertional polymorphisms of full-length endogenous retroviruses in humans.* Current Biology, 2001. **11**(19): p. 1531-1535.
45. Suarez, N.A., A. Macia, and A.R. Muotri, *LINE-1 retrotransposons in healthy and diseased human brain.* Developmental neurobiology, 2018. **78**(5): p. 434-455.
46. Muotri, A.R., et al., *Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition.* nature, 2005. **435**(7044): p. 903-910.
47. Stenz, L., *The L1-dependant and Pol III transcribed Alu retrotransposon, from its discovery to innate immunity.* Molecular biology reports, 2021. **48**(3): p. 2775-2789.
48. Sun, W., et al., *Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies.* Nature Neuroscience, 2018. **21**(8): p. 1038-1048.
49. Guo, C., et al., *Tau activates transposable elements in Alzheimer's disease.* Cell reports, 2018. **23**(10): p. 2874-2880.
50. Smolka, M., et al., *Comprehensive structural Vvriant detection: From mosaic to population-level*. 2022, Cold Spring Harbor Laboratory.
51. Goate, A., et al., *Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.* Nature, 1991. **349**(6311): p. 704-706.
52. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* Fly, 2012. **6**(2): p. 80-92.

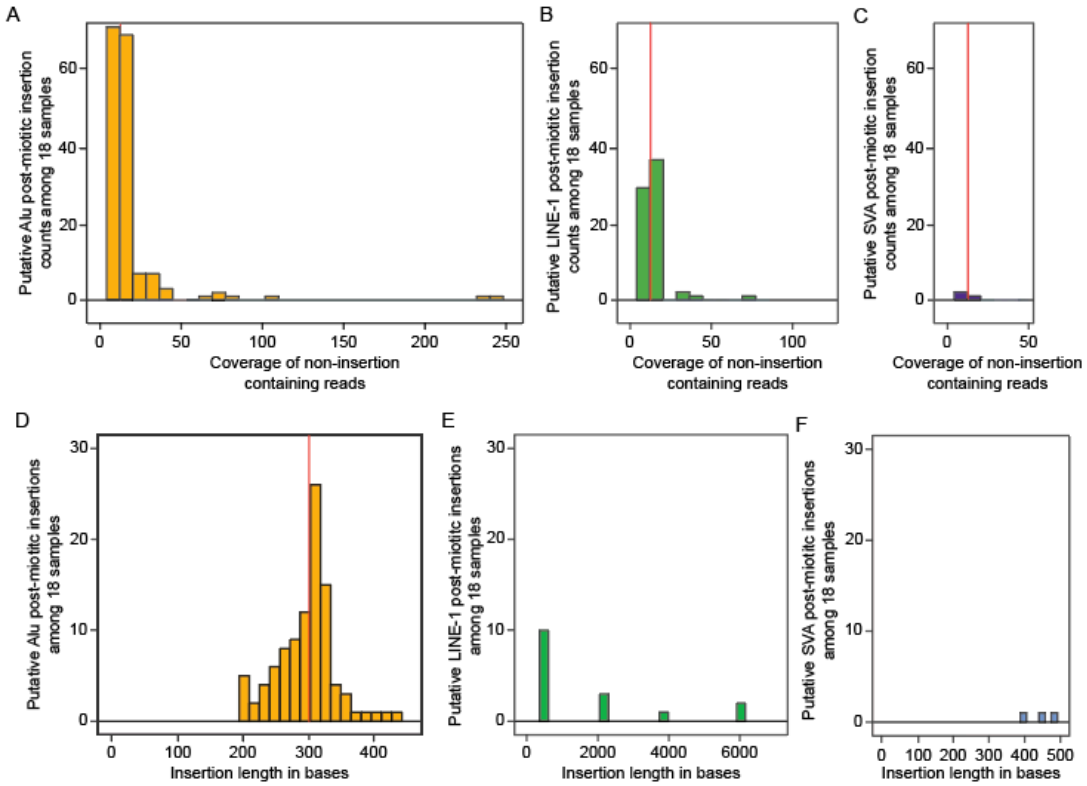
53. Chong, S.S., et al., *Gametic and somatic tissue–specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1.* Nature Genetics, 1995. **10**(3): p. 344-350.
54. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk.* Nature Genetics, 2019. **51**(3): p. 404-413.
55. Kawatani, K., et al., *ABCA7 deficiency causes neuronal dysregulation by altering mitochondrial lipid metabolism.* Molecular Psychiatry, 2023.
56. Fu, Y., et al., *Sex-specific lipid dysregulation in the Abca7 knockout mouse brain.* Brain Communications, 2022. **4**(3).
57. Kim, W.S., et al., *Deletion of ABCA7 increases cerebral Amyloid-β accumulation in the J20 mouse model of Alzheimer's Disease.* The Journal of Neuroscience, 2013. **33**(10): p. 4387-4394.
58. Sakae, N., et al., *ABCA7 Deficiency Accelerates Amyloid-β Generation and Alzheimer's Neuronal Pathology.* The Journal of Neuroscience, 2016. **36**(13): p. 3848-3859.
59. Satoh, K., et al., *ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing.* Journal of Biological Chemistry, 2015. **290**(40): p. 24152-24165.
60. Lövkvist, C., et al., *DNA methylation in human epigenomes depends on local topology of CpG sites.* Nucleic Acids Research, 2016. **44**(11): p. 5123-5132.
61. Chopra, P., et al., *Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice.* BMC Genomics, 2014. **15**(1): p. 131.
62. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nature Genetics, 2007. **39**(4): p. 457-466.
63. Luquez, T., et al., *Cell type-specific changes identified by single-cell transcriptomics in Alzheimer's disease.* Genome Medicine, 2022. **14**(1).
64. Johnson, T.S., et al., *Combinatorial analyses reveal cellular composition changes have different impacts on transcriptomic changes of cell type specific genes in Alzheimer's Disease.* Scientific Reports, 2021. **11**(1).
65. Sanchez-Luque, F.J., et al., *LINE-1 evasion of epigenetic repression in humans.* Molecular cell, 2019. **75**(3): p. 590-604. e12.
66. Grundy, E.E., N. Diab, and K.B. Chiappinelli, *Transposable element regulation and expression in cancer.* The FEBS journal, 2022. **289**(5): p. 1160-1179.
67. Xue, B., L.A. Sechi, and D.J. Kelvin, *Human endogenous retrovirus K (HML-2) in health and gisease.* Frontiers in Microbiology, 2020. **11**.
68. Li, H., *Minimap2: pairwise alignment for nucleotide sequences.* Bioinformatics, 2018. **34**(18): p. 3094-3100.
69. Pedersen, B.S. and A.R. Quinlan, *Mosdepth: quick coverage calculation for genomes and exomes.* Bioinformatics, 2018. **34**(5): p. 867-868.
70. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Research, 2016. **44**(D1): p. D733-D745.
71. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.
72. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update.* Nucleic Acids Research, 2022. **50**(W1): p. W345-W351.
73. Lawrence, M., et al., *Software for computing and annotating genomic ranges.* PLoS Computational Biology, 2013. **9**(8): p. e1003118.
74. Cavalcante, R.G. and M.A. Sartor, *annotatr: genomic regions in context.* Bioinformatics, 2017. **33**(15): p. 2381-2383.
75. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-842.
76. Robinson, J.T., et al., *Integrative genomics viewer.* Nature Biotechnology, 2011. **29**(1): p. 24-26.
77. Bao, W., K.K. Kojima, and O. Kohany, *Repbase Update, a database of repetitive elements in eukaryotic genomes.* Mobile DNA, 2015. **6**(1).
78. Loman, N.J., J. Quick, and J.T. Simpson, *A complete bacterial genome assembled de novo using only nanopore sequencing data.* Nature Methods, 2015. **12**(8): p. 733-735.
79. Akalin, A., et al., *methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles.* Genome Biology, 2012. **13**(10): p. R87.
80. Su, S., et al., *NanoMethViz: An R/Bioconductor package for visualizing long-read methylation data.* PLOS Computational Biology, 2021. **17**(10): p. e1009524.

81. Feng, H. and H. Wu, *Differential methylation analysis for bisulfite sequencing using DSS.* Quantitative Biology, 2019. **7**(4): p. 327-334.
82. Gigante, S., et al., *Using long-read sequencing to detect imprinted DNA methylation.* Nucleic Acids Research, 2019. **47**(8): p. e46-e46.
83. Fishilevich, S., et al., *GeneHancer: genome-wide integration of enhancers and target genes in GeneCards.* Database, 2017. **2017**.
84. Kent, W.J., et al., *The Human Genome Browser at UCSC.* Genome Research, 2002. **12**(6): p. 996-1006.
85. Gel, B., et al., *regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests.* Bioinformatics, 2016. **32**(2): p. 289-291.
86. Collins, R.L., et al., *A structural variation reference for medical and population genetics.* Nature, 2020. **581**(7809): p. 444-451.
87. Niu, Y., et al., *Characterizing mobile element insertions in 5675 genomes.* Nucleic Acids Research, 2022. **50**(5): p. 2493-2508.
88. Kanehisa, M., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
89. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The Innovation, 2021. **2**(3): p. 100141.
90. Wickham, H., *ggplot2.* Use R!, 2016.
91. Chen, S., et al., *A genome-wide mutational constraint map quantified from variation in 76,156 human genomes*. 2022, Cold Spring Harbor Laboratory.
92. Borges-Monroy, R., et al., *Whole-genome analysis reveals the contribution of non-coding de novo transposon insertions to autism spectrum disorder.* Mobile DNA, 2021. **12**: p. 1-15.
93. Gardner, E.J., et al., *The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology.* Genome Research, 2017. **27**(11): p. 1916-1929.
94. Chu, C., et al., *Comprehensive identification of transposable element insertions using multiple sequencing technologies.* Nature Communications, 2021. **12**(1).
95. Dawson, T., et al., *Locus specific endogenous retroviral expression associated with Alzheimer's disease.* Frontiers in Aging Neuroscience, 2023. **15**.
96. Macciardi, F., et al., *A retrotransposon storm marks clinical phenoconversion to late-onset Alzheimer's disease.* GeroScience, 2022. **44**(3): p. 1525-1550.
97. Scopa, C., et al., *JUN upregulation drives aberrant transposable element mobilization, associated innate immune response, and impaired neurogenesis in Alzheimer's disease.* Nature Communications, 2023. **14**(1).
98. Wahl, D., et al., *The reverse transcriptase inhibitor 3TC protects against age‐related cognitive dysfunction.* Aging Cell, 2023. **22**(5).
99. Ochoa, E., et al., *Pathogenic tau–induced transposable element–derived dsRNA drives neuroinflammation.* Science Advances, 2023. **9**(1): p. eabq5423.
100. Jha, A.R., et al., *Human Endogenous Retrovirus K106 (HERV-K106) was infectious after the emergence of anatomically modern humans.* PLoS ONE, 2011. **6**(5): p. e20234.
101. Marchi, E., et al., *Unfixed endogenous retroviral insertions in the human population.* Journal of Virology, 2014. **88**(17): p. 9529-9537.
102. Wildschutte, J.H., et al., *Discovery of unfixed endogenous retrovirus insertions in diverse human populations.* Proceedings of the National Academy of Sciences, 2016. **113**(16): p. E2326-E2334.
103. Wallace, T.A., et al., *Elevated HERV-K mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older men and smokers.* Carcinogenesis, 2014. **35**(9): p. 2074-2083.
104. Xue, B., et al., *Identification of the distribution of human endogenous retroviruses K (HML-2) by PCR-based target enrichment sequencing.* Retrovirology, 2020. **17**(1).
105. Moyes, D., D.J. Griffiths, and P.J. Venables, *Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease.* Trends in genetics, 2007. **23**(7): p. 326-333.
106. Wang, H., et al., *The role of structural variations in Alzheimer's disease and other neurodegenerative diseases.* Frontiers in Aging Neuroscience, 2023. **14**.
107. Dib, S., J. Pahnke, and F. Gosselet, *Role of ABCA7 in Human Health and in Alzheimer's Disease.* International Journal of Molecular Sciences, 2021. **22**(9): p. 4603.
108. De Roeck, A., et al., *An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease.* Acta neuropathologica, 2018. **135**: p. 827-837.

109. Reitz, C., et al., *Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4, and the risk of Late-Onset Alzheimer Disease in African Americans.* JAMA, 2013. **309**(14): p. 1483.
110. Aikawa, T., et al., *ABCA7 haplodeficiency disturbs microglial immune responses in the mouse brain.* Proceedings of the National Academy of Sciences, 2019. **116**(47): p. 23790-23796.
111. Mastrorosa, F.K., D.E. Miller, and E.E. Eichler, *Applications of long-read sequencing to Mendelian genetics.* Genome Medicine, 2023. **15**(1).
112. Li, Q.S., Y. Sun, and T. Wang, *Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions.* Clinical Epigenetics, 2020. **12**(1).
113. Pellegrini, C., et al., *A Meta-Analysis of brain DNA methylation across sex, age, and Alzheimer's Disease points for accelerated epigenetic aging in neurodegeneration.* Frontiers in Aging Neuroscience, 2021. **13**.
114. Smith, R.G., et al., *A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex.* Nature Communications, 2021. **12**(1).
115. Black, E.M. and S. Giunta, *Repetitive fragile dites: centromere satellite DNA as a source of genome instability in human diseases.* Genes, 2018. **9**(12): p. 615.
116. Enukashvily, N.I., et al., *Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells.* Cytogenetic and Genome Research, 2007. **118**(1): p. 42-54.
117. Barra, V. and D. Fachinetti, *The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA.* Nature Communications, 2018. **9**(1).
118. Peze-Heidsieck, E., et al., *Retrotransposons as a Source of DNA Damage in Neurodegeneration.* Frontiers in Aging Neuroscience, 2022. **13**.
119. Rollins, R.A., et al., *Large-scale structure of genomic methylation patterns.* Genome research, 2006. **16**(2): p. 157-163.
120. Jacobs, F.M., et al., *An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons.* Nature, 2014. **516**(7530): p. 242-245.
121. Imbeault, M., P.-Y. Helleboid, and D. Trono, *KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks.* Nature, 2017. **543**(7646): p. 550-554.
122. Srinivasachar Badarinarayan, S. and D. Sauter, *Not all viruses cause disease: HERV-K(HML-2) in healthy human tissues.* PLOS Biology, 2022. **20**(10): p. e3001884.
123. Andrade-Moraes, C.H., et al., *Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles.* Brain, 2013. **136**(12): p. 3738-3752.
124. Mathys, H., et al., *Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology.* Cell, 2023. **186**(20): p. 4365-4385.e27.
125. Hernández, H.G., et al., *Global Long Interspersed Nuclear Element 1 DNA Methylation in a Colombian Sample of Patients With Late-Onset Alzheimer's Disease.* American Journal of Alzheimer's Disease & Other Dementias®, 2014. **29**(1): p. 50-53.
126. Bollati, V., et al., *DNA methylation in repetitive elements and Alzheimer disease.* Brain, Behavior, and Immunity, 2011. **25**(6): p. 1078-1083.
127. Göke, J. and H.H. Ng, *CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome.* EMBO reports, 2016. **17**(8): p. 1131-1144.
128. Thompson, J., Peter, S. Macfarlan, Todd, and C. Lorincz, Matthew, *Long Terminal Repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire.* Molecular Cell, 2016. **62**(5): p. 766-776.
129. Cohen, C.J., W.M. Lock, and D.L. Mager, *Endogenous retroviral LTRs as promoters for human genes: a critical assessment.* Gene, 2009. **448**(2): p. 105-114.
130. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution.* Nature Reviews Genetics, 2009. **10**(10): p. 691-703.
131. María, G.-C., et al., *Human endogenous retroviruses and cancer.* Cancer Biology & Medicine, 2016. **13**(4): p. 483.
132. Gao, Y., X.-F. Yu, and T. Chen, *Human endogenous retroviruses in cancer: expression, regulation and function.* Oncology Letters, 2020. **21**(2).
133. Agoni, L., J. Lenz, and C. Guha, *Variant splicing and influence of ionizing radiation on Human Endogenous Retrovirus K (HERV-K) transcripts in cancer cell Lines.* PLoS ONE, 2013. **8**(10): p. e76472.

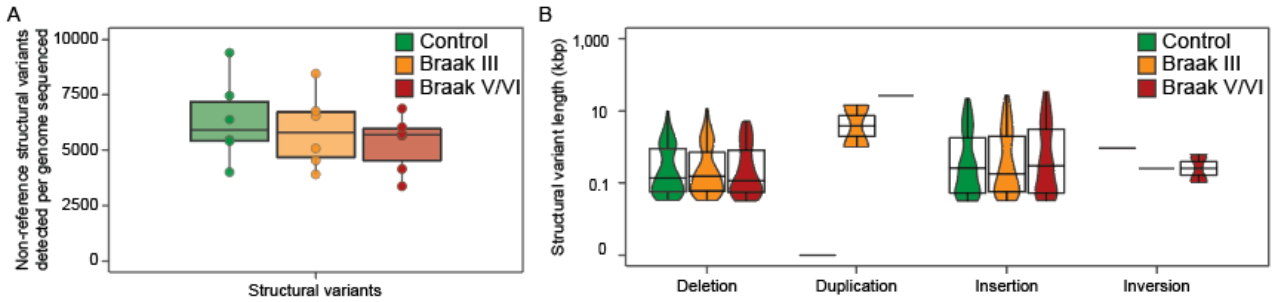
134. Nelson, P.N., et al., *Rheumatoid arthritis is associated with IgG antibodies to human endogenous retrovirus gag matrix: a potential pathogenic mechanism of disease?* The journal of rheumatology, 2014. **41**(10): p. 1952-1960.
135. Bhardwaj, N., et al., *Differential expression of HERV-K (HML-2) proviruses in cells and virions of the Teratocarcinoma cell line Tera-1.* Viruses, 2015. **7**(3): p. 939-968.
136. Li, W., et al., *Human endogenous retrovirus-K contributes to motor neuron disease.* Science Translational Medicine, 2015. **7**(307): p. 53-307.
137. Terry, S.N., et al., *Expression of HERV-K108 envelope interferes with HIV-1 production.* Virology, 2017. **509**: p. 52-59.
138. De Coster, W., M.H. Weissensteiner, and F.J. Sedlazeck, *Towards population-scale long-read sequencing.* Nature Reviews Genetics, 2021. **22**(9): p. 572-587.
139. Miga, K.H., et al., *Centromere reference models for human chromosomes X and Y satellite arrays.* Genome research, 2014. **24**(4): p. 697-707.
140. Altemose, N., et al., *Complete genomic and epigenetic maps of human centromeres.* Science, 2022. **376**(6588): p. eabl4178.
141. *Implications of the first complete human genome assembly.* Genome Research, 2022. **32**(4): p. 595-598.
142. Hård, J., et al., *Long-read whole-genome analysis of human single cells.* Nature Communications, 2023. **14**(1): p. 5164.
143. Kolmogorov, M., et al., *Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation.* Nature Methods, 2023. **20**(10): p. 1483-1492.

SUPPLEMENTAL FIGURES



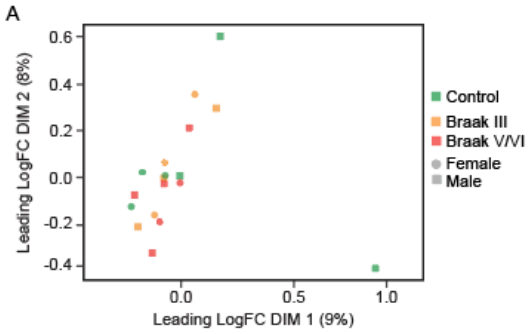
Supplemental Figure 1 | Non-reference genome retrotransposon insertions

A, B, C) Bar graph of single read supported Alu, L1 and SVA insertions and their respective number of non-insertions at the genomic insertion site. Red vertical line represents the cutoff of 10 non insertion containing reads for an insertion to be considered putative somatic. D, E, F) Bar graph of the length of insertions for Alu, L1, and SVA.



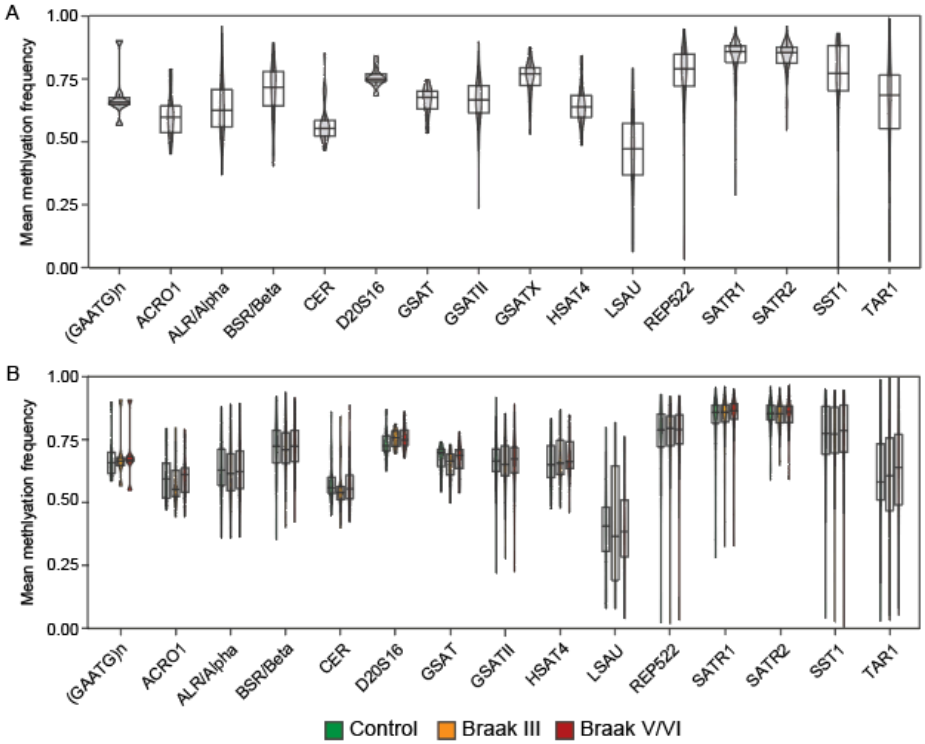
Supplemental Figure 2 | Non-reference structural variants

A) Boxplots of all non-reference structural variants insertion counts normalized to sequencing depth. B) Violin plot of structural variant length (within the range 500 bp – 1 mbp) for structural variant length.

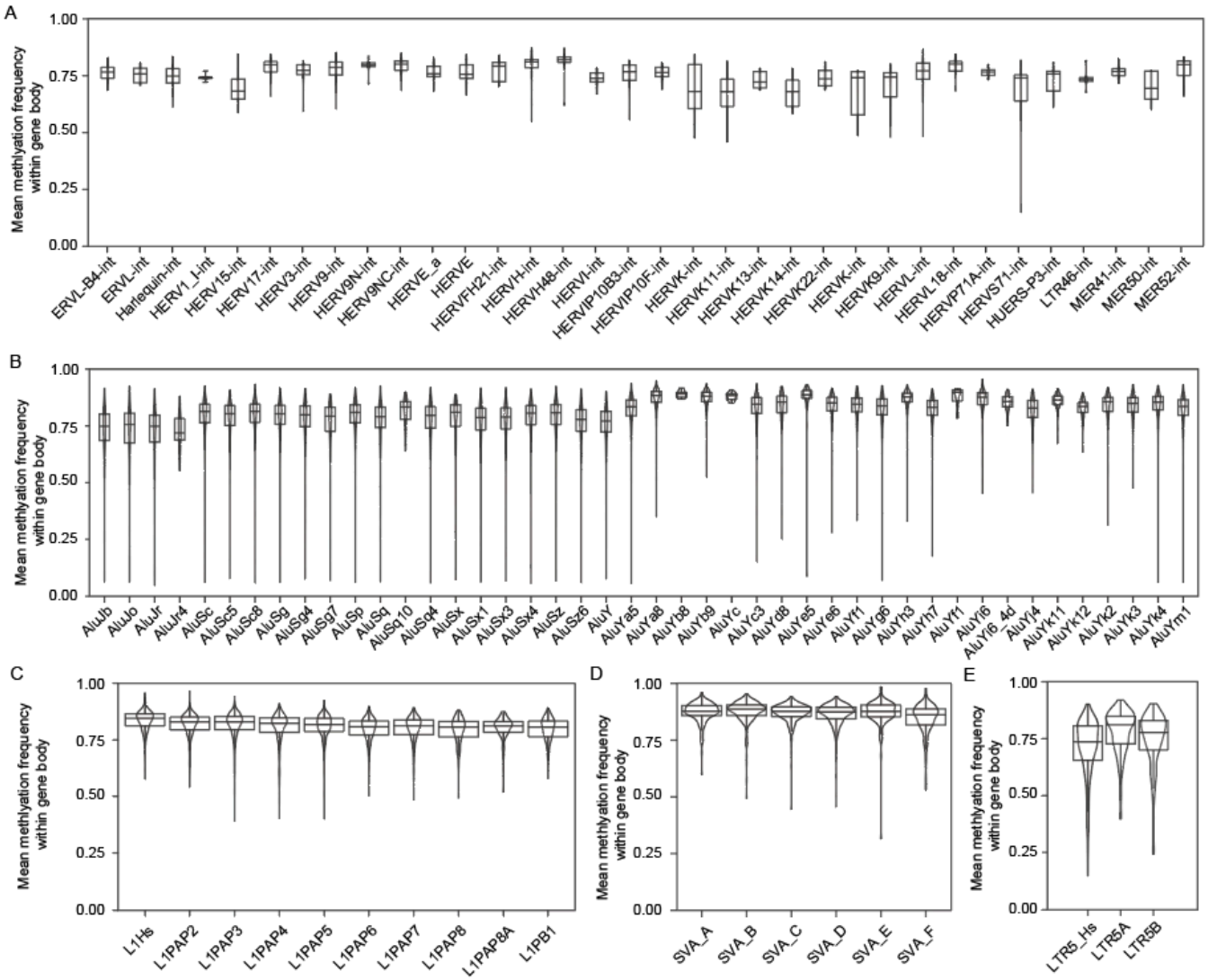


Supplemental Figure 3 | Nanopore methylation

A) MDS plot of LMR of autosomal CpG sites.

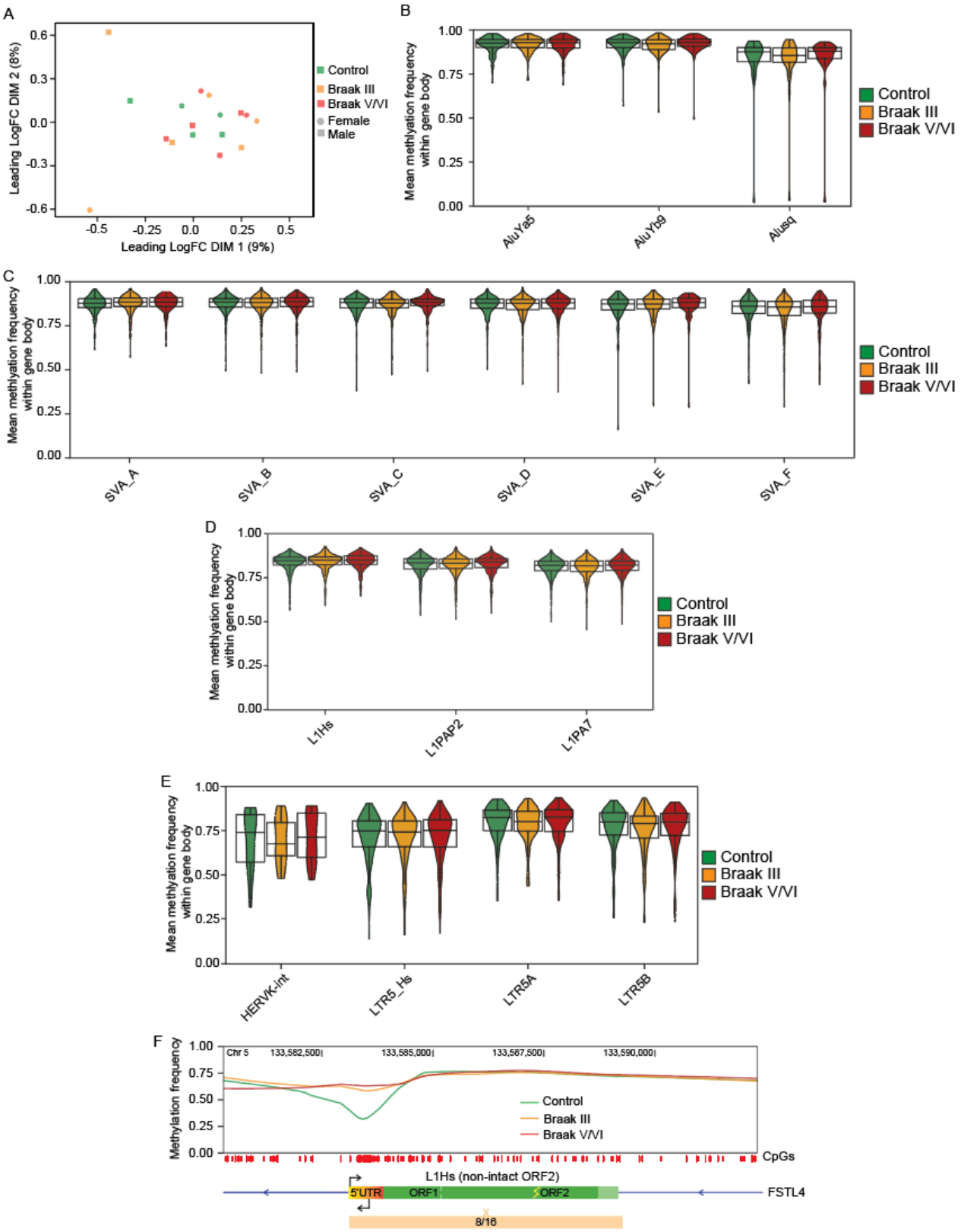


Supplemental Figure 4 | Nanopore methylation of satellite regions across all brain and Alzheimer's disease
Violin and box plots of the mean methylation frequency within satellite regions greater than 100 bp in length and with at least 10 CpG calls for A) all samples combined and B) by Braak Stage.



Supplemental Figure 5 | Nanopore methylation of retrotransposon

A) Violin and box plots of the mean methylation frequency within retrotransposon with at least 10 CpG site calls for all samples combined and a minimum length of 5 kbp in ERV retrotransposon family, 290 bp in Alu family, 5.9 kbp in L1 family, 1000 bp in SVA family, 900 bp in LTR5 subfamily.



Supplemental Figure 6| Nanopore methylation of retrotransposon

A) MDS plot of LMR of autosomal CpG sites within repetitive elements. B, C, D, E) Violin and box plots of the mean methylation frequency within gene body of Alu, SVA, L1, and HERV-K/ LTR sequences by Braak Stage for loci with at least 10 CpG site calls with a minimum length of 5.9 kbp in L1, 900 bp in LTR5Hs and 290 bp in AluYa5/Yb8. F) Methylation profile of L1-Ta body (within the gene FSTL4) for Braak III, Braak V/VI, and control.

SUPPLEMENTARY TABLES

Sample type	Sample name	Braak stage	Thal stage	Age	Sex	PMI	Race
Control	CTRL-1	0	Unknown	84	Male	Unknown	Unknown
Control	CTRL-2	0	Unknown	83	Male	18	Caucasian
Control	CTRL-3	0	Unknown	68	Female	Unknown	Unknown
Control	CTRL-4	0	Unknown	79	Male	10	Caucasian
Control	CTRL-5	0	Unknown	68	Female	Unknown	Unknown
Control	CTRL-6	0	Unknown	74	Female	3	Caucasian
Braak III	Braak_III_1	2.5	0	88	Male	Unknown	Caucasian
Braak III	Braak_III_2	3	0	92	Female	Unknown	Caucasian
Braak III	Braak_III_3	3	0	79	Female	8	Caucasian
Braak III	Braak_III_4	3	2	75	Male	12	Caucasian
Braak III	Braak_III_5	3.5	0	85	Female	24	Caucasian
Braak III	Braak_III_6	3	2	75	Male	Unknown	Caucasian
Braak V/VI	Braak_V-VI_1	6	Unknown	71	Male	Unknown	Unknown
Braak V/VI	Braak_V-VI_2	6	Unknown	67	Female	12	Caucasian
Braak V/VI	Braak_V-VI_3	5	Unknown	79	Male	13	Caucasian
Braak V/VI	Braak_V-VI_4	5	Unknown	70	Male	Unknown	Unknown
Braak V/VI	Braak_V-VI_5	6	Unknown	73	Female	5	Caucasian
Braak V/VI	Braak_V-VI_6	6	Unknown	68	Male	5	Caucasian

Supplemental Table 1| Demographic of brain samples

PMI = Post mortem interval in months.

Sample type	Sample name	Mean read quality	Total reads	Read length n50	Total gbp	Genome coverage
Control	CTRL-1	11.6	1,023,179	30,505	14.19	4.09
Control	CTRL-2	11.9	1,627,426	30,006	18.24	5.30
Control	CTRL-3	11.5	6,478,290	11,670	35.20	10.00
Control	CTRL-4	11.8	2,047,730	28,548	21.61	6.32
Control	CTRL-5	11.8	3,840,393	21,898	25.73	7.43
Control	CTRL-6	11.8	3,638,404	25,346	25.78	7.51
Braak III	Braak_III_1	11.8	1,455,635	27,275	15.69	4.57
Braak III	Braak_III_2	11.8	4,185,578	16,613	27.67	8.02
Braak III	Braak_III_3	11.7	2,443,785	28,230	35.66	10.50
Braak III	Braak_III_4	11.5	2,733,026	23,740	21.07	6.11
Braak III	Braak_III_5	11.7	1,937,159	25,730	20.02	5.88
Braak III	Braak_III_6	11.6	2,014,256	25,769	30.55	8.93
Braak V/VI	Braak_V-VI_1	11.7	1,504,113	28,758	19.63	5.75
Braak V/VI	Braak_V-VI_2	11.9	8,720,004	9,668	34.08	9.74
Braak V/VI	Braak_V-VI_3	11.7	6,542,225	17,291	41.64	12.12
Braak V/VI	Braak_V-VI_4	11.8	3,226,405	16,556	24.18	7.01
Braak V/VI	Braak_V-VI_5	11.9	3,583,034	19,083	24.58	7.05
Braak V/VI	Braak_V-VI_6	11.8	1,903,934	31,580	22.55	6.59

Supplemental Table 2| Nanopore sequencing metrics

Table of Promethion nanopore sequencing metrics per sample. Total reads= total amount of reads sequenced. Read length n50 = read length n50 in kbp. Total gbp = total amount of DNA sequenced for sample. Genome coverage = sequencing coverage of sample.

Name	Coordinates	Stage	CpGs	Control	Braak	Δ 5mc	GeneHancer genes
AluYa5	chr11:102601093-102601457	Braak III	13	0.81	0.6	-0.21	MMP7, MMP20
AluYa5	chr2:20598293-20598970	Braak III	17	0.53	0.7	0.17	Hs13BP, RHOB, GDF7
LTR5_Hs	chr11:63530136-63530437	Braak III	15	0.52	0.32	-0.2	GH11J063530, RARRESE3
LTR5_Hs	chr8:144795817-144796432	Braak III	10	0.35	0.18	-0.17	ZNF34, RPL8, ZNF517
LTR5_Hs	chr11:63529265-63530678	Braak V/VI	15	0.62	0.45	-0.18	GH11J063530, RARRESE3

Supplemental Table 3 | Retrotransposon loci methylation metrics and predicted gene associations

Table of retrotransposon loci with the largest changes of methylation in Braak III or Braak V/VI. Coordinates = coordinates in hg38 reference genome. Stage = Braak stage at which methylation change occurs. CpGs = number of CpG sites present in the differentially methylated region. Control = mean methylation frequency at differentially methylated region for control. Braak = mean methylation frequency at differentially methylated region for respective Braak group. Δ 5mc =change in methylation. GeneHancer genes = list of genes which of which associated retrotransposon are predicted to interact with through an enhancer-based function via the GeneHancer track in the UCSC genome browser.

SUPPLEMENTAL FILES

Supplemental File 1| Non-reference genome retrotransposon insertion calls
Sheet 1 - Main output file of TLDR annotated with supplemental information, Sheet 2 – Genic locations of TLDR insertions.

Supplemental File 2| Non-reference structural variants tables
Sheet 1 - Main output file of SNIFFLES2 combined VCF file converted into table format, Sheet 2 - Blast results to identify Sheet 3 - pseudogenized gene insertions, insertions and deletions identified in dark regions.

Supplemental File 3| Differential methylation of promoter regions and dark loci
Output of DSS annotated with promoter regions and dark loci regions. Sheet 1 – Differential methylation of promoter regions for Braak III vs Control. Sheet 2 - Differential methylation of promoter regions for Braak V/VI vs Control. Sheet 3 – Differential methylation of dark regions for Braak III vs Control. Sheet 4 - Differential methylation of dark regions for Braak V/VI vs Control.

Supplemental File 4| Differentially methylated repetitive regions of Braak III vs control and Braak V/VI vs control
Output of DSS annotated with Repeat Masker defined repetitive regions. Sheet 1 – Differential methylation of repetitive regions for Braak V/VI vs Control

Supplemental File 5| Non-reference genome retrotransposon insertion calls
Raw output of TLDR analysis.

Supplemental File 6| Non-reference genome structural variants insertion calls
Raw VCF output of SNIFFLES2 analysis.

Supplemental File 7| Raw output of differentially methylated loci of Braak III vs control
Raw Output of DSS for the Braak III vs control comparison.

Supplemental File 8| Raw output of differentially methylated loci of Braak V/VI vs control
Raw Output of DSS for the Braak V/VI vs control comparison.