

RESEARCH ARTICLE

Long-read sequencing reveals genomic and epigenomic variation in the dark genome of human Alzheimer's disease

Paulino Ramirez^{1,2,3,4} | Wenyan Sun⁵ | Shiva Kazempour Dehkordi^{2,3} | Habil Zare^{2,3} | Giovanni Pascarella⁶ | Piero Carninci⁶ | Bernard Fongang^{2,7} | Kevin F. Bieniek^{2,8} | Bess Frost⁴¹Barshop Institute for Longevity and Aging Studies, San Antonio, Texas, USA²Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, Texas, USA³Department of Cell Systems and Anatomy, University of Texas Health San Antonio, San Antonio, Texas, USA⁴Center for Alzheimer's Disease Research, Brown University, Providence, Rhode Island, USA⁵Clinical Neuroscience Research Center, Department of Neurosurgery, School of Medicine, Tulane University, New Orleans, Louisiana, USA⁶Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan⁷Department of Biochemistry & Structural Biology, University of Texas Health San Antonio, San Antonio, Texas, USA⁸Department of Pathology, University of Texas Health San Antonio, San Antonio, Texas, USA

Correspondence

Bess Frost, Salame-Feraud Director, Center for Alzheimer's Disease Research, Brown University, 185 Meeting St., Sidney Frank Hall, Rm 372, Providence, RI 02912, USA.
Email: bess_frost@brown.edu

Funding information

BrightFocus Foundation; National Institute of General Medical Sciences, Grant/Award Number: R25 GM095480; National Institute of Neurological Disorders and Stroke, Grant/Award Number: RF1 NS112391; National Institute on Aging, Grant/Award Number: R01AG078964

Abstract

INTRODUCTION: Faulty DNA repair and epigenetic regulation contribute to neurodegeneration in Alzheimer's disease. Long-read sequencing enables analysis of "dark regions" that are difficult to study via traditional sequencing methodologies.**METHODS:** Using nanopore whole-genome DNA sequencing of *post mortem* brain from early- and late- stage Alzheimer's disease cases and controls, we analyzed retrotransposition, non-allelic homologous recombination (NAHR), structural variation, and DNA methylation within repetitive regions.**RESULTS:** Retrotransposon insertions and NAHR were enriched in centromeric/pericentromeric and ribosomal DNA (rDNA) in the aged brain. Putatively somatic AluY retrotransposon insertions trended upward in late-stage disease. Enrichment of NAHR between repetitive regions and DNA demethylation were detected in early disease. Differential methylation of dark regions within specific Alzheimer's disease risk genes and repetitive elements occurred across disease stage.**DISCUSSION:** This study provides the first long-read analysis of repetitive elements in the aged human brain and identifies these regions as hotspots for genomic variation in Alzheimer's disease.This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.© 2025 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

KEYWORDS

Alzheimer's disease, epigenomics, genomics, repetitive elements

Highlights

- Long-read sequencing enables analysis of “dark” regions Alzheimer's disease brains
- Somatic AluY retrotransposon insertions may be elevated in late-stage disease
- Retrotransposon-associated non-allelic homologous recombination (NAHR) is enriched in repetitive regions in early disease
- DNA demethylation within the centro/pericentromere and ribosomal DNA (rDNA) occur across disease stage
- Dark regions of risk genes are differentially methylated across disease stage

1 | INTRODUCTION

While landmark studies have revealed genetic associations contributing to Alzheimer's disease risk,^{1–3} the inability of traditional DNA sequencing methods to analyze genomic “dark regions” limits our understanding of the full genetic architecture of disease. Dark regions are highly repetitive elements that are not resolved via traditional 50–150 bp sequencing,^{4–8} many of which are lacking in the GRCh38 human reference genome. These regions include low-complexity microsatellites, transposable element-rich sequences, centromeric DNA, and rDNA arrays. Combining the new human CHM13 telomere-to-telomere reference genome with long-read DNA sequencing overcomes this challenge by generating long reads that enable confident mapping to repetitive loci.⁹ Nanopore sequencing of native, un-amplified DNA further facilitates analysis of DNA modifications such as 5-methylcytosine (5mC) within repetitive sequences.

The brain accumulates structural variants that result in genomic mosaicism between cells as organisms age.¹⁰ Structural variants can occur in progenitor cells, proliferating glial and epithelial cells, and in post-mitotic neurons, and can arise from errors in DNA repair, homologous recombination, and retrotransposition. Retrotransposition is a process in which retrotransposon-encoded RNA is reverse transcribed into complementary DNA integrates back into the genome, generating a novel insertion.¹¹ Studies in *Drosophila*, cultured cells, mouse brain, and human Alzheimer's disease report that retrotransposon RNA is significantly elevated in settings of tauopathy.^{12–14} While evidence in *Drosophila* and HeLa cells suggests that pathogenic forms of tau cause retrotransposition,^{13,15} it is currently unknown if retrotransposons mobilize to a greater degree in the human brain affected by Alzheimer's disease.

While retrotransposons make up 35% of the human genome, most retrotransposons have lost the ability to mobilize due to mutations that have accumulated over the course of evolution. Specific subfamilies of long interspersed nuclear element (LINE) retrotransposons are capable of autonomous retrotransposition in humans; specific families of

short interspersed nuclear element (SINE) and SINE-VNTR-Alu (SVA) retrotransposons are active and leverage LINE-encoded protein for trans-retrotransposition.^{16–19} Retrotransposition of the active human-specific LINE-1 element (L1Hs) occurs in neurons and glia and is a proposed physiological feature of developing neurons.^{20–22} In addition to *de novo* retrotransposition events, the repetitive nature of retrotransposon DNA has the potential to facilitate non-allelic homologous recombination (NAHR), leading to new genomic variants.^{23–27} NAHR, documented in Alzheimer's disease,²⁸ can arise from double stranded DNA breaks that occur during instances of high cellular stress and contribute to the loss of genomic integrity.²⁸

Here we utilized Oxford Nanopore Technologies native DNA sequencing to identify putatively somatic retrotransposition events, NAHR, structural variants, and DNA methylation changes in frontal cortex of individuals at early and late stages of Alzheimer's disease and age-matched controls, with a particular focus on dark regions of the human genome. The pathological distribution of tau in Alzheimer's disease is defined by Braak neurofibrillary tau tangle staging. Compared to a brain lacking tau tangles (Braak 0), tau pathology progresses from brainstem to transentorhinal/entorhinal regions (Braak I), to the primary hippocampus (Braak III) and then association and primary neocortical regions of the brain (Braak V–VI).²⁹ Among all brains analyzed, we find that repetitive segments of the genome are particularly susceptible to genomic changes compared to coding regions. Putatively somatic retrotransposon insertions primarily occur within centromeric and pericentromeric regions of the aged brain. While NAHR is also elevated in centromeric and pericentromeric regions, we find that 47S rDNA exhibits a particularly high frequency of NAHR compared to other genomic regions. In cases with Alzheimer's disease, we detect a trending increase in rare, putatively somatic retrotransposition events involving the SINE AluY family in advanced disease, as well as disease stage-specific differences in NAHR and DNA methylation of repetitive elements and retrotransposons. Overall, our study represents the first long-read DNA sequencing-based analysis of retrotransposons, NAHR, structural variants, and DNA methylation in genomic dark regions in brains affected by Alzheimer's disease and identifies retrotransposons,

RESEARCH IN CONTEXT

1. **Systematic review:** Evidence suggests that defects in DNA repair and epigenetic control mediate neurodegeneration in Alzheimer's disease. DNA sequencing studies have been applied to study these processes but are unable to characterize dark regions of the genome. Advances in sequencing technologies now enable more comprehensive genome analyses.
2. **Interpretation:** Analysis of retrotransposon insertions and non-allelic homologous recombination (NAHR) revealed enrichment in centromeric/pericentromeric and ribosomal DNA (rDNA) in the aged brain. Putatively somatic AluY retrotransposon insertions trended upward in late-stage disease. DNA demethylation in centro/pericentromeres and rDNA, along with differential methylation of dark regions within specific Alzheimer's disease risk genes and repetitive elements, was observed across disease stages.
3. **Future directions:** Our study provides a basis for future work focusing on genomic and epigenomic changes in the dark regions, the centro/pericentromere and rDNA in the context of Alzheimer's disease pathology.

centromeric/pericentromeric regions and rDNA as key hotspots of genomic variation.

2 | METHODS

2.1 | Sample collection

18 *post mortem* 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.

2.2 | DNA processing and read alignment

Approximately 80 mg of tissue was homogenized with a Dounce and 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, randomized into groups of three per PromethION flow cell, and sequenced to obtain a desired target coverage of 10X

per sample. Reads were basecalled with Guppy version 4.0.11 (ONT). Nanopore-derived reads were aligned to the GRCh38 genome (University of California at Santa Cruz [UCSC]) and the HS1-CHM13 genome using MiniMap2³⁰ using nanopore sequencing-specific parameters to obtain bam files. Mosdepth³¹ was used to determine the average coverage per sample. An additional 30X coverage NeuN-sorted nanopore dataset for Alzheimer's disease ($n = 10$, Braak 4.6–6) and controls ($n = 10$, Braak 1–2.5)²⁸ was downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) repository (PRJNA636606) and was subject to the same post-processing steps.

2.3 | 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 insertion file with calls across all 18 samples. Insertions called by TLDR were then filtered to retain insertions with a target site duplication (TSD), an UnmapCover (fraction of inserted sequence covered by retrotransposon sequence) value of 0.8 or more, an intact 3' end of the retrotransposon of interest and a PASS quality score to eliminate low confidence/quality insertions. We defined a putative L1 endonuclease (L1en)-mediated singleton in a post-mitotic cell as a call that was only supported by a single insertion-spanning read of 10 kbp or longer in a single individual who had at least 10 additional reads across the region, all of which lack the insertion, along with a TSD greater or equal to 5 bp and less than or equal to 31 bp and an L1en motif at the insertion site. All singleton insertion alignments were further visualized with the Integrative Genomics Viewer (IGV) to confirm L1en insertion characteristics.³² The NeuN-sorted dataset was analyzed similarly.

2.4 | NAHR analyses

Nanopore-derived reads were aligned to both the GRCh38 and CHM13 genomes using the LAST alignment software with parameters suggested by the TE-reX pipeline.^{28,33} NAHR events were then called using TE-reX²⁸ and recommended analysis pipeline to obtain singleton and polymorphic events. Events were only retained if they had a value equal to or less than $1E-4$ and were identified as a singleton event. NAHR enrichment values were calculated by dividing the genome into 100 kbp windows and obtaining the ratio of NAHR events to the total repetitive elements per window.²⁸

2.5 | Structural variant calling

Individual bam files were subjected to structural variant calling using SNIFFLES2³⁴ (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 in the GRCh38 were determined using snpEFF. Insertion nucleotide sequences were extracted from the variant file and searched against the NCBI nucleotide and RefSeq³⁵ databases using BLASTn³⁶ (default parameters) within the Galaxy platform.³⁷ 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 genotyped across all 18 samples. A variant that was only found in samples belonging to a single Braak group was considered unique to that group.

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

TLDR insertions, TE-reX NAHR events, and SNIFFLES2 variants unique to a Braak group were compared via 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).

2.7 | Characterization of retrotransposon insertions and structural variants

Coordinates from the TLDR, TE-reX, and SNIFFLES2 output were processed with the R GenomicRanges package.³⁸ Overlap of insertions and variants with regions (genic/non-genic, dark regions, tandem repeat regions) were determined with AnnotatR.³⁹ Bedtools⁴⁰ was used to identify variants near single nucleotide polymorphisms. Gene graphs were first generated by IGV³² and then modified in Adobe Illustrator 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 (L1Hs-Ta) diagnostic ACA and G nucleotides in the 3' untranslated region (UTR). HERV-K open reading frames were determined by aligning HERV-Ks of interest to the consensus HERV-K sequence (DFAM ID: DF0000188). HERV-K non-reference insertions were manually analyzed for coding potential and completeness compared to the full-length HERV-K consensus sequence. To identify other HERV insertions, SNIFFLES2-identified insertions were searched against a custom HERV internal and LTR5 database with sequences from RepBase⁴¹ using BLASTn³⁶ (default parameters) within the Galaxy platform.

2.8 | CpG 5mC methylation analysis

The 5mC modifications were called using the Nanopolish⁴² call-methylation pipeline with single bam files described above on both the GRCh38 and CHM13 genomes. Alignments with a mapping quality score of 20 or more were included in calculating log-likelihood methylation ratios. Nanopolish-generated log-likelihood ratio values were

then converted to methylation frequencies with accompanying Nanopolish scripts. The default log-likelihood ratio threshold suggested by the authors was used to filter out low-quality calls. Samples were then analyzed via multidimensional scaling (MDS) to visualize sample clustering. Sample CTRL-1 had a previous cancer diagnosis and was identified as an outlier; this sample was removed from further analysis. 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 methylArtist,⁴³ methylkit,⁴⁴ and NanomethViz.⁴⁵

2.9 | Differential methylation analysis

Differentially methylated CpG sites were identified using DSS,⁴⁶ 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. As Nanopolish groups CpGs within 10 bp of each other into a single methylation call, we only assigned a methylation value to the most 5' CpG within a given cluster to avoid artificial inflation of depth. Individual differentially methylated CpG sites were considered significant if they reached a false discovery rate (FDR) of 0.05 or less for both the GRCh38 and CHM13 genomes. DSS was then used to identify and aggregate significant neighboring CpG sites (uncorrected *p* value less than 0.05) into differentially methylated regions (DMRs) compared to the control.⁴⁷ AreaStat, which represents the sum of test statistics for CpG sites within a DMR, was calculated for Braak III vs. Braak 0 and Braak V/VI vs. Braak 0. AreaStat allows the ranking of DMRs, with higher values indicating a higher likelihood that the CpG is differentially methylated. A sample CpG site methylation frequency was only included if the call was supported by three or more Nanopolish-called sites. Comparisons were made for Braak III vs. Braak 0 and Braak V/VI vs. Braak 0. Alu elements and LTR5Hs with the largest changes in methylation between groups were then manually inspected using the GRCh38 GeneHancer Regulatory Elements and the Gene Interaction table⁴⁸ within the UCSC genome browser⁴⁹ 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 (> 6000 bp), LTR5_Hs/A/B (> 900 bp), LINE1-Hs (> 5900 bp), AluYa5/Yb8 (> 280 bp), and SVA-E/F (> 1000 bp).⁴³ Other repetitive elements such as centromeric satellite regions were required to have 10 or more CpG sites called within a given region. A promoter region was only counted once in each direction even if multiple DMRs fell within the promoter region.

2.10 | Metachromosome analysis

Metachromosomes were generated using a custom R script following previously described methods.^{50,51} First, a bed file was created with Bedtools⁴⁰ to obtain 100 kb windows across the CHM13 genome

(excluding sex chromosomes and mitochondria). GRange objects³⁸ were then created for centromere regions, cytoband information, and biological values to plot DMR or insertion coordinates. The overlap of centromeres, cytobands, and genomic windows was then determined by ranking each genomic window based on its chromosome arm and position to standardize distances from the p arm telomere (rank = 1) to the centromere (rank = 0) and q arm telomere (rank = -1) to the centromere. Windows that fully overlapped with the active alpha satellite region of the centromere were assigned a rank value of zero. For plotting of DMR centromere and non-centromere ranges, a rolling average of methylation changes was calculated per window over an 11-window span (~1 Mb) for both centromere and non-centromere ranges. Results were combined and normalized by segmenting chromosome positions into 500 meta-bins ranging from 0 (q arm telomere) to 500 (p arm telomere), allowing for the grouping of chromosomal windows across chromosomes of different lengths. For each meta-bin group, the average methylation change was calculated and plotted across the bins. For each meta-bin group, a line plot for each chromosome shows the average methylation change across meta-bins, marking centromeric (bins 249–251) and telomeric boundaries (bins 0–1 and 499–500), and overlaying reference lines for average DMR. The same pipeline was followed for analysis of singleton insertions in Braak 0, Braak III or Braak V/VI, except that raw counts were taken per window rather than rolling averages. The average insertion count per bin was then calculated and normalized to the mean insertion count per window (excluding windows with no insertions).

2.11 | Enrichment analysis and statistical tests

A permutation test (regioner function permTest⁵²) was used to determine if there were more differentially methylated regions within centromeres and rDNA arrays than what would be expected by chance. All graphs were generated with R packages ggplot2,⁵³ TreeMap or adapted from IGV.³² Lola⁵⁴ was used for genomic region enrichment with UCSC datasets. KaryotypeR was used for karyotype graphs.⁵⁵ Gene lists for promoter methylation analyses were analyzed for KEGG pathway⁵⁶ enrichment using ClusterProfiler.⁵⁷ One-way analysis of variance (ANOVA) and Student's t-test were performed in base R or with the ggstatsplot⁵⁸ R package.

3 | RESULTS

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

DNA was extracted from isolated nuclei of *post mortem* frontal cortex of six individuals at Braak 0, six individuals at Braak stage III and six individuals at Braak stage V/VI with a clinical diagnosis of Alzheimer's disease (Table S1). The 18 human frontal cortex samples utilized in this study had an average age of 76.6 years (Table 1). Sequencing was performed using the Oxford Nanopore Technologies PromethION

sequencing platform. We obtained a mean read quality score of 11, with a mean sample N50 of 23.2 kbp and an average genomic coverage of 7.38X (Table S2). ANOVA testing reveals no statistically significant difference among Braak 0, III, and V/VI for *post mortem* interval (PMI), total bases sequenced, read quality, read length or genome coverage. While the age of individuals at Braak III or Braak V/VI do not significantly differ from Braak 0, post-hoc t-test indicates that the age of individuals at Braak III is significantly higher than those at Braak V/VI.

3.2 | Detection of retrotransposon singleton insertions in the aged human brain

We first utilized our long-read sequencing data across all 18 human brain samples to identify retrotransposon sequences that are not captured in the new CHM13 human telomere-to-telomere genome (File S1). TLDR,⁴³ a tool designed to detect retrotransposon insertions using long-read sequencing data, identified 642 retrotransposon sequences that were unique to a single sample and are lacking in the CHM13 reference genome (Figure S1A). These non-reference genome retrotransposons may result from retrotransposition events that occurred in an individual in utero or during life or could represent retrotransposons that are polymorphic within the human genome that the individuals used to generate reference genomes happened to lack.

We next identified potential somatic retrotransposition events marked by singleton insertions (insertions only supported by a single read) following methods similar to Siudeja and colleagues.⁵⁹ As the site of L1-mediated retrotransposon insertion is random, *de novo* insertions should be present as (1) singletons if they occur in post-mitotic cells, or (2) low frequency variants if they occur in dividing cells. Of all non-reference insertions unique to a given individual, we identified 54 as singletons based on presence of at least ten additional reads covering the location of the insertion, all of which lacked the insertion, and presence of TSD-like flanking sequences indicative of L1-mediated transposition (File S2). Intact TSDs are a signature of a recent target-primed reverse transcription event that consists of a duplication of a short DNA sequence near the target site (canonically AA/TTT) of insertion on either side of the inserted element.^{60–64} Among retrotransposon families, the Alu subfamily of SINE elements was the most abundant class of TSD-harboring singleton insertions among our human brain samples; AluYa5 was the most active member within the Alu subfamily (Figure 1A). Interestingly, the majority of singleton insertions (unlike non-singleton insertions) were located in centromeric and pericentromeric regions of the genome, and predominantly affected chromosomes 11, 14, 16, and 20 (Figure S1B, File S2). Further analysis of singleton Alu insertions with respect to centromeric repeats (Figure 1B) revealed that they were predominantly located in alpha satellite active highly ordered repeat (HOR) regions, while singleton L1 insertions were more frequent in inactive HORs (Figure 1C).

As some LINE and SINE retrotransposons are transpositionally active in somatic cells, including postmitotic neurons,^{65–68} we used previously established criteria^{43,69,70} to characterize L1 endonuclease

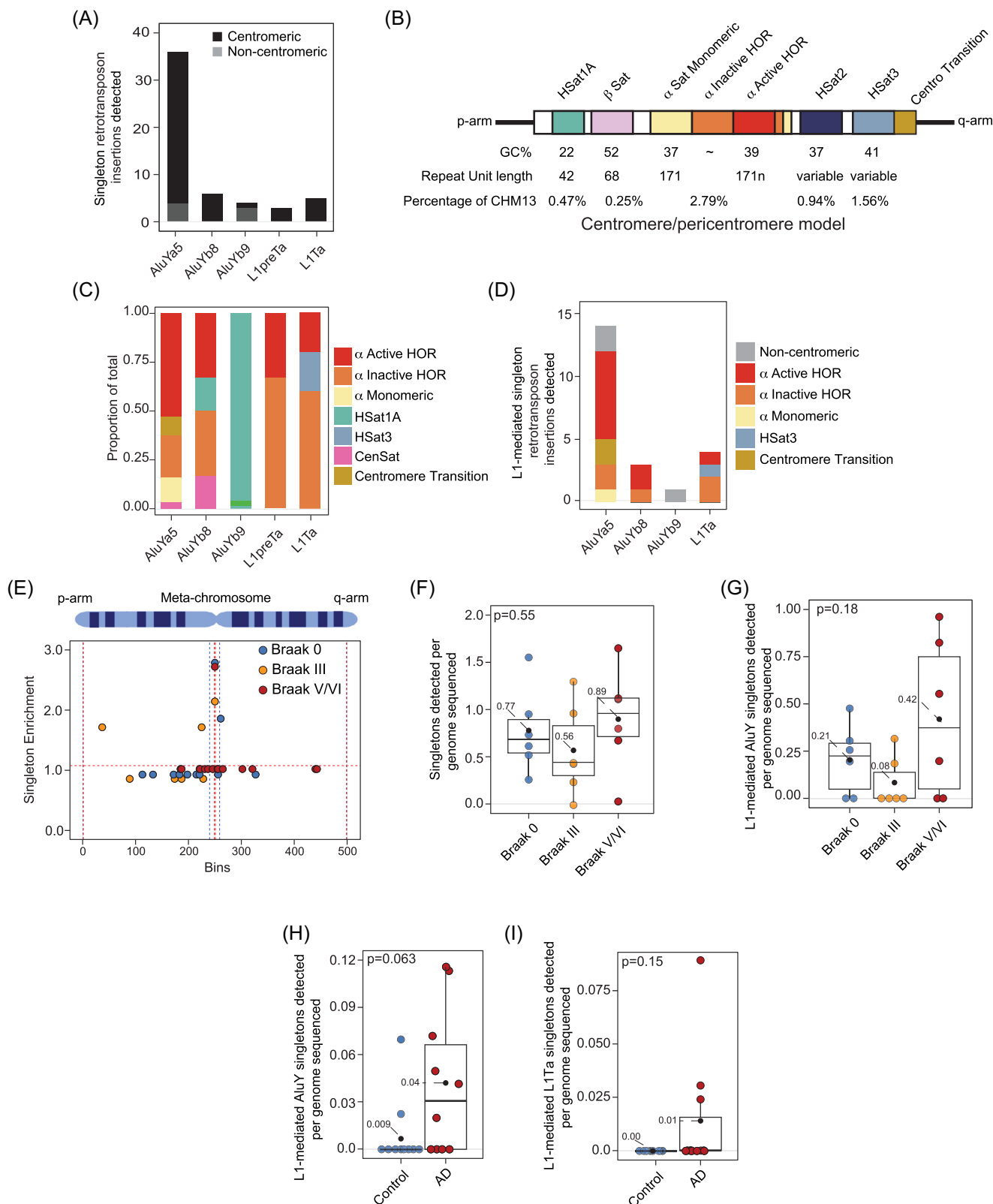


FIGURE 1 Enrichment of putative singleton retrotransposon insertions in the centromere of the aged human brain. (A) Total counts of all singleton retrotransposon insertions by subfamily grouped by whether they are present in the centromere. (B) Graphic model of centromeric and pericentromeric regions. (C) Barplot of the fraction of singleton insertions detected within functional centromeric repeats. (D) Total counts of all L1en-mediated singleton retrotransposon insertions by subfamily, grouped by centromeric repeat. (E) Mean count of retrotransposon singleton insertions detected across all autosomes, normalized for chromosome length plotted across a metachromosome. (F) Boxplots of singleton retrotransposon insertion counts normalized to sequencing depth. (G) Boxplots of putative L1en-like mediated singleton AluY insertion counts normalized to sequencing depth. (H) Boxplots of NeuN-sorted putative L1en-like mediated singleton AluY insertion counts normalized to

TABLE 1 Brain demographics and sequencing metrics

Parameter	Braak 0	Braak III	Braak V/VI	All	Range	p-Value
Total # sequenced	6	6	6	18		
Age	76 years	82.3 years	71.3 years	76.6 years	67–92 years	0.03*, Braak III vs. V/VI
% Female	50%	50%	33%	44%		
Post mortem interval	10.33 h	14.67 h	8.75 h	11.89 h	5–24 h	0.63
	Mean	Range	p-Value			
Total bases sequenced	25.5 Gbp	14.2–41.6 Gbp	0.66			
Read quality	11.7	11.5–11.9	0.21			
Read length N50 (kbp)	23.2 kbp	9.66–31.58 kbp	0.66			
Genome coverage	7.4X	4.1X–12.1X	0.65			

Note: Sample characteristics and associated sequencing statistics.

*Represents the *p* value obtained from the post hoc *t*-test.

(L1en)-mediated insertion events from evolutionarily “young” elements. After hand-filtering our list of singleton insertion events to include reads of 10 kbp or longer that harbor characteristics of L1en-mediated transposition, including presence of a polyA tail, an L1en motif at the predicted insertion location and manual inspection of insertions using IGV,³² we detected 22 putative L1en singleton insertions across the 18 brains analyzed, most of which belong to the AluYa5 subfamily (Figure 1D). While most detected Alu insertions were intact, we found that L1en-mediated L1Hs singleton insertions were severely truncated at the 5' end (File S2), consistent with a 3' end bias of insertion.⁷¹ While we attempted to determine the genomic source of L1Hs singleton insertions via Blat alignment of the insertion sequences against the CHM13 genome, alignments were ambiguous due to lack of transduction sequences. We observed significant enrichment (*p* value of 1E-15, Fisher's exact test) of L1en-mediated insertions in centromeric regions of the genome, with 17 occurring in HOR segments of the centromere.

To assess retrotransposition in brains affected by Alzheimer's disease, we next compared overall singleton insertions and L1en-mediated singleton insertions in brains at Braak III and IV/V versus controls. All three groups featured enrichment of singleton insertions within the centromere when plotted across a metachromosome (Figure 1E). While abundance of general singleton insertions did not differ among Braak stages (Figure 1F), restricting our analysis to L1en-mediated singletons revealed a trending increase (*p* value of 0.18, total of 18 insertions) of AluY, but not L1Hs (*p* value of 0.92, total of 4 insertions) insertions in brains with the highest degree of tau burden (Figure 1G, Figure S1C).

To determine if somatic retrotransposition occurs specifically in neurons affected by Alzheimer's disease, we analyzed nanopore long-read DNA sequencing data of NeuN-sorted neuronal nuclei from a

publicly available Alzheimer's disease dataset with 30X sequencing coverage²⁸ (File S2). Increased coverage and analysis specifically in neurons revealed a stronger trend of increased AluY retrotransposition in Alzheimer's disease compared to control (*p* value of 0.063, total of 24 insertions) (Figure 1H). While potentially somatic insertions of L1Hs did not significantly differ between Alzheimer's disease cases vs. controls in neurons (*p* value of 0.15, total of three insertions), L1Hs singletons were only detected in samples affected by Alzheimer's disease (Figure 1I). L1en-mediated singleton insertions were only detected in the centromere/pericentromere (File S2). While centromeric/pericentromeric singletons had relative low mapping quality (MAPQ) scores of one, the absence of L1en-mediated singletons in many samples suggests that detection of such insertions is not simply a technical artifact. Mean estimated L1en-mediated singleton insertion rates for 7X coverage bulk sequencing and 30X coverage neuronal datasets were 0.74 and 0.0245 per cell, respectively. Taken together, our analyses suggest that specific AluY and LINE-1 family members are mobile in cells of the aged human brain and/or their precursor cells, that *de novo* insertion events are rare and occur more frequently in the centromere/pericentromere, and that the AluY element may mobilize to a greater extent in brains affected by Alzheimer's disease.

3.3 | NAHR between repetitive elements in the aged human brain

Repetitive element-derived structural variants can arise from NAHR events between homologous high-copy repeats in the genome.⁷² We utilized TE-reX,²⁸ a tool developed to detect NAHR events between transposable elements, to identify singleton retrotransposon-

sequencing depth. (I) Boxplots of NeuN-sorted putative L1en-like mediated singleton L1Hs insertion counts normalized to sequencing depth. Statistical analysis was done via analysis of variance analysis for F and G, and a Student's *t*-test for H and I. Black dots on boxplots represent the group mean normalized insertion value.

associated NAHR events within our set of 18 human brains. We utilized CHM13 and GRCh38 human reference genomes to leverage their respective strengths: assembly completeness (CHM13) vs. more thorough annotation (GRCh38).

We first identified recombination hotspots from a total of 84,280 putative singleton events across the CHM13 genome by counting singleton NAHR events within 0.1 Mbp windows and normalizing counts by the number of repetitive elements present within the window.²⁸ To obtain a genome-wide view of NAHR, enrichment values were plotted along a metachromosome, which revealed that centromeric/pericentromeric regions exhibited the highest normalized enrichment values (Figure 2A), consistent with previous work.²⁸ We then determined the distribution of NAHR enrichment values for large chromosomal regions, which revealed increased recombination activity primarily in acrocentric chromosome arms, rDNA arrays, and centromeric/pericentromeric regions compared to other genomic areas (Figure 2B, Figure S2A). Most detected recombination events were predicted to result in interchromosomal rearrangements mediated by SINEs (Figure S2B, C), consistent with previous findings in neurons isolated from *post mortem* human brain.²⁸

Given the high enrichment of NAHR in centromeric and pericentromeric regions, we analyzed NAHR within the functional repeat arrays that compose these regions. NAHR events were detected in HOR repeats (monomeric and divergent), classical human satellite regions, beta satellites, and other centromeric repeat regions (Figure 2C). Interestingly, we found that NAHR within HORs were composed primarily of L1-L1 recombination events, while NAHR within beta and classical human satellites reflected recombination between Alu elements (Figure 2D). We also detected unique individual regions of enrichment in many peritelomeric and telomeric regions such as the telomeric region of chromosome 17, as well as intergenic segmental duplications such as those present in chromosomes 4 and 6 (File S3).

Strikingly, 475 rDNA arrays were found to have higher median recombination enrichment compared to other NAHR-prone regions such as telomeres or the centromere. While the majority of detected NAHR rDNA events occurred between different chromosomes, we also detected intrachromosomal NAHR events measuring 35 kbp, 45 kbp, and 95 kbp in length (Figure 2E, Figure S2D) that reflect 47S rDNA that has been deleted, inverted, or duplicated. While insertion and inversion NAHR rDNA events were marked by Alu-Alu recombination, deletion events appeared to stem primarily from LINE-LINE NAHR events, with additional contribution of rRNA-HSA repeats (File S3).

While the overall frequency of repeat-associated NAHR events did not significantly differ among Braak stages (Figure S3A-O), varying values of NAHR enrichment could be discerned in distinct regions of the genome based on Braak stage (examples provided in Figure 2F, G, full karyotype maps provided in Figure S4A). UCSC dataset enrichment analysis further revealed significant changes in NAHR localization (112,108 singleton events) to genomic regions consisting of simple repeats, microsatellites, CpG islands, and annotated segmental duplications at Braak III (Figure 2F, G).

3.4 | Structural variants in the aged human brain

We next leveraged our dataset to identify complex, likely germline-derived repetitive structural variants in the human brain using the more thoroughly annotated GRCh38 genome (Figure S5A, File S4). Structural variants were most frequently detected in introns and exons (Figure S5B). Length distribution of structural variants was highly variable, with insertions having a larger median length compared to other variants (Figure S5C). We next utilized SNPeff⁷³ to predict how these variants may affect gene function. Most structural variants were classified as modifier variants in non-coding genes/intergenic regions or moderate (non-disruptive) impact variants (Figure S5D). We detected structural variants in genes associated with neurodegeneration, including an intronic deletion/inversion in *APP*, a tandem duplication in *ATXN1*,⁷⁴ and a deletion in *MAPT* (File S5).

We next identified non-reference structural variants that fell within the 6054 dark regions of the human GRCh38 reference genome.⁴ Among the 18 brains analyzed, we detected 1123 insertions in the dark regions of 787 genes, 60 of which occurred in dark regions of exons. 1781 deletions were associated with dark regions among the 18 brains analyzed. Interestingly, 23 Alzheimer's disease risk genes contained dark regions with previously-undocumented structural variants, some of which were in close proximity to single nucleotide polymorphisms (SNPs) defined as risk variants for neurodegenerative diseases based on the Alzheimer's Disease Sequencing Project (ADSP)⁷⁵ 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 *ATP binding cassette subfamily A member 7* (*ABCA7*) (Figure S5E), which encodes an ABC transporter that regulates lipid metabolism,^{76,77} amyloid processing, and clearance.⁷⁸⁻⁸⁰ A deletion and an insertion were detected within a dark region of an intronic SVA retrotransposon within vasoactive intestinal peptide receptor 2 (*VIPR2*), near an Alzheimer's disease-associated SNP overlapping a deletion present in 10 out of 18 brains analyzed (Figure S5F). Similarly, most brains analyzed harbored a deletion overlapping a dark intronic AluYa5 element within chromodomain helicase DNA binding protein 2 (*CHD2*) in close proximity to an Alzheimer's disease-associated SNP (Figure S5G, File S5). Dark region insertions were also detected near SNPs associated with amyotrophic lateral sclerosis, multiple system atrophy, multiple sclerosis, and spinocerebellar ataxia. Many non-reference insertions and deletions within dark regions were shared among all samples, likely reflecting common circulating variants, while others were only found in single samples, reflecting less common circulating variants, new variants that occurred in the germline, or variants of somatic origin.

We next identified structural variants unique to Braak 0, Braak III or Braak V/VI samples using the CHM13 and GRCh38 reference genomes. We required that the given locus was genotyped in all 18 samples to ensure each variant was unique to a particular group. The number of structural variants detected per sample did not significantly differ

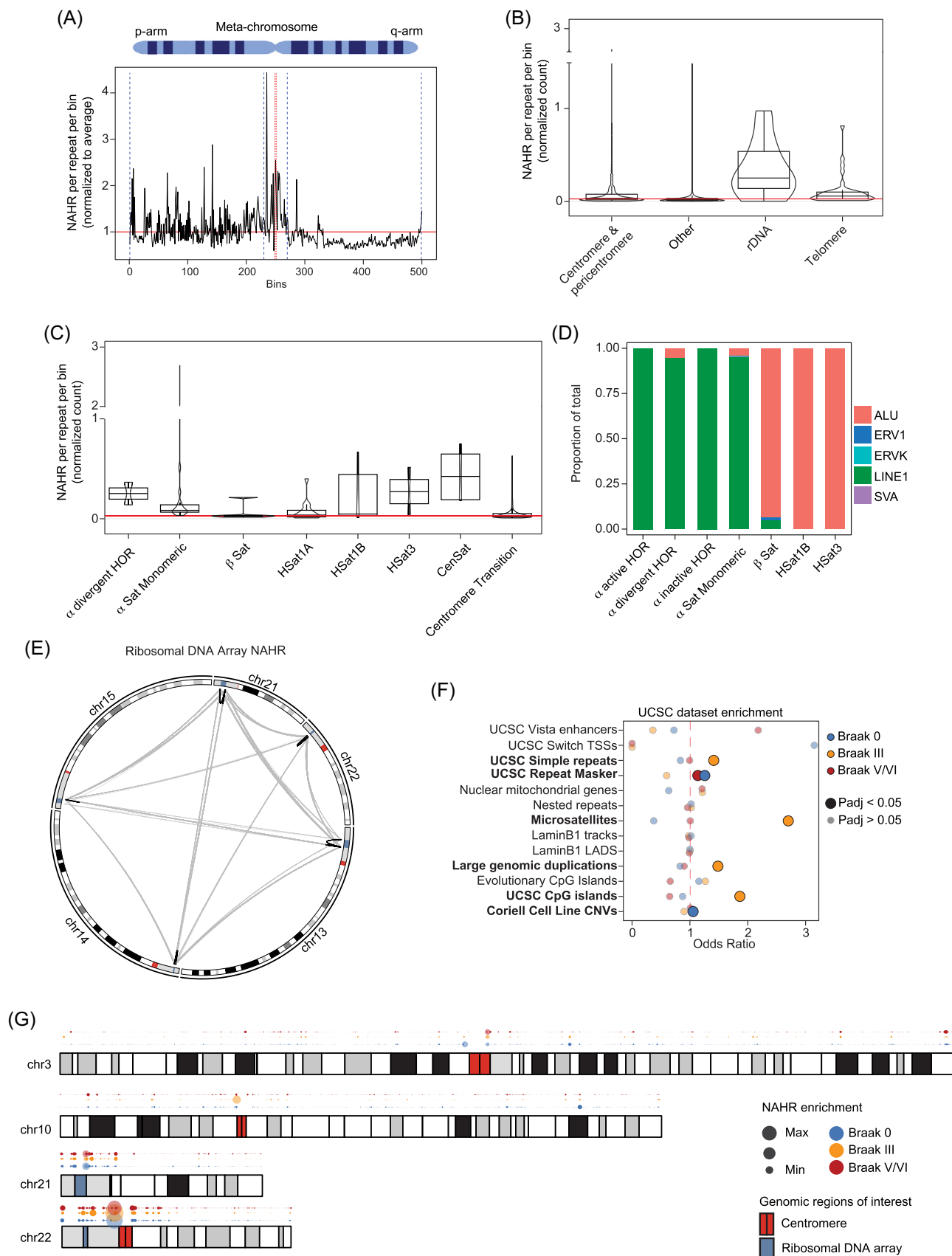


FIGURE 2 NAHR in the aged human brain. (A) Metachromosome of the rolling mean of NAHR enrichment values across all autosomes, normalized for chromosome length. (B) Boxplots of singleton NAHR enrichment values for different chromosomal structures. (C) Boxplots of singleton NAHR enrichment values for different centromeric repeat structures. (D) Barplot of transposable element-associated NAHR events

between Braak 0, Braak III, and Braak V/VI brains (Figure S6A, B), nor did structural variant length (Figure S6C). These results also hold true when Braak III and V/VI were combined into one group and compared to Braak 0. Despite the lack of change between Braak stages among the small number of brains analyzed, the overlap of structural variants within dark regions of genes associated with human disease highlights the value of approaches that resolve complex regions of the genome and identifies new variants that can be specifically targeted for analysis in a larger cohort of human samples.

3.4.1 | DNA methylation in the aged human brain

We next analyzed 5mC DNA methylation using the CHM13 and GRCh38 genomes. A sample from an individual with a previous cancer diagnosis was an outlier in terms of 5mC methylation and was thus removed from subsequent analysis (Figure S7A). We focused on CpG sites with a 5mC call in at least three samples, resulting in 24.8 million and 23 million CpG sites for analysis based on the CHM13 and GRCh38 genomes, respectively.

We used the DDS callDMR function to leverage single-nucleotide CpG resolution and the documented increase in detection power when combining differentially methylated CpG sites. Changes in uncorrected DMRs across the CHM13 genome were analyzed for Braak III vs. Braak 0 (93,441 DMRs) and Braak V/VI vs. Braak 0 (88,423 DMRs) (Figure S7B, C, Files S6, S7), and the rolling average of the change in methylation of statistically significant DMRs was plotted across a metachromosome (excluding sex chromosomes) for each comparison. While the average change in methylation was decreased in Braak III and elevated in Braak V/VI compared to Braak 0 (Figure 3A), some regions shared a conserved directional change in Braak III and Braak V/VI compared to Braak 0. Centromeric/pericentromeric regions, for example, were demethylated at Braak III and V compared to Braak 0, albeit to varying degrees. Indeed, a permutation test revealed that centromeric methylation was significantly dysregulated in tau-affected brains compared to control (Figure 3B, Figures S7D-H).

The DSS area statistic (areaStat), which sums the test statistics for CpG sites within a DMR, was calculated for Braak III vs. Braak 0 and Braak V/VI vs. Braak 0, taking into account both the length and the number of CpG sites with differential methylation.⁴⁶ A larger area suggests a higher probability that the region is a genuine DMR. We determined the distribution of DMRs across autosomes in Braak III and Braak V/VI cases by plotting areaStat of DMRs on a karyotype graph (Figure S8A, B; Figure 3C, D include examples within chromosomes 13, 15, 21, and 22). We identified many DMRs that are shared between early- and late-stage disease based on magnitudes of change

in the same or opposite direction, some of which overlapped with rDNA array model regions of the CHM13 genome. Indeed, a permutation test revealed that 475 rDNA regions were significantly enriched for DMRs in Braak III and Braak V/VI compared to Braak 0 (Figure 3E). DMRs overlapped some but not all 475 rDNA copies across all five rDNA array regions, with the largest magnitude of disease-associated change being hypomethylation.

We next analyzed 5mC changes within promoters across Braak stages using the GRCh38 genome. We detected differential methylation of 20,352 unique promoters of 7852 genes at Braak III compared to Braak 0, while 19,638 promoters of 5765 genes were differentially methylated at Braak V/VI (File S8). We observed a subtle change in the bimodal distribution of 5mC modifications in differentially methylated promoters (Figure S9A, B), in line with promoter element MDS analysis.⁸¹ DMRs in brains at Braak III had higher densities at 0.25 and 0.75 compared to Braak 0, while brains at Braak V/VI shift towards no methylation (0) or complete methylation (1). While promoter methylation differs among brains at Braak 0, Braak III, and Braak V/VI, we identified over 7072 hypomethylated promoters that were unique to brains at Braak III related to KEGG pathways⁵⁶ related to phospholipase D, FC gamma receptor mediated phagocytosis, gonadotropin-releasing hormone (GnRH) signaling pathway, Rap1 signaling pathway and actin cytoskeleton regulation (Figure S9C). The ~3000 hypermethylated promoters that were shared between Braak III and Braak V/VI brains related to pathways involving calcium and oxytocin signaling, focal adhesions, phospholipase D and glutamatergic synapses (Figure S9D).

We next focused on DNA methylation within “dark” regions of the genome to fully leverage our long-read approach. We detected 1150 and 1333 DMRs at Braak III vs. Braak 0, and Braak V/VI vs. Braak 0, respectively. Four DMRs in Braak III vs. Braak 0 fell within dark regions of Alzheimer's disease risk genes (*ABACA7*, *AMY1A*, *CHRFAM7A*, *CR1*). DMRs within *AMY1A*, *CHRFAM7A*, and *CR1* were also present at Braak V/VI (File S8). 29 5' UTRs were differentially methylated at Braak III; 37 5' UTRs were differentially methylated at Braak V/VI, including hypomethylation of Alzheimer's disease risk genes *AMY1A* and *FMR1* within dark regions in their 5' UTR at Braak V/VI.

3.4.2 | DNA methylation within transposable element sequences of the human brain

Consistent with prior observations,^{82–84} we found that DNA methylation levels correlated with the evolutionary age of retrotransposons, with younger elements exhibiting higher methylation frequencies than older elements. With the exception of human endogenous retrovirus

among different centromeric repeat structures. (E) Circular karyotype map depicting interchromosomal (gray) and intrachromosomal (black) NAHR between rDNA arrays. (F) Dotplot of NAHR event enrichment on different UCSC genomic sets based on Braak stage. (G) Karyotype map depicting hotspots of NAHR based on Braak stage. Red and blue rectangles within chromosomes represent centromeres and rDNA arrays, respectively. Red lines in A, B, C indicate the average NAHR value across the genome. NAHR, non-allelic homologous recombination; rDNA, ribosomal DNA; UCSC, University of California at Santa Cruz.

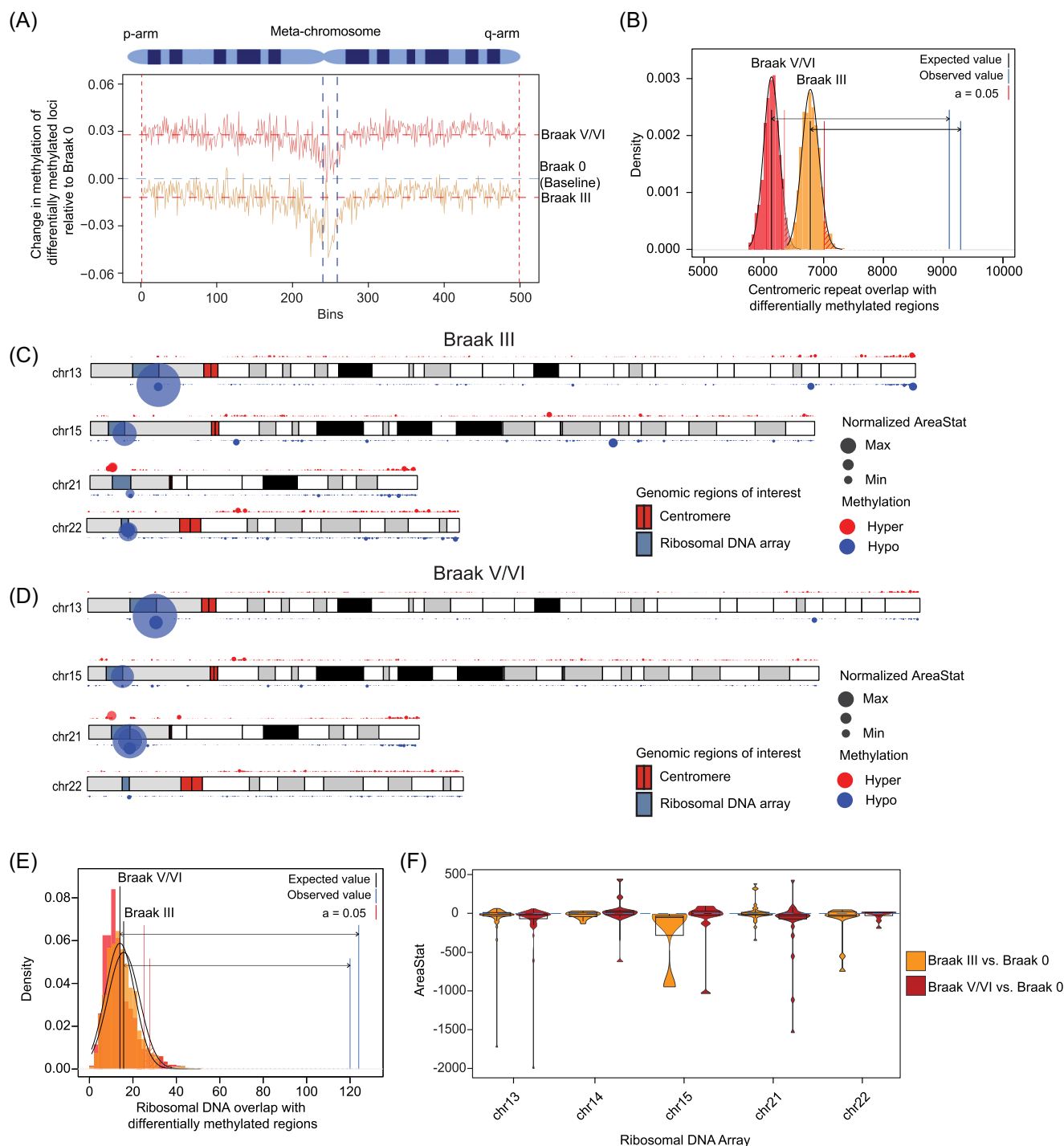


FIGURE 3 DNA methylation in brains affected by Alzheimer's disease. (A) Metachromosome of the rolling mean of the difference in methylation values of differentially methylated regions across all autosomes and normalized for chromosome length for Braak III vs. Braak 0 and Braak V/VI vs. Braak 0. Orange and red horizontal dashed lines represent the average change across the metachromosome for Braak III and Braak V/VI, respectively. The blue horizontal dashed line represents the baseline value of Braak 0. (B) Permutation test of differentially methylated regions localized to the centromere in Braak III vs. Braak 0 and Braak V/VI vs. Braak 0, $p = 0.0001$. (C, D) Karyotype of example chromosomes with significant changes in methylation in Braak III and Braak V/VI. Red and blue rectangles within chromosomes represent centromeres and rDNA arrays, respectively. (E) Permutation test of differentially methylated regions localized to rDNA arrays in Braak III vs. Braak 0 and Braak V/VI vs. Braak 0, $p = 0.0001$. (F) Box/violin plot of areaStat values for differentially methylated regions between Braak III vs. Braak 0 and Braak V/VI vs. Braak 0. rDNA, ribosomal DNA.

(HERV)-K elements, retrotransposons were found to be heavily methylated in the aged human brain, with a mean methylation value exceeding 0.75 (Figure S10A-E). Young Alu and LINE retrotransposon family members had generally higher levels of 5mC compared to older Alu and LINE family members, perhaps reflective of an increased need to repress their transpositional capacities. HERV families and other long terminal repeat (LTR) retrotransposons featured less methylation of younger elements.

While previous studies report widespread changes in 5mC at different stages of Alzheimer's disease in the frontal cortex,⁸⁵ methylation within dark regions has not been studied in brains of patients with Alzheimer's disease. An MDS plot of CpG methylation frequency in autosomal repetitive elements revealed no clustering based on sex or Braak stage (Figure S11A). While we did not detect differences in 5mC methylation of retrotransposon subfamilies between Braak stages when retrotransposons were grouped *en masse*, similar to previous studies in the blood of individuals with Alzheimer's disease,^{86,87} methylation within repetitive regions was clearly dysregulated at the level of individual retrotransposon loci (Figure S11B-H). Distribution of 5mC modification in differentially methylated repetitive elements shifted in opposite directions according to Braak stage, with generally less methylation at Braak III and generally increased methylation at Braak V/VI (Figure S11B).

We analyzed the methylation areaStat for select families and subfamilies of retrotransposons to more precisely quantify the degree of differential methylation. We observed changes in areaStat for L1 elements (with some loci accounting for large magnitude changes), with most of the more significant changes reflective of hypomethylation (Figure 4A). We then separated the L1 family into L1Hs (young) and ancestor-shared (old) L1 subfamilies; distribution of L1Hs and ancestral L1 subfamily methylation was similar among Braak stages (Figure S11C). We observed similar changes in methylation areaStat for AluYa/Yb, HERV-K, and SVA subfamilies (Figure 4B-D). Given that LTR elements can function as alternative promoters for genes and as promoters for HERVs, we analyzed methylation patterns of the LTR12 and LTR5 families. Within the LTR12 family, LTR12C exhibited the most pronounced changes in methylation of specific loci compared to LTR12E and other LTR subfamilies. Similarly, the younger LTR5 family demonstrated milder, though significant, methylation changes compared to LTR12 subfamilies (Figure 4E). We also detected differential methylation in several solo LTR5Hs and Alu elements that are predicted to be regulatory regions based on the GeneHancer database (Table S3).

We then analyzed changes in methylation of selected differentially methylated non-fragmented (> 5900 bp) L1Hs retrotransposons on a per-sample basis. Four L1Hs loci (Figure S12A-E) featured demethylation at the 5' UTR, suggesting that these elements may be source loci for L1Hs transcripts in the human brain. When comparing among Braak stages, we found statistically significant hypomethylation in the promoter of individual L1Hs loci such as an intergenic L1-Ta (transcriptionally active) (7q34) and an intronic L1-Ta in *SPOCK3* (4q32.3) at Braak III compared to Braak 0 (Figure 4F, G, File S9). Our analysis also revealed examples of LINE hypermethylation at Braak III (Figure 4H).

We detected promoter demethylation of an intact L1-Ta element promoter,⁴³ regardless of Braak stage, that is reported to be a source of somatic retrotransposition in the developing hippocampus (Figure S12C).^{20,88} Several LTR5Hs loci, in both provirus and solo forms, were differentially methylated at Braak III and/or Braak V/VI compared to Braak 0 (Figure S13A, B, File S9), with some hypomethylated 5' LTR5Hs loci adjacent to proviral HERV-K elements. For example, the 5' LTR5Hs of HERV-K107 (7p22.1) was found to be modestly hypomethylated in Braak III and Braak V/VI and has been previously implicated in the pathogenesis of human diseases.⁸⁹⁻⁹¹ Overall, our analyses provide novel insights into methylation of individual retroelements and highlight the importance of analyzing such elements at the locus level rather than *en masse*.

4 | DISCUSSION

While previous studies have identified risk variants and epigenetic dysregulation in Alzheimer's disease,^{1,92,93} technical limitations result in exclusion of many dark/repetitive regions from analysis. Long-read sequencing, associated bioinformatic tools, and a more complete human genome assembly help overcome these challenges. Nanopore sequencing-based analysis of retrotransposon insertions, retrotransposon-associated NAHR, structural variants, and 5mC modification in *post mortem* human Alzheimer's disease brains revealed genomic and epigenomic changes that concentrate in retrotransposon-rich centro/pericentromeric regions and rDNA arrays. Our findings build upon a long-read sequencing study of DNA isolated from neuronal nuclei of control and late-stage Alzheimer's disease brains.²⁸ Inclusion of Braak III tissue allowed us to capture changes occurring in early disease; leveraging the telomere-to-telomere reference genome and methylation data allowed us to comprehensively interrogate genomic dark regions.

Long-read sequencing-based quantification of L1en-mediated retrotransposition events in the aged brain revealed insertion rates of 0.145 and 0.0147 per cell for 7X coverage bulk sequencing and 30X coverage neuronal sequencing, respectively. This rate of insertion is higher than L1Hs estimates based on whole genome amplification^{20,21,94} but less than estimates of 13.7 per neuron and 6.5 per glia based on retrotransposon capture sequencing.²² We note that our estimations include L1Hs and Alu elements, whereas previous studies focused only on L1Hs. We find that AluYa5 is the most transpositionally active retrotransposon in the human brain, consistent with data from the 1000 genomes project.⁹⁵ In line with previous research indicating that centromeres function as sinks for retrotransposon insertion,⁹⁶⁻⁹⁸ we identified significant enrichment of retrotransposon insertions within centromeric alpha satellite HOR arrays.

Elevation of retrotransposon transcripts in Alzheimer's disease and progressive supranuclear palsy, a primary tauopathy, along with increased transposition in tau transgenic *Drosophila*^{12-14,99-102} suggested that retrotransposons are derepressed and may mobilize in human disease. Our shallow (7X) long read sequencing reveals a

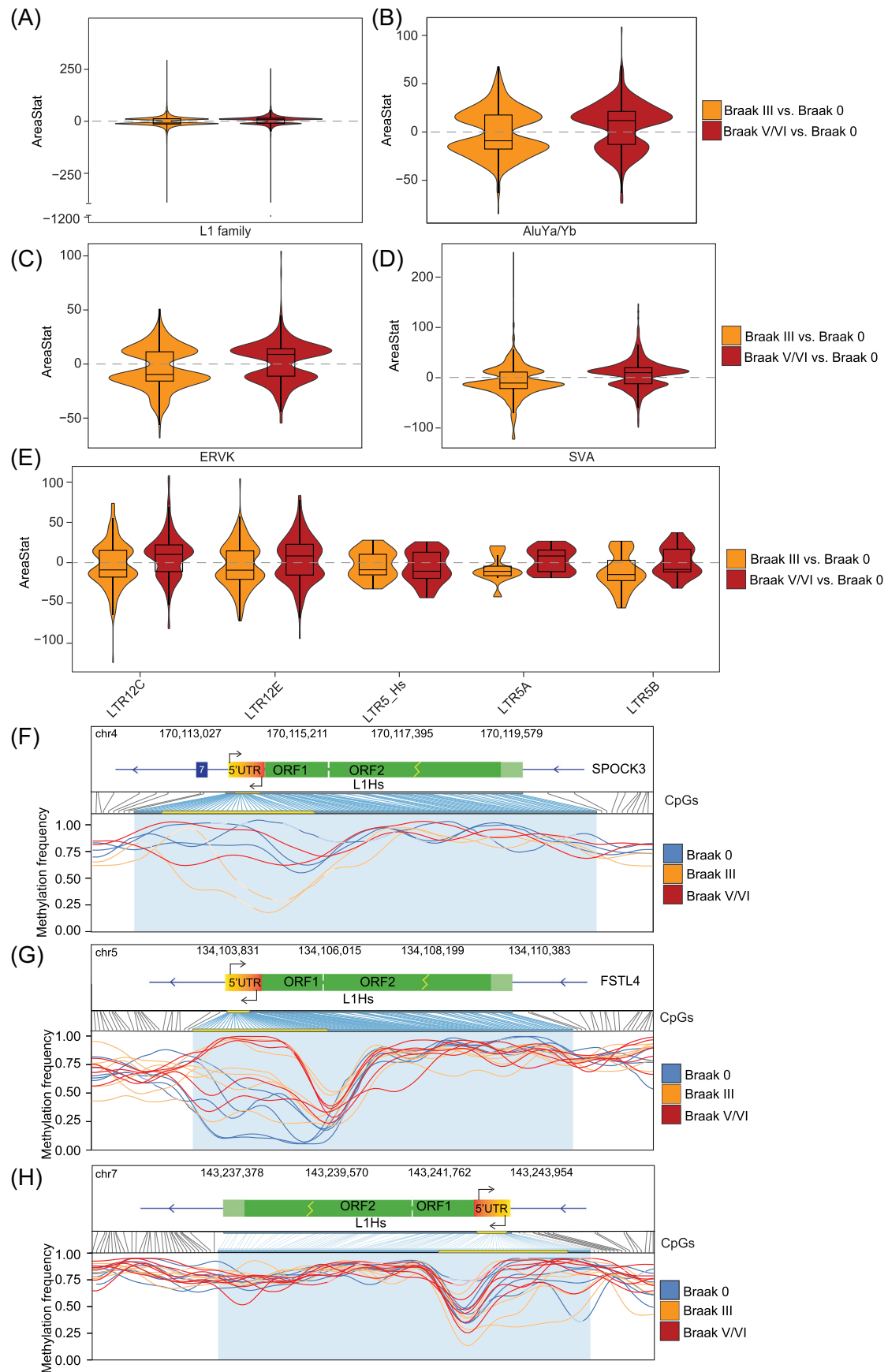


FIGURE 4 DNA methylation patterns of retrotransposons in Alzheimer's disease. Box/violin plots of areaStat values for DMRs between Braak III vs. Braak 0 (orange) and Braak V/VI vs. Braak 0 (red) for (A) LINE-1 family (B) AluYa/Yb, (C) ERVK subfamily, (D) SVA subfamily, (E) and LTR sequences. (F, G, H) Methylation frequency of CpG sites across the indicated genomic locus per sample, colored by Braak group. Statistically

trending increase in singleton/L1en-mediated AluY insertions at Braak V/VI; deeper (30X) sequencing of neurons reveals a stronger, but still trending, increase in singleton/L1en-mediated AluY insertions at Braak V/VI, suggesting that AluY elements may be more transpositionally active in Alzheimer's disease. New copies of retrotransposon DNA are generated via reverse transcription of retrotransposon RNA; the reverse transcriptase inhibitor lamivudine (3TC) reduces retrotransposition and neuronal death in laboratory models of Alzheimer's disease and related tauopathies.^{13,15,102,103} In a Phase IIa clinical trial, 3TC was found to significantly reduce levels of glial fibrillary acidic protein (GFAP) in cerebrospinal fluid of individuals with early Alzheimer's disease, suggestive of reduced neuroinflammation.¹⁰⁴ A retrospective analysis of public health records further suggests that reverse transcriptase inhibitors, including 3TC, decrease incidence of Alzheimer's disease.¹⁰⁵ Given the low rate of L1en-mediated singleton insertions detected in our analyses, most of which fall within centro/pericentromeric regions, we speculate that somatic retrotransposon insertion is not a primary factor driving neurodegeneration, and that blocking retrotransposon insertion is not the major mechanism underlying the protective effects of 3TC. As studies in *Drosophila* report that most new DNA copies generated via reverse transcription exist as extrachromosomal DNA,¹⁰⁶ we speculate that retrotransposon-derived extrachromosomal cytoplasmic DNA drives innate immune activation in Alzheimer's disease and related tauopathies, and that 3TC reduces neurotoxicity by limiting production of neuroinflammatory extrachromosomal DNA.

NAHR analyses revealed L1-L1 recombination in alpha satellite monomeric repeats and Alu-Alu recombination in centro/pericentromeric regions. While previous nanopore-based analysis detected increased NAHR between retrotransposons in neurons in late-stage Alzheimer's disease,²⁸ the overall degree of NAHR did not differ between Alzheimer's disease cases vs. controls in our study. This discrepancy may result from neuron-specific changes that are lost when DNA is extracted from total brain lysate. We nevertheless found that NAHR location differs at Braak III, with simple repeats/microsatellites particularly affected. We speculate that NAHR enrichment within repetitive regions at Braak III is due to loss of 5mC and/or other silencing mechanisms, thus increasing access to recombination machinery.

Array-based quantification of DNA methylation in *post mortem* human Alzheimer's disease brain has identified increases and decreases in many different genes,^{85,92,107,108} in line with our analysis of protein-coding regions. In contrast, we found that methylation changes within repetitive regions were directionally consistent. 5mC methylation was globally decreased at Braak III, with centro/pericentromeric regions most affected; centro/pericentromeric 5mC was also reduced compared to chromosome arms at Braak V/VI. Studies utilizing late-stage Alzheimer's disease brain report DNA

hypomethylation based on 5mC immunoreactivity,^{109,110} whereas array-based approaches suggest DNA hypermethylation in Alzheimer's disease.^{85,111,112} We speculate that divergent findings result from the high degree of methylation within centro/pericentromeric regions of control brains, which is detected using immunoreactivity-based approaches but is lost with arrays that do not include repetitive genomic regions. Previous array-based platforms also cannot differentiate between 5mC and 5hmC, which is considered an activating DNA modification that occurs in opposition to 5mC.¹¹³

Our findings add a new dimension to studies reporting loss of histone modifications and proteins associated with constitutive heterochromatin in tauopathy models and in neurons isolated from late-stage human Alzheimer's disease brain.^{114–116} While we do not know the impact of centro/pericentromeric DNA demethylation in Alzheimer's disease, studies in cancer and Immunodeficiency, Centromeric Instability, and Facial Anomalies syndrome, a disorder caused by mutation of the centromere-associated *DNA methyltransferase 3B* gene, suggest that DNA methylation regulates centromere function and stability by shaping local chromatin structure and suppressing recombination.^{117–119} As centro/pericentromeric demethylation is associated with aneuploidy and chromosomal rearrangements, we speculate that demethylation of these regions contributes to aneuploidy in Alzheimer's disease.^{120–123} If centromeric demethylation induces production of centro/pericentromere-derived RNA (cenRNA), such species could drive a senescence-associated secretory phenotype.¹²⁴

Aligned with studies in other tissues,^{125,126} we identified rDNA arrays as hotspots for genomic variation. We detected numerous Alu-Alu NAHR events within and between rDNA arrays, consistent with enrichment of DNA breakpoints near Alu sequences within rDNA of cancer cell lines.¹²⁷ Presence of intrachromosomal rDNA NAHR events within 47S rDNA arrays suggest that rDNA duplications and deletions may stem from recombination of full-length 47S rDNA copies and could underlie variations in rDNA copy number reported in other organisms.^{128,129} Our finding that rDNA gene bodies are hypomethylated at early and late stages of Alzheimer's disease contrasts with a targeted bisulfite mapping-based study reporting that rDNA promoters are hypermethylated in Alzheimer's disease.¹³⁰ Differing findings may result from the inability of bisulfite sequencing to differentiate between 5mC and 5hmC. Future studies focused on rDNA NAHR, rDNA demethylation, and the potential for these mechanisms to produce circular rDNA^{128,131} may reveal novel mechanisms that drive neurodegeneration, neuroinflammation, and senescence in Alzheimer's disease. Genomic changes in rDNA could also impact translation rate in settings of tauopathy^{132–135} or translational fidelity.

While previous analysis of structural variants, often relying on DNA copy number and content, suggest that structural variation is

significant changes in CpG methylation detected in promoter regions are highlighted in yellow; yellow lightning bolts denote truncated ORFs. Lighter colors within lines in F and H indicate gaps in sequencing. DMR, differentially methylated region; LTR, long terminal repeat; ORF, open reading frame.

elevated in Alzheimer's disease,¹³⁶ these approaches can miss complex variants entirely. This results in a lack of information regarding structural variants in dark regions and/or retrotransposons that are in linkage disequilibrium with SNPs. We identified several insertions in a dark region composing a variable nucleotide repeat expansion highly correlated with a common high-penetrant risk SNP (rs3764650) for Alzheimer's disease in ABCA7.^{137,138} We also detect overlap between an Alzheimer's disease risk SNP (rs115550680) and a polymorphic AluYa8 element and identify several variants within other known risk genes. While these variants were not restricted to cases with tau pathology, our analyses reflect the tip of the iceberg regarding potential contributions of genetic variants within dark regions to disease. Future studies leveraging deeper long read sequencing and/or targeted sequencing methods in a larger set of individuals will enable population-level variant analysis.¹³⁹

Shallow sequencing combined with non-cell-type specific DNA sequencing presents several limitations. Low genomic coverage likely results in underestimation of genetic variants. For example, we detected a stronger trend of AluY retrotransposition events in individuals with Alzheimer's disease when DNA was sequenced at 30X compared to 7X coverage. Our overall estimated number of 5mC-modified CpG sites (CHM13: 24.8 million, GRCh38: 32.28 million) is lower than the previously estimated number of CpG sites in the CHM13 (32.28 million) and GRCh38 human genome reference (29.17 million),¹⁴⁰ likely reflecting relatively low coverage sequencing of our samples. Additionally, variation in cell type composition may obscure cell-type-specific methylation changes associated with disease; frontal cortex at Braak V/VI, for example, has less neurons and more glia than Braak 0 or III.¹⁴¹ Despite applying stringent criteria to remove false positives for singleton insertions and NAHR and manually inspecting each singleton insertion within alignment files, some observations may represent false positives. This could occur if sequenced reads at specific loci do not reflect the expected 50%/50% makeup of maternal vs. paternal genotype. False positives could result in inflated normalized sequencing values, resulting in estimates higher than those reported in the literature. Our analyses nevertheless provide new insight into genomic and epigenomic changes in Alzheimer's disease, as well as rationale for molecular and sequencing studies focused on genetic and epigenetic changes occurring in the dark genome.

ACKNOWLEDGMENTS

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 was provided by NINDS RF1 NS112391 (BF), NIA R01 AG078964 (BF), the Rainwater Foundation (BF), UTHSA Pepper Center Grant Pilot (BF), NIGMS R25 GM095480 (PR) and BrightFocus Foundation (WS).

CONFLICT OF INTEREST STATEMENT

Bess Frost serves on the Scientific Advisory Board of Transposon Therapeutics. Other co-authors have no competing interests. Author disclosures are available in the [supporting information](#).

CONSENT STATEMENT

Not applicable; *post mortem* human brain tissue was obtained from existing biorepositories.

DATA AVAILABILITY STATEMENT

All sequencing data is publicly available through the NCBI SRA database (PRJNA1083482).

CODE AVAILABILITY

Example code used to generate data is deposited in the Zenodo repository (10.5281/zenodo.15831306).

REFERENCES

1. Lambert J-C, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*. 2013;45(12):1452-1458.
2. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics*. 2009;41(10):1088-1093.
3. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics*. 2011;43(5):429-435.
4. Ebber MTW, Jensen TD, Jansen-West K, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol*. 2019;20:1-23.
5. Dwarshuis N, Kalra D, McDaniel J, et al. The GIAB genomic stratifications resource for human reference genomes. *Nat Commun*. 2024;15(1):9029.
6. Amemiya H, Kundaje A, Boyle A. The ENCODE blacklist: Identification of problematic regions of the genome. *Sci Rep*. 2019;9:9354.
7. Pan B, Ren L, Onuchic V, et al. Assessing reproducibility of inherited variants detected with short-read whole genome sequencing. *Genome Biol*. 2022;23:1-26.
8. Harris RA, Wang T, Coarfa C, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature Biotechnol*. 2010;28(10):1097-1105.
9. Gershman A, Sauria Michael EG, Guitart X, et al. Epigenetic patterns in a complete human genome. *Science*. 2022;376(6588):eabj5089.
10. Costantino I, Nicodemus J, Chun J. Genomic mosaicism formed by somatic variation in the aging and diseased brain. *Genes*. 2021;12(7):1071.
11. Huang CRL, Burns KH, Boeke JD. Active transposition in genomes. *Annual Review of Genetics*. 2012;46(1):651-675.
12. Ramirez P, Zuniga G, Sun W, et al. Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system. *Progress in Neurobiol*. 2022;208:102181.
13. Sun W, Samimi H, Gamez M, et al. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat Neurosci*. 2018;21(8):1038-1048.
14. Guo C, Jeong H-H, Hsieh Yi-C, et al. Tau activates transposable elements in Alzheimer's disease. *Cell Reports*. 2018;23(10):2874-2880.
15. Vallés-Saiz L, Ávila J, Hernández F. Lamivudine (3TC), a nucleoside reverse transcriptase inhibitor, prevents the neuropathological alterations present in mutant tau transgenic mice. *Int J Mol Sci*. 2023;24(13):11144.

16. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*. 2003;35(1):41-48.
17. Raiz J, Damert A, Chira S, et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res*. 2012;40(4):1666-1683.
18. Hancks DC, Goodier JL, Mandal PK, et al. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Human Mol Genetics*. 2011;20(17):3386-3400.
19. Beck CR, Collier P, Macfarlane C, et al. LINE-1 retrotransposition activity in human genomes. *Cell*. 2010;141(7):1159-1170.
20. Evrony GD, Lee E, Mehta BK, et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron*. 2015;85(1):49-59.
21. Erwin JA, Paquola ACM, Singer T, et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat Neurosci*. 2016;19(12):1583-1591.
22. Upton KR, Gerhardt DJ, Jesuadian JS, et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell*. 2015;161(2):228-239.
23. Elliott B, Richardson C, Jasin M. Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol Cell*. 2005;17(6):885-894.
24. Morales ME, White TB, Strevia VA, et al. The contribution of alu elements to mutagenic DNA double-strand break repair. *PLoS Genetics*. 2015;11(3):e1005016.
25. Song X, Beck CR, Du R, et al. Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Res*. 2018;28(8):1228-1242.
26. Parks MM, Lawrence CE, Raphael BJ. Detecting non-allelic homologous recombination from high-throughput sequencing data. *Genome Biol*. 2015;16:1-19.
27. Startek M, Szafranski P, Gambin T, et al. Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res*. 2015;43(4):2188-2198.
28. Pascarella G, Hon CC, Hashimoto K, et al. Recombination of repeat elements generates somatic complexity in human genomes. *Cell*. 2022;185(16):3025-3040.e6.
29. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259.
30. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100.
31. Pedersen BS, Quinlan AR. Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34(5):867-868.
32. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology*. 2011;29(1):24-26.
33. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21(3):487-493.
34. Smolka M, Paulin LF, Grochowski CM, et al. *Comprehensive structural variant detection: From mosaic to population-level*. Cold Spring Harbor Laboratory; 2022.
35. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-D745.
36. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410.
37. Afgan E, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*. 2022;50(W1):W345-W351.
38. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118.
39. Cavalcante RG, Sartor MA. annotatr: Genomic regions in context. *Bioinformatics*. 2017;33(15):2381-2383.
40. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
41. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. 2015;6(1).
42. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*. 2015;12(8):733-735.
43. Ewing AD, Smits N, Sanchez-Luque FJ, et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol Cell*. 2020;80(5):915-928.e5.
44. Akalin A, Kormaksson M, Li S, et al. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012;13(10):R87.
45. Su S, Gouil Q, Blewitt ME, et al. NanoMethViz: An R/Bioconductor package for visualizing long-read methylation data. *PLoS Comput Biol*. 2021;17(10):e1009524.
46. Feng H, Wu H. Differential methylation analysis for bisulfite sequencing using DSS. *Quant Biol*. 2019;7(4):327-334.
47. Gigante S, Gouil Q, Lucattini A, et al. Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res*. 2019;47(8):e46-e46.
48. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017.
49. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
50. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020;581(7809):444-451.
51. Niu Y, Teng X, Zhou H, et al. Characterizing mobile element insertions in 5675 genomes. *Nucleic Acids Res*. 2022;50(5):2493-2508.
52. Gel B, Díez-Villanueva A, Serra E, et al. regioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*. 2016;32(2):289-291.
53. Wickham H, ggplot2. Use R! 2016.
54. Sheffield NC, Bock C. LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32(4):587-589.
55. Gel B, Serra E. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 2017;33(19):3088-3090.
56. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
57. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. 2021;2(3):100141.
58. Patil I. Visualizations with statistical details: The 'ggstatsplot' approach. *J Open Source Softw*. 2021;6(61):3167.
59. Siudeja K, van den Beek M, Riddiford N, et al. Unraveling the features of somatic transposition in the *Drosophila* intestine. *The EMBO J*. 2021;40(9).
60. Flasch DA, Macia Á, Sánchez L, et al. Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell*. 2019;177(4):837-851.e28.
61. Feng Q, Moran JV, Kazazian HH, et al. Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell*. 1996;87(5):905-916.
62. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci*. 1997;94(5):1872-1877.
63. Sultana T, van Essen D, Siol O, et al. The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. *Mol Cell*. 2019;74(3):555-570.e7.
64. Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry*. 1998;37(51):18081-18093.
65. Suarez NA, Macia A, Muotri AR. LINE-1 retrotransposons in healthy and diseased human brain. *Dev Neurobiol*. 2018;78(5):434-455.

66. Muotri AR, Chu ViT, Marchetto MCN, et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005;435(7044):903-910.
67. Baillie JK, Barnett MW, Upton KR, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011;479(7374):534-537.
68. Stenz L. The L1-dependant and Pol III transcribed Alu retrotransposon, from its discovery to innate immunity. *Mol Biol Rep*. 2021;48(3):2775-2789.
69. Smits N, Rasmussen J, Bodea GO, et al. No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing. *Cell Reports*. 2021;36(7):109530.
70. Faulkner GJ, Billon V. L1 retrotransposition in the soma: A field jumping ahead. *Mobile DNA*. 2018;9(1):22.
71. Ostertag EM, Kazazian HH. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res*. 2001;11(12):2059-2065.
72. Hoang ML, Tan FJ, Lai DC, et al. Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. *PLoS Genetics*. 2010;6(12):e1001228.
73. Cingolani P, Platts A, Wang LeL, 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):80-92.
74. Chong SS, McCall AE, Cota J, et al. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics*. 1995;10(3):344-350.
75. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*. 2019;51(3):404-413.
76. Kawatani K, Holm ML, Starling SC, et al. ABCA7 deficiency causes neuronal dysregulation by altering mitochondrial lipid metabolism. *Mol Psychiatry*. 2023;29(3):809-819.
77. Fu Y, He Y, Phan K, et al. Sex-specific lipid dysregulation in the Abca7 knockout mouse brain. *Brain Commun*. 2022;4(3):fcac120.
78. Kim WS, Li H, Ruberu K, et al. Deletion of ABCA7 increases cerebral Amyloid- β accumulation in the J20 mouse model of Alzheimer's Disease. *The J Neurosci*. 2013;33(10):4387-4394.
79. Sakae N, Liu C-C, Shinohara M, et al. ABCA7 Deficiency Accelerates Amyloid- β Generation and Alzheimer's Neuronal Pathology. *The J Neurosci*. 2016;36(13):3848-3859.
80. Satoh K, Abe-Dohmae S, Yokoyama S, et al. ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing. *J Biol Chem*. 2015;290(40):24152-24165.
81. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*. 2007;39(4):457-466.
82. Rollins RA, Haghighi F, Edwards JR, et al. Large-scale structure of genomic methylation patterns. *Genome Res*. 2006;16(2):157-163.
83. Jacobs FMJ, Greenberg D, Nguyen N, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*. 2014;516(7530):242-245.
84. Imbeault M, Helleboid P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*. 2017;543(7646):550-554.
85. Zhang L, Silva TC, Young JI, et al. Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nat Commun*. 2020;11(1):6114.
86. Hernández HG, Mahecha MF, Mejía A, et al. Global Long Interspersed Nuclear Element 1 DNA Methylation in a Colombian Sample of Patients With Late-Onset Alzheimer's Disease. *American J Alzheimers Dis & Other Dementias®*. 2014;29(1):50-53.
87. Bollati V, Galimberti D, Pergoli L, et al. DNA methylation in repetitive elements and Alzheimer disease. *Brain, Behavior, and Immunity*. 2011;25(6):1078-1083.
88. Sanchez-Luque FJ, Kempen M-JHC, Gerdes P, et al. LINE-1 evasion of epigenetic repression in humans. *Mol Cell*. 2019;75(3):590-604. e12.
89. Gonzalez-Cao M, Iduma P, Karachaliou N, Santaripa M, Blanco J, Rosell R. Human endogenous retroviruses and cancer. *Cancer Biol Med*. 2016;13(4):483.
90. Gao Y, Yu X-F, Chen T. Human endogenous retroviruses in cancer: Expression, regulation and function. *Oncol Lett*. 2020;21(2):121.
91. Agoni L, Lenz J, Guha C. Variant splicing and influence of ionizing radiation on Human Endogenous Retrovirus K (HERV-K) transcripts in cancer cell Lines. *PLoS ONE*. 2013;8(10):e76472.
92. Smith RG, Pishva E, Shireby G, et al. A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nat Commun*. 2021;12(1).
93. Shireby G, Dempster EL, Policicchio S, et al. DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. *Nat Commun*. 2022;13(1):5620.
94. Evrony GD, Cai X, Lee E, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell*. 2012;151(3):483-496.
95. Konkel MK, Walker JA, Hotard AB, et al. Sequence analysis and characterization of active human Alu subfamilies based on the 1000 Genomes Pilot Project. *Genome Biol Evol*. 2015:evv167.
96. Birchler JA, Presting GG. Retrotransposon insertion targeting: A mechanism for homogenization of centromere sequences on nonhomologous chromosomes. *Genes & Development*. 2012;26(7):638-640.
97. Sultana T, Zamborlini A, Cristofari G, et al. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet*. 2017;18(5):292-308.
98. Klein SJ, O'Neill RJ. Transposable elements: Genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res*. 2018;26:5-23.
99. Dawson T, Rentia U, Sanford J, et al. Locus specific endogenous retroviral expression associated with Alzheimer's disease. *Front Aging Neurosci*. 2023;15:1186470.
100. Macciardi F, Giulia Bacalini M, Miramontes R, et al. A retrotransposon storm marks clinical phenocopy to late-onset Alzheimer's disease. *GeroScience*. 2022;44(3):1525-1550.
101. Scopa C, Barnada SM, Cicardi ME, et al. JUN upregulation drives aberrant transposable element mobilization, associated innate immune response, and impaired neurogenesis in Alzheimer's disease. *Nat Commun*. 2023;14(1):8021.
102. Wahl D, Smith ME, McEntee CM, et al. The reverse transcriptase inhibitor 3TC protects against age-related cognitive dysfunction. *Aging Cell*. 2023;22(5):e13798.
103. Sun Z, Kwon JS, Ren Y, et al. Endogenous recapitulation of Alzheimer's disease neuropathology through human 3D direct neuronal reprogramming. *bioRxiv*. 2023:2023.05.24.542155.
104. Sullivan AC, Zuniga G, Ramirez P, et al. A Phase IIa clinical trial to evaluate the effects of anti-retroviral therapy in Alzheimer's disease (ART-AD). *npj Dementia*. 2025;1(1):2.
105. Magagnoli J, Ambati M, Cummings TH, et al. Association of nucleoside reverse transcriptase inhibitor use with reduced risk of Alzheimer's disease risk. *Alzheimer's & Dementia*. 2025;21(5):e70180.
106. Yang Fu, Su W, Chung OW, et al. Retrotransposons hijack alt-EJ for DNA replication and eccDNA biogenesis. *Nature*. 2023;620(7972):218-225.
107. Pellegrini C, Pirazzini C, Sala C, et al. A Meta-Analysis of brain DNA methylation across sex, age, and Alzheimer's Disease points for accelerated epigenetic aging in neurodegeneration. *Front Aging Neurosci*. 2021;13:639428.

108. Li QS, Sun Y, Wang T. Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. *Clin Epigenetics*. 2020;12(1):149.
109. Mastroeni D, Grover A, Delvaux E, et al. Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. *Neurobiol Aging*. 2010;31(12):2025-2037.
110. Chouliaras L, Mastroeni D, Delvaux E, et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging*. 2013;34(9):2091-2099.
111. Watson CT, Roussos P, Garg P, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med*. 2016;8(1):5.
112. Bakulski KM, Dolinoy DC, Sartor MA, et al. Genome-Wide DNA methylation differences between Late-Onset Alzheimer's Disease and cognitively normal controls in human frontal cortex. *J Alzheimers Dis*. 2012;29(3):571-588.
113. Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*. 2012;151(7):1417-1430.
114. Frost B, Hemberg M, Lewis J, et al. Tau promotes neurodegeneration through global chromatin relaxation. *Nat Neurosci*. 2014;17(3):357-366.
115. Klein H-U, McCabe C, Gjoneska E, et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat Neurosci*. 2019;22(1):37-46.
116. Mertens J, Herdy JR, Traxler L, et al. Age-dependent instability of mature neuronal fate in induced neurons from Alzheimer's patients. *Cell Stem Cell*. 2021;28(9):1533-1548. e6.
117. Shih H-T, Chen W-Yi, Wang H-Y, et al. DNMT3b protects centromere integrity by restricting R-loop-mediated DNA damage. *Cell Death & Disease*. 2022;13(6):546.
118. Narayan A, Ji W, Zhang X-Y, et al. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer*. 1998;77(6):833-838.
119. Lengauer C, Kinzler KW, Vogelstein B. DNA methylation and genetic instability in colorectal cancer cells. *Proc Natl Acad Sci*. 1997;94(6):2545-2550.
120. Mosch B, Morawski M, Mittag A, et al. Aneuploidy and DNA replication in the normal human brain and Alzheimer's disease. *J Neurosci*. 2007;27(26):6859-6867.
121. Iourov IY, Vorsanova SG, Liehr T, et al. Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: Differential expression and pathological meaning. *Neurobiol Dis*. 2009;34(2):212-220.
122. Arendt T, Brückner MK, Mosch B, et al. Selective cell death of hyperploid neurons in Alzheimer's disease. *Am J Pathol*. 2010;177(1):15-20.
123. Yurov YB, Vorsanova SG, Liehr T, et al. X chromosome aneuploidy in the Alzheimer's disease brain. *Mol Cytogenet*. 2014;7(1):20.
124. Miyata K, Imai Y, Hori S, et al. Pericentromeric noncoding RNA changes DNA binding of CTCF and inflammatory gene expression in senescence and cancer. *Proc Natl Acad Sci*. 2021;118(35):e2025647118.
125. Warmerdam DO, van den Berg J, Medema RH. Breaks in the 45S rDNA lead to recombination-mediated loss of repeats. *Cell Reports*. 2016;14(11):2519-2527.
126. Kim J-H, Dilthey AT, Nagaraja R, et al. Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. *Nucleic Acids Res*. 2018;46(13):6712-6725.
127. Kupriyanova NS, Netchvolodov KK, Sadova AA, et al. Non-canonical ribosomal DNA segments in the human genome, and nucleoli functioning. *Gene*. 2015;572(2):237-242.
128. Sinclair DA, Guarente L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell*. 1997;91(7):1033-1042.
129. Cohen S, Yacobi K, Segal D. Extrachromosomal circular DNA of tandemly repeated genomic sequences in *Drosophila*. *Genome Res*. 2003;13(6a):1133-1145.
130. Pietrzak M, Rempala G, Nelson PT, et al. Epigenetic silencing of nucleolar rRNA genes in Alzheimer's disease. *PloS one*. 2011;6(7):e22585.
131. Cohen S, Agmon N, Sobol O, et al. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA*. 2010;1:1-11.
132. Zuniga G, Katsumura S, De Mange J, et al. Pathogenic tau induces an adaptive elevation in mRNA translation rate at early stages of disease. *Aging Cell*. 2024;23(10):e14245.
133. Meier S, Bell M, Lyons DN, et al. Pathological tau promotes neuronal damage by impairing ribosomal function and decreasing protein synthesis. *J Neurosci*. 2016;36(3):1001-1007. /bib>
134. Evans HT, Taylor D, Kneynsberg A, et al. Altered ribosomal function and protein synthesis caused by tau. *Acta Neuropathol Commun*. 2021;9(1):110.
135. Koren SA, Hamm MJ, Meier SE, et al. Tau drives translational selectivity by interacting with ribosomal proteins. *Acta Neuropathol*. 2019;137:571-583.
136. Wang H, Wang Li-S, Schellenberg G, et al. The role of structural variations in Alzheimer's disease and other neurodegenerative diseases. *Front Aging Neurosci*. 2023;14:1073905.
137. Dib S, Pahnke J, Gosselet F. Role of ABCA7 in Human Health and in Alzheimer's Disease. *Int J Mol Sci*. 2021;22(9):4603.
138. De Roeck A, Duchateau L, Van Dongen J, et al. An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease. *Acta Neuropathol*. 2018;135:827-837.
139. Mastroianni FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med*. 2023;15(1).
140. Lövkvist C, Dodd IB, Sneppen K, et al. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res*. 2016;44(11):5123-5132.
141. Andrade-Moraes CH, Oliveira-Pinto AV, Castro-Fonseca E, et al. Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles. *Brain*. 2013;136(12):3738-3752.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ramirez P, Sun W, Dehkordi SK, et al. Long-read sequencing reveals genomic and epigenomic variation in the dark genome of human Alzheimer's disease. *Alzheimer's Dement*. 2025;21:e70852. <https://doi.org/10.1002/alz.70852>