AWARD NUMBER: W81XWH-20-1-0028

TITLE: Systematic Analysis of Genetic Mosaicism in FTD/ALS Brains

PRINCIPAL INVESTIGATOR: Dr. Zinan Zhou, PhD

CONTRACTING ORGANIZATION: Children's Hospital, Boston

REPORT DATE: MAY 2023

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2023	FINAL	1FEB2020 - 31JAN2023

4. TITLE AND SUBTITLE

Systematic Analysis of Genetic Mosaicism in FTD/ALS Brains

5a. CONTRACT NUMBER
W81XWH-20-1-0028

5b. GRANT NUMBER
W81XWH-20-1-0028

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Zinan Zhou, PhD

E-Mail: zinan.zhou@childrens.harvard.edu/

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Children's Hospital, Boston
300 Longwood Ave, Boston, MA 02115

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) are two closely related neurodegenerative diseases with TAR DNA-binding protein 43 (TDP-43) inclusions as a common pathological hallmark. They are characterized as age-related neurodegeneration, yet the mechanisms by which age and genetic risk interact, as well as the ultimate cause of neuronal loss, remains murky. Clinical features of FTD/ALS, including focal onset, stereotyped patterns of spread, and increased risk due to smoking, have suggested to some a role of somatic mutations as causative for some unexplained cases of FTD/ALS. Our major goal is to reveal the potential roles of genetic mosaicism in the etiology and progression of FTD/ALS. Our study identifies clonal and non-clonal somatic mutations in FTD/ALS brains using ultra-deep targeted sequencing and single-cell whole genome sequencing. In the last funding period, we had finished the analyses of our MIPs panel sequencing data and showed enrichments of somatic mutations in FTD/ALS brains, as well as the interesting topological and cell-type distribution of the identified somatic mutations. We had also analyzed the additional scWGS data. Although the results had not been published, we anticipate three publications in the near future.

15. SUBJECT TERMS

amyotrophic lateral sclerosis, double-strand break, frontotemporal dementia, Indel, Molecular Inversion Probes, retrotransposon insertion, single-cell whole genome sequencing, somatic mutation, single-nucleotide variant, TDP-43

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	15	USAMRDC

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

1. Introduction

Frontotemporal dementia (FTD), a neurodegenerative disorder characterized by behavioral, language, and memory dysfunction, is increasingly recognized to have clinical, genetic and pathological overlap with amyotrophic lateral sclerosis (ALS), a disease in which premature loss of upper and lower motor neurons leads to fatal paralysis, since the seminal discovery of TDP-43 inclusions as a common pathological hallmark in FTD and ALS. FTD and ALS are characterized as age-related neurodegeneration, yet the mechanisms by which age and genetic risk interact, as well as the ultimate cause of neuronal loss, remains murky. Clinical features of FTD/ALS, including focal onset, stereotyped patterns of spread, and increased risk due to smoking, have suggested to some a role of somatic mutation in FTD/ALS. Although many causative genes involved in various pathways have been identified to predispose to both familial and sporadic FTD/ALS over the past two decades, the possibility that clonal mutations of these genes—i.e., present initially in some cells but not all cells—may contribute to some cases of sporadic FTD/ALS has not been evaluated.

Whereas clonal mosaicism creates risk for some diseases, a recent study suggested that nonclonal mutations accumulate in postmitotic neurons with age, and may contribute to age-related degenerative conditions. For example, Cockayne Syndrome (CS) and Xeroderma Pigmentosum (XP), two human progeroid diseases characterized by accelerated aging and premature neuronal loss, exhibit faster accumulation of somatic single-nucleotide variant (sSNV), as well as distinct mutational signatures compared to age-matched controls through single-cell whole genome sequencing (scWGS) of post-mitotic neurons. These results suggest that accelerated accumulation of DNA mutations is a potential mechanism of other forms of neuronal degeneration and emerging evidence has implicated DNA damage in FTD/ALS.

TDP-43 is an aggregation-prone RNA- and DNA-binding protein identified as the major component of the cytoplasmic ubiquitinated inclusions formed in neurons and glial cells of almost all ALS and about 40% of FTD patients. In TDP-43 proteinopathies, the protein is detectable in the nuclei of unaffected neurons but partially cleared from nuclei in neurons containing cytoplasmic aggregations supporting that pathogenesis is driven, at least in part, by a loss of TDP-43 nuclear function. TDP-43 interacts with Fused in Sarcoma (FUS), a DNA/RNA binding protein shown to participate in homologous recombination (HR) and non-homologous end joining (NHEJ) mediated double-strand break (DSB) repair. Mutations of both *TDP-43* and *FUS* have been identified in FTD/ALS patients. TDP-43 and FUS both localize at sites of transcription-associated DNA damage and depletion of either TDP-43 or FUS leads to increased DNA breakage. TDP-43 has also been shown to be a critical component of the NHEJ-mediated DNA double-strand break (DSB) repair pathway. Hence TDP-43 would be normally suppressing transcription-associated DNA damage and DSB, both of which could otherwise increase somatic mutations in post-mitotic neurons and lead to cell death. In addition, *C9orf72* repeat expansion is the most common genetic cause of familial FTD/ALS, accounting for approximately 30% of familial cases. Recent studies have shown that *C9orf72* repeat expansion causes DNA breakage and leads to neuronal cell death. Taken together, these data suggest our hypothesis that clonal and non-clonal somatic mutations could be associated with FTD/ALS, which might arise from both development and postnatal DNA damage that go unrepaired.

To discover the presence of rare clonal mutations in FTD/ALS brains directly, the first goal of our project will be devoted to high-throughput screening of clonal mutations in a panel of 90 neurodegeneration related genes in the prefrontal cortex, motor cortex and cerebellum of a cohort of sporadic FTD/ALS and age-matched normal cases. Clonal mutations of previously identified FTD/ALS genes will be identified and evaluated based on their predicted pathogenicity. Potential crossover between FTD/ALS and other neurogenerative diseases can also be evaluated because our targeted gene sequencing panel also contains genes that are related to Alzheimer's Disease, Parkinson's Disease and other rare types of dementia. Results of our project will reveal the frequency of pathogenic clonal mutation in FTD/ALS and may change our understanding of the genetics of the disease.

Our second goal is to compare the rate and pattern of non-clonal somatic mutations in neurons of FTD/ALS brains to neurons of normal brains. Various types of mutation, including SNV, Indel, retrotransposon insertion and double-strand break, will be evaluated through single-cell whole-genome and targeted sequencing. These approaches allow us to define ultra-rare mutations at single-cell level. An increased burden of somatic mutation will suggest that DNA damage exceeds the repair capacity. The sources of exogenous and exogenous DNA damage can be identified by analyzing the specific patterns of somatic mutations in FTD/ALS brains, which may potentially reveal the environmental risk factors of FTD/ALS.

2. Keywords

Amyotrophic Lateral Sclerosis
double-strand break
Frontotemporal Dementia
Indel
Molecular Inversion Probes
retrotransposon insertion
single-cell whole genome sequencing
somatic mutation
single-nucleotide variant
TDP-43

3. Accomplishments

What were the major goals of the project?

Specific Aim 1: Ultra-deep targeted sequencing of neurodegenerative genes in sporadic FTD/ALS brains using Molecular Inversion Probes (MIPs)

Major Task 1

Milestone at 3 months: HRPO Approval
We submitted HRPO forms for review in September 2020 and got approved in February 2021. This step was dramatically delayed due to the COVID-19 pandemic and also caused the delay of subsequent experiments.
Milestone at 6 months: Extracted gDNA from all brain samples
We finished gDNA extraction of all samples in February 2021.
Milestone at 10 months: Raw sequencing data
All samples were sequenced in August 2021.
Milestone at 14.5 months: Validated pathogenic germline and somatic variants
We have validated a set of somatic variants that were predicted to be pathogenic in December 2021. Additionally, during the extension period, we have examined the regional and cell-type distributions of the validated pathogenic variants.

Specific Aim 2: Genome-wide burden of sSNV and retrotransposon insertion in single neurons of FTD/ALS brains

Major Task 2: Determine the burden of sSNV through scWGS

Milestone at 3 month: Raw data of low-coverage sequencing
A new whole genome amplification method, primary template-directed amplification (PTA), became commercially available in 2020 and showed a much better quality of whole-genome amplification of single cells in our hands. Using this new method, we no longer needed to do low-coverage sequencing for QC analyses. All PTA amplified cells were directly used for high-coverage sequencing before data analyses.
Milestone at 6 month: Raw data of high-coverage sequencing
High-coverage sequencing were generated in September 2021.
Milestone at 11 month: Complete comparison of somatic SNV between FTD/ALS and normal brains
Somatic SNVs and Indels were initially called and compared in January 2022. In addition to the original planned scWGS, we sequenced more normal neurons and neurons with TDP-43 pathology from the PFC of three more FTD brains. The additional scWGS data has been analyzed. We are performing further analyses to better understand potential mechanisms underlying the observed changes.

Major Task 3: Determine the rate of retrotransposon insertion through L1 targeted sequencing

Milestone at 6 months: Raw data of L1 targeted sequencing
Using the newly developed PTA-HATseq method, we finished the sequencing for FTD/ALS neurons in December 2021.
Milestone at 12 months: Complete comparison of somatic L1 insertion between FTD/ALS and normal brains
We have made the initial calls of L1 insertions. Unexpectedly, the initial analysis showed that normal neurons had higher burden of somatic L1 insertions than neurons with TDP-43 pathology. To better interpretate the result, we performed PTA-HATseq for neurons from 6 age-matched normal brains. This additional data would set the baseline of somatic L1 insertions in normal neurons. We are now calling and validating somatic L1 insertions in the additional neurons

Specific Aim 3: Genome-wide burden of DSB in neurons of FTD/ALS brains

Major Task 4: Determine the burden of somatic Indels through scWGS

Milestone at 4 months: Complete comparison of somatic Indel between FTD/ALS and normal brains
The initial Indel calls were made in December 2021. Indel calling for additional cells were done in April 2023. The unexpected finding of an FTD/ALS-associated mutational signatures of somatic Indels was intriguing and led us to perform additional computational analyses to understand the underlying mechanisms of the accumulation of somatic Indels in FTD/ALS neurons.

Major Task 5: Identify DSB using BLISS

Milestone at 6 months: Raw data of the BLISS sequencing
Due to the technical challenges of BLISS using postmortem tissues, we couldn't generate clear results of double-strand breaks in postmortem brains. We don’t think this technical problem can be solved within the time period of this grant.

What was accomplished under these goals?

Specific Aim 1: Ultra-deep targeted sequencing of neurodegenerative genes in sporadic FTD/ALS brains using Molecular Inversion Probes (MIPs)

We have collected 1,817 human postmortem brain tissues of 548 individuals from brain banks and gotten the HPRO forms approved. gDNA from all these samples were extracted. We have designed a MIPs panel targeting the exons and exon-intron junctions of 88 neurodegeneration related genes, which includes 34 FTD/ALS genes, 10 AD genes, 28 PD genes as well as genes associated with other rare dementia. We performed MIPs panel sequencing for all the 1,817 samples and achieved ~1,800X average sequencing depth. We did not observe much variations of sequencing depth across batches, disease conditions and tissue regions. We first identified pathogenic germline mutations in our cohort using GATK followed by multiple variant filters. The functional impacts and predicted pathogenicity of identified germline mutations were annotated using ANNOVAR and multiple clinical databases. We also genotyped the C9ORF72 repeat expansion in all FTD/ALS cases by a repeat-primed PCR assay. Overall, 25.7% of ALS and 32.7% of FTD cases had germline mutations in ALS and FTD genes that have been previously reported as pathogenic, or were predicted to be damaging by our pipeline (Fig.1)

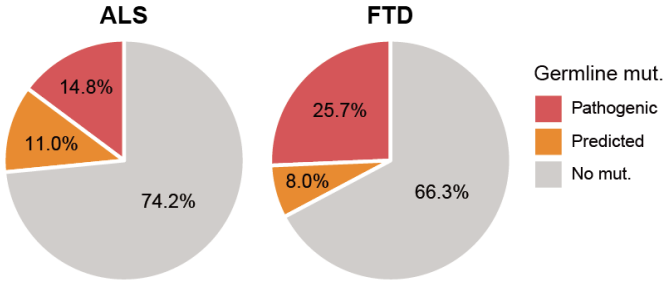

Figure 1. Proportion of FTD/ALS cases with germline variants in FTD/ALS genes. Pathogenic mutations were previously reported and predicted mutations were predicted by our customized annotation pipeline.

To check the performance of our MIPs panel in calling somatic mutations, we first performed a spike-in experiment using commercially available human gDNA from the Genome in a Bottle Consortium (GIAB). The GIAB cases have been well characterized and the WGS data of them are available, which has been widely used for benchmarking variant calling. The spike-in test sequencing was done by mixing one normal individual's gDNA (GIAB12878) into another normal individual's gDNA (GIAB24695) at five different variant allele frequencies (VAFs), 10%, 5%, 2.5 %, 1% and 0.5%. The observed VAFs of somatic SNVs were very well in line with the target VAFs at all five VAF ratios (Fig. 2a). To identify somatic mutations in our MIP sequencing data, we developed a custom pipeline integrating RePlow, Mutect2, and Pisces for calling somatic SNVs and indels. We selected somatic mutations identified by two of the three callers (double-called mutations) followed by multi-step variant filters to remove false positive candidates. High sensitivity and precision were achieved for the remaining RePlow-based double-called mutations (RePlow-Mutect2 and RePlow-Pisces) while maintaining a low false positive rate across the low VAFs compared to the somatic mutations called by each caller (Fig. 2b). The MIP sequencing and our custom pipeline together allowed us to confidently identify somatic mutations with a low false positive rate at 0.5% VAF.

Using this customized pipeline, we identified 167 somatic SNVs and Indels from our MIPs sequencing data. We selected 41 somatic candidates for validation and 87.8% of the them were validated by amplicon sequencing. The VAFs of validated candidates in amplicon sequencing showed a strong correlation with their original VAFs in the MIPs sequencing data (Fig. 3a). Out of the 167 somatic candidates, 95 unique somatic mutations were identified from 216 ALS and 76 FTD cases without carrying any known or predicted pathogenic germline mutations (referred to as germline-free cases), and 144 normal cases. Most of the 95 unique somatic mutations were focal, identified only in one tissue region of the same individual (Fig. 3b), and at very low VAFs, suggesting that they likely arose during late stages of development or in small fractions of glial cells in the postnatal life. These somatic mutations also showed a clear separation between the disease and control groups (Fig. 3c). Interestingly, only one protein-altering (missense/ nonsense/ frameshift) somatic mutation was observed among all of the normal samples, while many were observed in both disease groups, suggesting an enrichment of potentially damaging somatic mutations in the targeted neurodegenerative genes in disease cases. Further enrichment analysis indeed showed that protein-altering somatic mutations in neurodegenerative genes in both germline-free FTD and ALS samples (Fig. 4a). We then assessed the somatic mutations in genes specifically related to FTD and ALS and found that somatic mutations in each set of disease-related genes were enriched in the corresponding diseases (Fig. 4b). On the other hand, no protein-altering mutation was observed in any of the FTD/ALS genes in control samples (Fig. 3c). The exclusive and diagnosis-specific enrichments of functional somatic mutations suggest their potential roles in contributing to the pathogenesis of somatic FTD and ALS in a modest but significant fraction of cases. We also found interesting topographic patterns of somatic mutations. Exonic somatic mutations were exclusively enriched in the disease-affected tissue regions of ALS and FTD germline-free cases — the primary motor cortex (PMC) for ALS and prefrontal cortex (PFC) for FTD (Fig. 4c). Enrichments of protein-altering somatic mutations in the PMC of ALS cases and PFC of FTD cases were even more significant. These findings further support the pathogenic roles of the identified somatic mutations.

Our pathogenicity prediction resulted in 8 predicted pathogenic somatic SNVs in previously known FTD/Tau-proteinopathy and ALS genes, which account for 2.6% and 3.2% of germline-free FTD and ALS cases respectively.

Among the 8 predicted pathogenic somatic SNVs, all pathogenic mutations in ALS cases were observed in the PMC or spinal cord, the most severely affected regions in ALS. In addition, a predicted pathogenic somatic SNV in APP

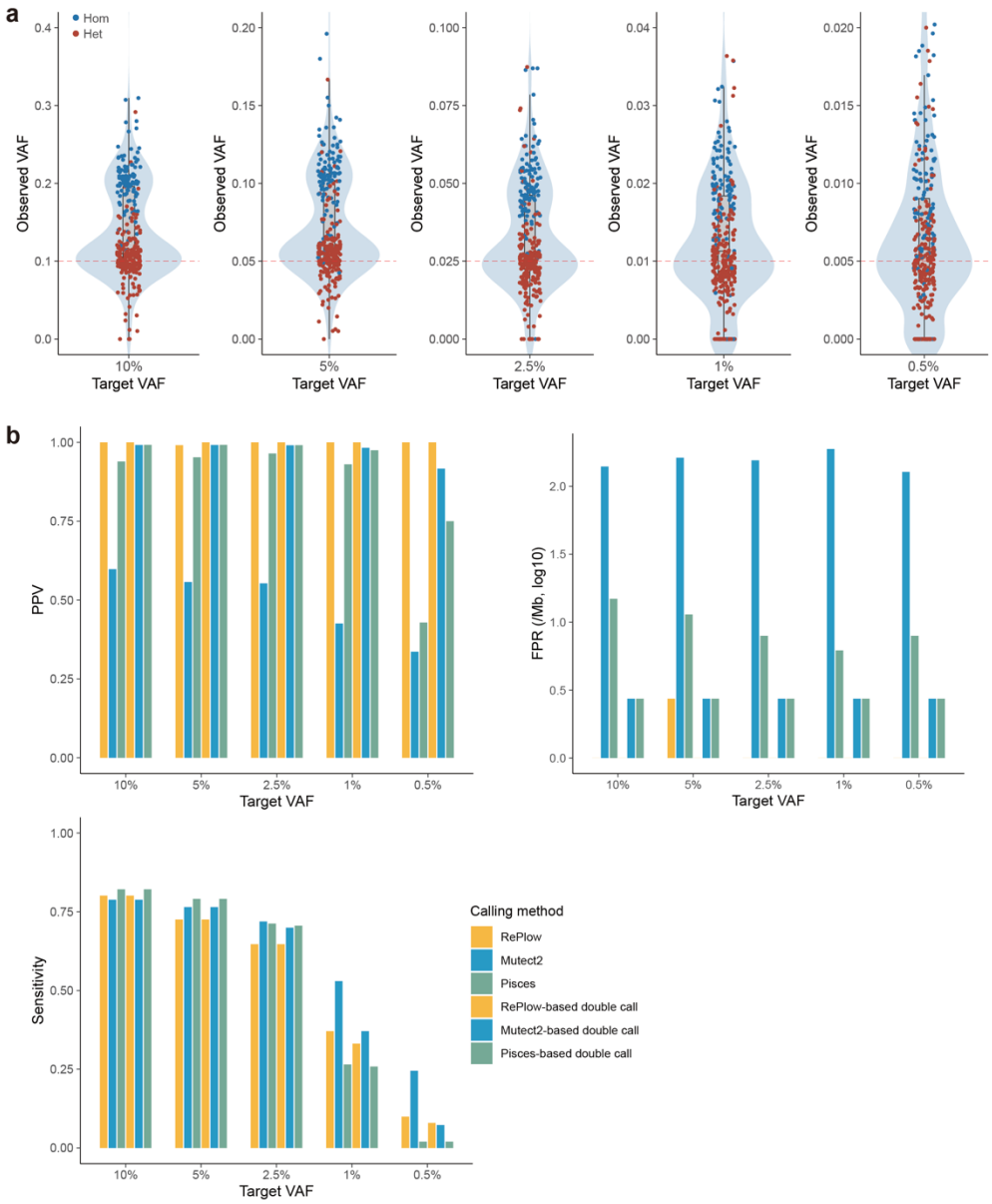

Figure 2. Evaluation of the variant calling method for somatic mutation using spike-in experiments. (a) Variant allele fraction (VAF) distribution of 165 benchmark variants from the MIP-sequenced spike-in data. Each violin plot demonstrates the VAF distribution of benchmark variants across spike-in mixtures that target five different VAF levels (10%, 5%, 2.5%, 1%, and 0.5%). Red dotted lines represent the targeted VAF of a given mixture. (b) Benchmarking of three different callers and their combinations using spike-in data. PPV, Positive predictive value; FPR, false positive rate.

(p.R328Q) was identified in the PMC of a sporadic case that exhibited both ALS and FTD. We selected four pathogenic somatic SNVs in FTD/ALS genes, i.e., TIA1 (p.H54R), MATR3 (p.K594I), ALS2 (p.T787R), and TARDBP (p.L248F), and also a pathogenic APP somatic SNV (p.R328Q) to further delineate their regional and cell-type distributions since we were able to get tissues of additional brain regions of the same brains. We first performed amplicon sequencing for these five pathogenic somatic SNVs across multiple brain and spinal cord regions of the individuals in which the somatic SNVs were originally detected. Interestingly, three of the five somatic SNVs (MATR3 (p.K594I), APP (p.R328Q), TARDBP (p.L248F)) were mainly restricted to the PMC (Fig. 5a). The other two somatic SNVs (TIA1 (p.H54R) and ALS2 (p.T787R)) had the highest VAFs in the spinal cord (2.16% for TIA1 (p. H54R) and 0.97% for ALS2 (p.T787R)), where they were originally identified, and were also present in other brain regions at very low VAFs (VAFs ranging from 0.15% to 1.05% for TIA1 (p.H54R) and from 0.16% to 0.59% for ALS2 (p.T787R)) (Fig. 5a). The low levels and restricted distribution of these somatic SNVs suggest that they probably arose during the late stage of development and were excluded in non-CNS tissues.

We then determined the cell-type specificity of these five somatic SNVs by performing amplicon sequencing for neuronal (NeuN+), glial (NeuN-), diploid, polyploid, and hypodiploid nuclei (which are interpreted as apoptotic) isolated by fluorescence-activated nuclei sorting (FANS) from the tissue regions in which they were originally identified. Interestingly,

the TIA1 (p.H54R), MATR3 (p.K594I), and ALS2 (p.T787R) mutations were enriched in hypodiploid nuclei, which likely indicated DNA fragmentation and cell death (Fig. 5b). Enrichments of these somatic mutations in dying cells further

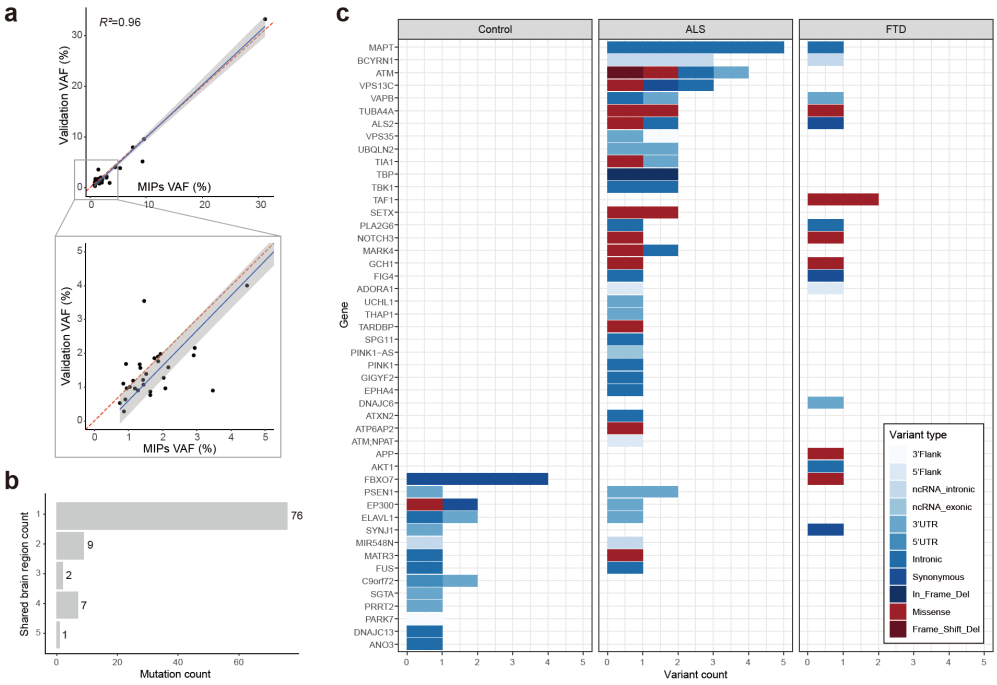

Figure 3. Somatic variants in MIP sequencing data tend to be focal and protein-altering somatic variants are almost exclusively present in disease cases. and distinct mutational distribution in disease and control tissues. (a) The observed VAFs of somatic variants in amplicon sequencing validation were consistent with the VAFs in original MIP sequencing. Forty somatic variants were validated and included in the plot. (b) Total somatic variant counts classified by the number of brain regions in which a given variant was identified. (c) Distribution of somatic variants in all neurodegenerative genes. Color codes indicate variant types. Note that somatic variants identified in controls are unlikely to be function-altering, with just one missense mutation (red) and the remaining being synonymous or noncoding substitutions.

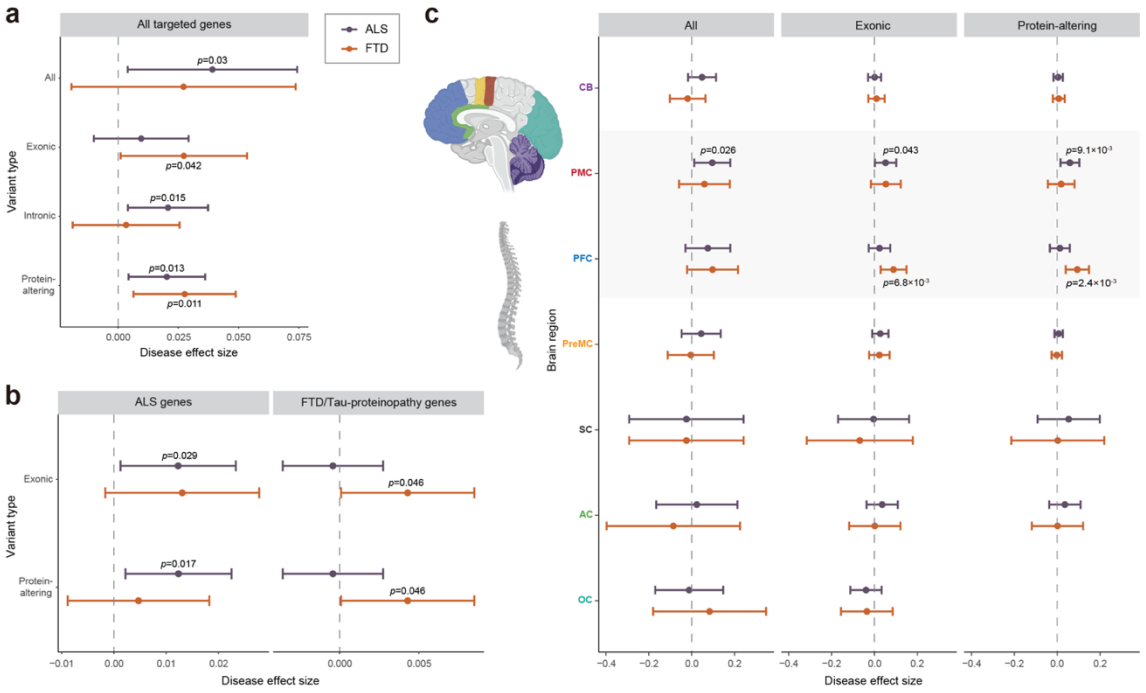

Figure 4. Somatic variants are enriched in ALS and FTD cases and disease related tissue regions. (a) Enrichment of somatic variants in different genomic regions of germline-free ALS and FTD cases compared to normal controls. (b) Enrichment of exonic and protein-altering somatic variants in two different groups of disease-related genes (ALS genes and FTD/Tau-proteinopathy genes) compared to normal controls. (c) Enrichment of somatic variants in different brain regions of germline-free ALS and FTD cases compared to normal controls. Significance of enrichment and 95% confidence interval was estimated while controlling for potential confounding factors including average read-depth, sequencing batch, sampled individual using a linear mixed model.

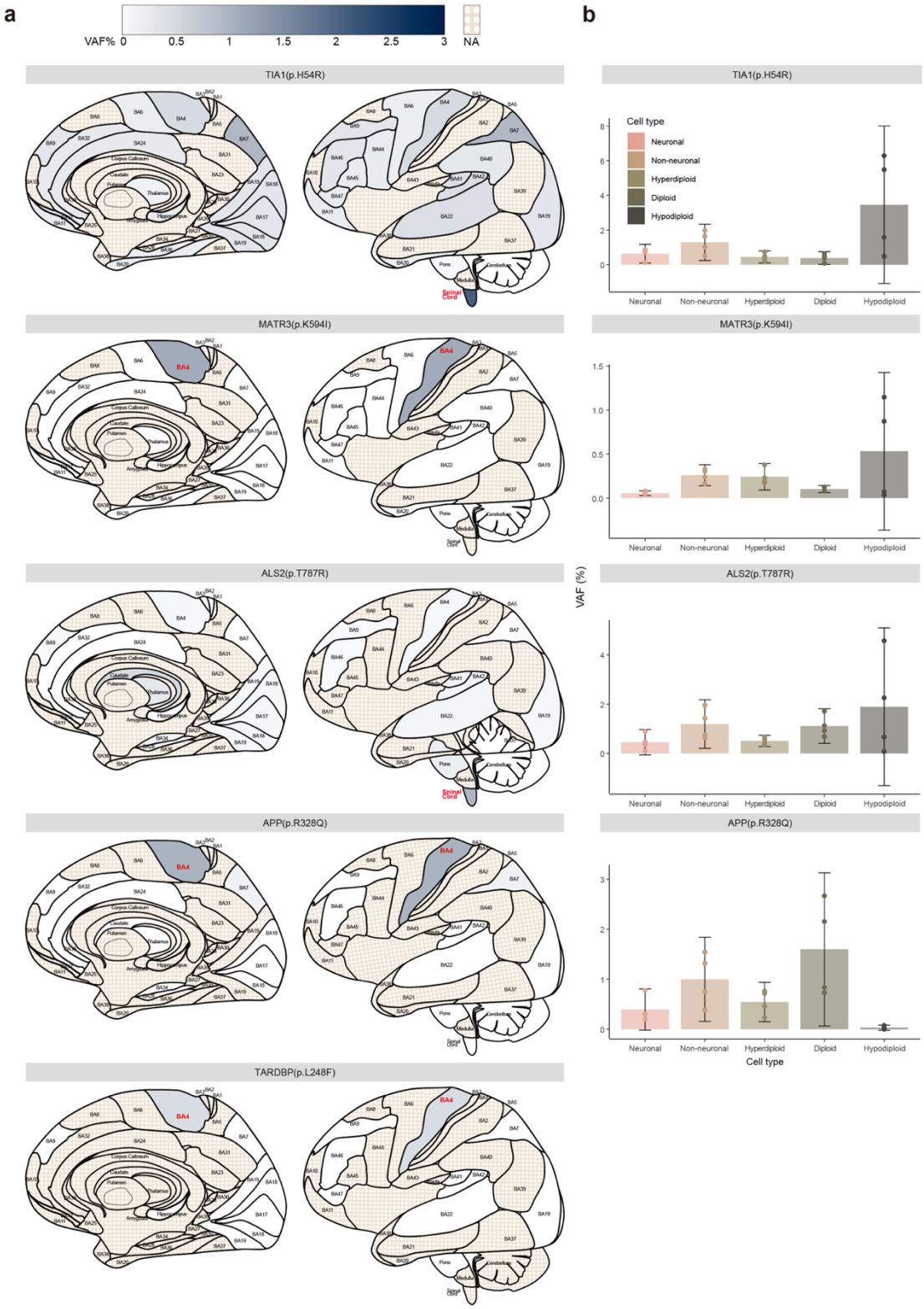

Figure 5. Pathogenic somatic mutations have restricted regional distributions and are enriched in hypodiploid cells. (a) Regional distribution of VAFs of somatic variants in individual brains and spinal cords. Brain cortex is annotated by Brodmann areas. The color spectrum indicates the VAFs of somatic variants in amplicon sequencing. Dots indicate unavailable regions and white indicates regions without the somatic variants. Red highlight indicates the region of initial detection by MIP sequencing. (b) VAFs of somatic variants in FANS sorted cell types. Five hundred neuronal (NeuN+), non-neuronal (NeuN-), diploid (DAPI), hyperdiploid (High DAPI) and hypodiploid (Low DAPI) cells were each sorted for amplicon sequencing with four replicates. Error bars, 95% CI.

support the pathogenicity of the mutations and suggest that these mutations act cell-autonomously to induce cell death. On the other hand, the APP (p.R328Q) mutation was depleted from hypodiploid cells but relatively enriched in non-neuronal cells (Fig. 5b), suggesting a non-cell-autonomous mechanism of the mutation in causing the disease. Surprisingly, the TARDBP (p.L248F) mutation was not detected in any isolated cell fractions, indicating that the mutation

was restricted to the area that it was initially discovered but not extended to even the nearby area. This was further confirmed by amplicon sequencing of a second tissue sample from the same brain region, where it was absent again.

Overall, our findings of this specific aim revealed the important roles of somatic mutations in the pathogenesis of sporadic FTD and ALS.

Specific Aim 2: Genome-wide burden of sSNV and retrotransposon insertion in single neurons of FTD/ALS brains

PTA, a new whole genome amplification (WGA) method, became commercially available in 2020. We compared scWGS data generated by PTA and MDA. PTA data had higher sensitivity, more even genome amplification, and fewer artifacts than other commonly used MDA data (Fig. 6). We decided to generate scWGS data for this project using PTA instead of MDA as we proposed originally in this grant. PTA not only improves the quality of scWGS data, but also saves time in doing quality control experiments.

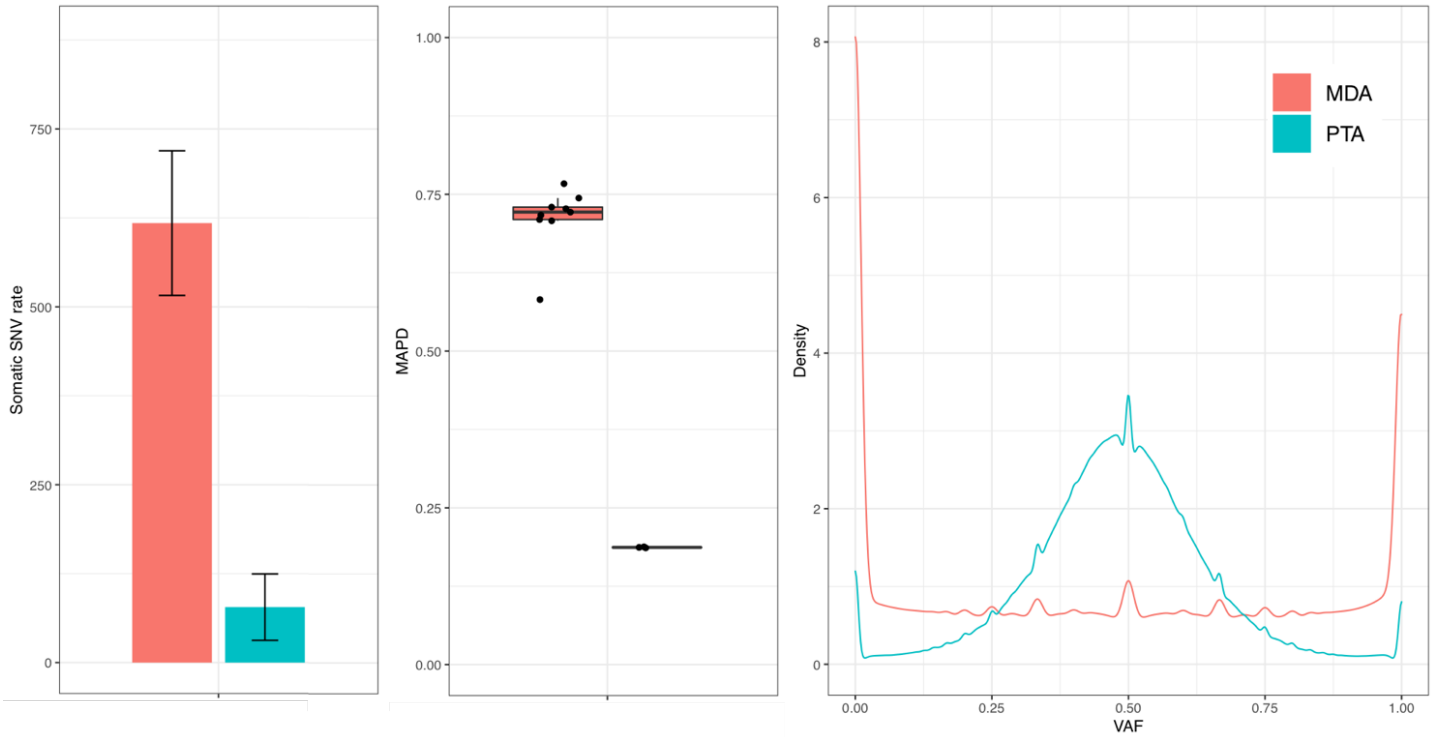

Figure 6. Comparison between PTA and MDA. 9 neurons amplified by MDA and 3 neurons amplified by PTA from the postmortem brain of a normal individual at the age of five months were compared. PTA-amplified neurons showed lower somatic SNV rate (left), lower variability in sequencing depth/better evenness indicated by lower median absolute pairwise deviation (MAPD) score (middle), and lower allelic bias indicated by better allele balance for germline heterozygous SNPs (right).

Single-cell whole genome sequencing (scWGS) was performed for 24 neurons from the BA6 region of 8 normal brains with ages ranging from 0 to 83 yo and established the base trendline of somatic mutation accumulation in neurons of BA6 region. The newly developed SCAN2 single-cell variant caller was used to identify single-cell private SNVs and Indels. BA6 neurons showed slightly higher sSNV rate than PFC neurons from the same normal individuals that our lab generated for another project (Data not shown). On the other hand, somatic Indels (sIndels) accumulate similarly in BA6 and PFC neurons. scWGS was then performed for 3 normal and 3 TDP43 depleted neurons from each of the BA6 of 3 ALS brains and PFC of 6 FTD brains. BA6 neurons of ALS brains showed no increase of sSNVs and a slight increase in sIndels (Fig. 6a, b). PFC neurons of FTD brains showed both significant increase of sSNVs and sIndels (Fig. 6c, d). Interestingly, normal PFC neurons exhibited higher burdens of sSNVs and sIndels compared to PFC neurons with TDP43 nuclear depletion. This pattern was also observed in sIndels in ALS neurons. These results suggest a cell-autonomous protecting role or non-cell-autonomous damaging role of TDP43 pathology. Mutational signature analyses and genomic region enrichment analyses are undergoing for these sSNVs and sIndels.

To identify somatic L1 retrotransposon insertions in single cells, we decided to combine human active transposon sequencing (HATseq) and PTA. Before applying it to FTD/ALS neurons, we first tested its performance. A test sequencing with the PTA-HATseq showed sensitivities of 98.3% and 93.1% in detecting known reference (KR) and known non-reference (KNR) insertions, respectively. This represents greater sensitivity than ever before reported for any single-cell method. The slightly lower sensitivity for the non-reference insertions is probably caused by the allele and locus dropout arising during single-cell WGA. Most non-reference insertions are heterozygous and more vulnerable to being missed due to allele and locus dropouts than reference insertions which are mostly present in two (homozygous) copies. This

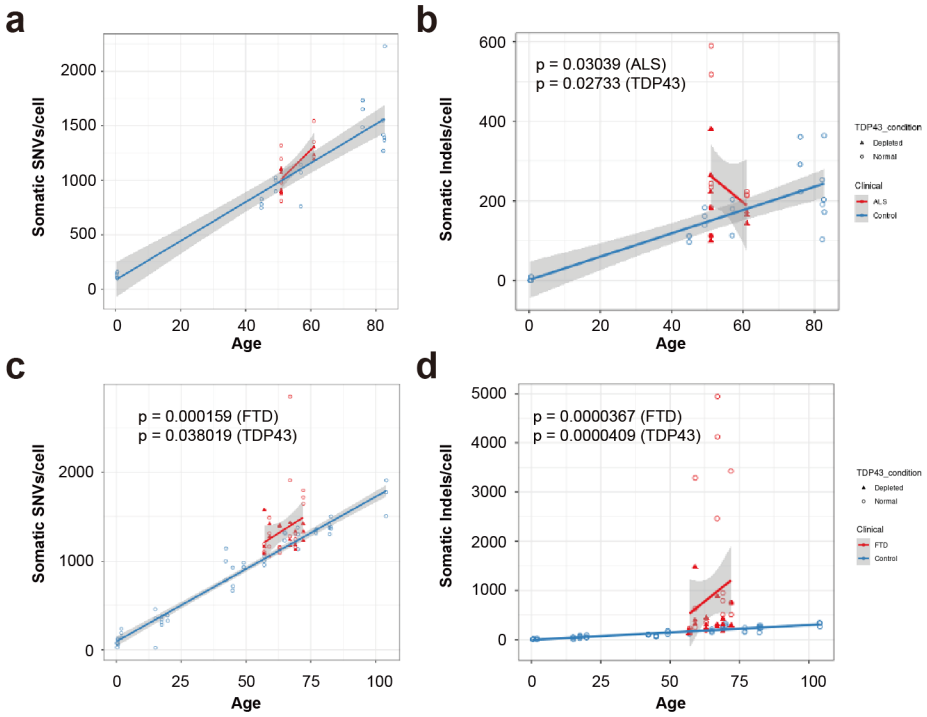

Figure 6. sSNVs and sIndels in scWGS data of FTD/ALS brains. a-b. sSNV burden (**a**) and sIndel burden (**b**) of BA6 neurons of ALS and control brains. Twenty-four control BA6 neurons are included. **c-d.** sSNV burden (**c**) and sIndel burden (**d**) of PFC neurons of FTD and control brains. Forty control PFC neurons are included. P value are calculated using a linear mixed model.

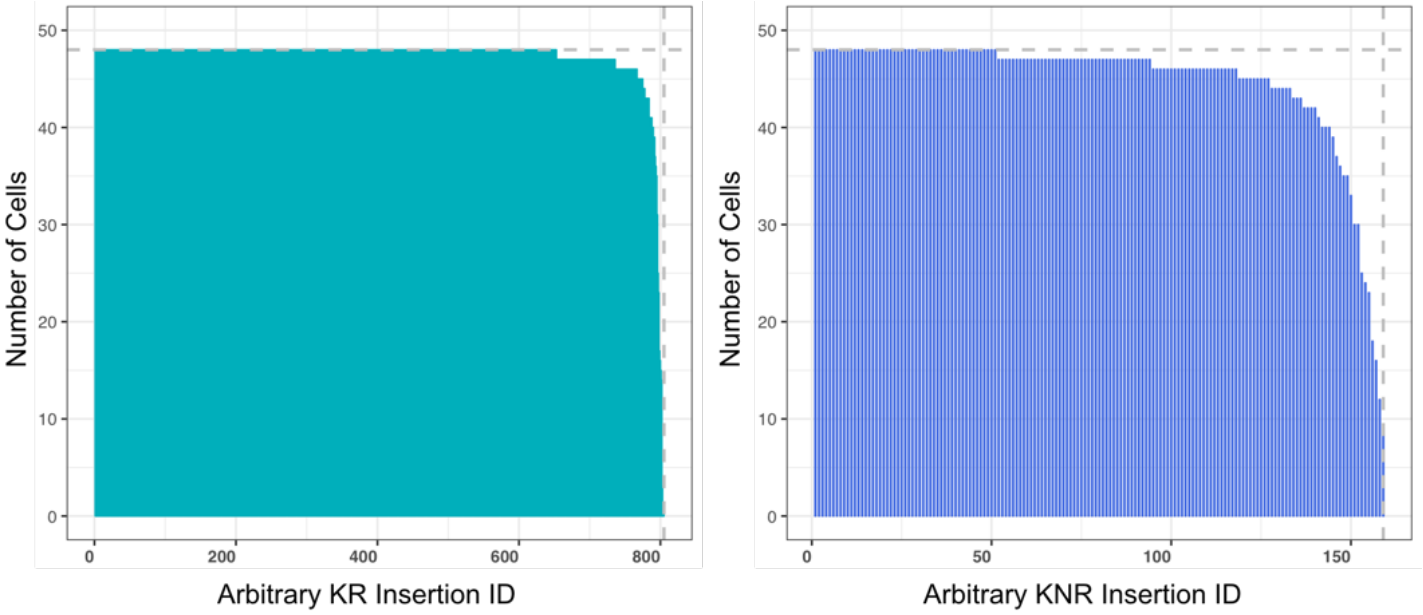

Figure 7. Recovery of germline KR and KNR insertions in single cells using PTA L1 targeted sequencing. 48 neurons of a normal human postmortem brain show great evenness of the recovery of germline KR and KNR insertions. Each vertical bar represents the number of neurons having each KR or KNR insertion was called. Since KR and KNR are germline insertions, all 48 single-neurons from the same donor are expected to have each insertion called.

combination approach also has very subtle variations in the sensitivity across single cells and the bulk DNA sample (Fig. 7). Our pilot test sequencing results suggest that the combination of PTA and L1 targeted sequencing can recover >93% of L1 insertions genome-wide with an average depth of 200X by multiplex-sequencing of 96 single-neuron HAT-seq libraries in one lane of the Illumina HiSeq X sequencer. The estimated rate of somatic L1 insertions is 0.242 [95%CI: 0.108, 0.377] per neuron using this approach which is slightly higher than earlier reports and generally consistent with more recent studies. Therefore, it is a robust and sensitive approach to quantitatively study somatic L1 insertions in single cells.

We then applied this method to single cells of both FTD/ALS brains. A bulk sample, 47 nomal neurons and 48 neurons with TDP43 nuclear depletion were sequenced using PTA-HATseq for each of the BA6 of 3 ALS brains and the PFC of 3 FTD brains. Similar sequencing depth and sensitivity were achieved across all samples regardless of the disease and TDP43 conditions. A higher L1 insertion rate was observed in normal neurons than neurons with TDP43

nuclear depletion (Fig. 8), which again suggested a non-cell-autonomous damaging role of TDP43 related dysregulation of retrotransposons. The identified L1 insertions will be validated through a combination of 5' and 3' junction PCR validations.

Specific Aim 3: Genome-wide burden of DSB in neurons of FTD/ALS brains

Due to the technical challenges of BLISS using postmortem tissues, we couldn't generate clear results of double-strand breaks in postmortem brains. Postmortem brains in general had very high background noise of double-strand breaks, likely due to the freeze and thaw cycles. In addition, the inter-sample variation was high. We have communicated with other labs that were also trying to use BLISS to study double-strand breaks in postmortem tissues and realized that it was a common problem. Although we are not able to determine DSBs directly, we expect that mutational signatures of the sIndels identified in Aim 2 should represent a mutational process of DSB and the associated DSB repair; *e.g.*, homologous recombination, non-homologous end joining and microhomology-mediated end joining. The mutational signature analysis is currently undergoing.

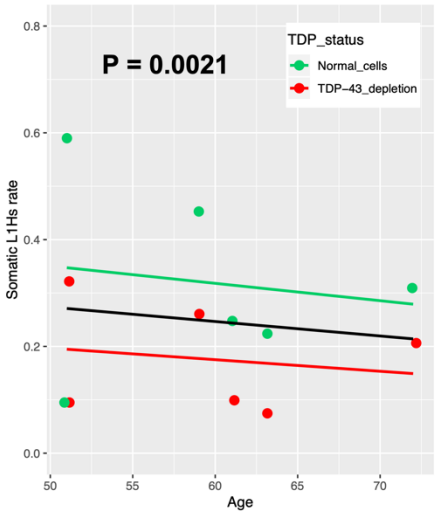

Figure 8. Somatic L1 insertions in FTD/ALS neurons. Normal neurons show higher rate of L1 insertions than neurons with TDP43 nuclear depletion in both FTD and ALS brains. Black line indicates the trendline of average L1 insertion rate of both normal and TDP43 depleted neurons.

What opportunities for training and professional development has the project provided?
Nothing to Report
How were the results disseminated to communities of interest?
Nothing to Report

What do you plan to do during the next reporting period to accomplish the goals?

We have identified clonal somatic mutations in FTD/ALS brain and spinal cord samples and validated some of the identified mutations. We have requested additional brain tissues and section slides of several cases carrying pathogenic somatic mutations for a more detailed regional distribution analysis. The regional distribution of pathogenic somatic mutations and TDP43 pathology through immunohistochemistry will also be compared for a potential correlation between the two. Cell-type analysis will be performed to see if the pathogenic somatic mutations have a preferential enrichment in neurons or glial cells. We have also requested the disease onset information of the FTD/ALS cases. The age of onset will be compared between cases carrying pathogenic germline mutations and somatic mutations.

The mutational signatures of sSNVs and sIndels from the scWGS data will be analyzed. Signatures will be made through both de novo signature decomposition and decomposition with COSMIC signatures. This should help figure out the potential mechanisms underlying the changes of sSNVs and sIndels in disease brains.

The candidate L1 retrotransposon insertions will be validated through both 5' and 3' junction PCR experiments. PTA-HATseq of neurons from aged-matched control brains will also be performed, which will be used as the baseline of L1 insertions.

Results generated from these studies will be wrapped up for two or three potential publications over the next year or so.

4. Impact

What was the impact on the development of the principal discipline(s) of the project?

Our study of clonal somatic mutations showed that a small fraction of FTD/ALS brains carry pathogenic somatic mutations that can explain the pathogenesis of the disease. Our findings will change the understanding of the pathogenesis of FTD/ALS and potentially change the diagnostic process of the diseases. Although our current results suggest that only 2-3% FTD/ALS cases can be potentially explained by pathogenic somatic mutations, the study could be limited by the detection sensitivity of ultra-low level somatic mutations and the limited number of brain regions involved in this study. Indeed, our results also suggest that the identified somatic mutations are very focal with VAFs around 0.5-1%. With more advanced sequencing approaches in the future, it should allow us to detect somatic mutations with VAFs <0.5% and presumably increase the fraction of FTD/ALS cases carrying pathogenic somatic mutations.

Our scWGS data showed that normal neurons exhibit higher rate of sSNVs, sIndels and somatic L1 insertions than neurons with TDP43 neuclear depletion in FTD/ALS brains. These results raise an interesting hypothesis that TDP43 pathology may have a non-cell-autonomous damaging role which leads to elevated cellular stress and somatic mutations in adjacent normal neurons. Further analyses and more experiments should be performed in the future to fully understand the mechanism.

What was the impact on other disciplines?
Nothing to Report
What was the impact on technology transfer?
Nothing to Report

What was the impact on society beyond science and technology?
Nothing to Report

5. Changes/Problems

Changes in approach and reasons for change

For whole-genome amplification of single cells, we performed primary template-directed amplification (PTA) instead of multiple displacement amplification (MDA). PTA is an improved method for whole-genome amplification of single cells which is based on MDA. PTA became commercially available in 2020 and showed a much better quality of genome amplification in our hands. Therefore, PTA was used for this project.

For the scWGS experiments, we decided to sequence additional neurons from the PFC of three more FTD brains. This was because of the unexpected findings that normal neurons tended to have higher burdens of sSNVs and sIndels than neurons with TDP-43 pathology, although the changes were not significant previously. Additional neurons indeed made the changes significant.

Actual or anticipated problems or delays and actions or plans to resolve them

Shortly after the grant started, our institution was forced to shut down due to the COVID-19 pandemic. Our institution reopened in July 2020, but has been only allowing a 50% capacity of lab members who can work at the same time. Brain banks that we requested samples from were also shut down in the meantime. This significantly slowed down our speed of getting the necessary human brain tissue samples and the associated documents from those brain banks. We immediately resumed the process of getting the samples after the brain banks reopened and submitted the Human Research Protection Office (HRPO) form for review in September 2020. The HRPO approval was received in February 2021. The COVID-19 pandemic has delayed our process of getting the samples and the downstream experiments using them for about 6 months. We requested an one-year no-cost extension for this grant, which should give us enough time to finish the proposed studies. In addition, the originally proposed BLISS sequencing experiments turned out to be technically challenging with postmortem tissues. We don't think this issue can be resolved before the end of this grant. We will focus on the mutational signature analyses of sIndels, which can potentially link specific indel signatures to DSB and associated DSB repair processes.

Changes that had a significant impact on expenditures
Nothing to Report
Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
Nothing to Report
Significant changes in use or care of human subjects
Nothing to Report
Significant changes in use or care of vertebrate animals
Nothing to Report
Significant changes in use of biohazards and/or select agents
Nothing to Report

6. Products

Publications, conference papers, and presentations
Nothing to Report
Website(s) or other Internet site(s)
Nothing to Report
Technologies or techniques
Nothing to Report
Inventions, patent applications, and/or licenses
Nothing to Report
Other Products
Nothing to Report

7. Participants & Other Collaborating Organizations

What individuals have worked on the project?
Zinan Zhou: no change
Junho Kim:

Name:	*Junho Kim*
Project Role:	*Collaborator*
Researcher Identifier (e.g. ORCID ID):	0000-0003-3513-401X
Nearest person month worked:	3
Contribution to Project:	*Dr. Kim has performed computational analyses for all the studies.*
Funding Support:	*Assistant Professor startup package from the Sungkyunkwan University*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Zinan Zhou is partially supported by the postdoctoral fellowship of the Edward R. and Anne G. Lefler Center at Harvard Medical School from June 2021 to June 2023.
Junho Kim became an Assistant Professor at Sungkyunkwan University starting from July 2021. He is fully supported by his startup package since then.
What other organizations were involved as partners?

- **Organization Name:** Sungkyunkwan University
- **Location of Organization:** South Korea
- **Partner's contribution to the project**
    - Collaboration: Dr. Junho Kim is now working on these projects as a collaborator. He has been helping with the computational analyses of the studies including somatic detections in both bulk and single-cell samples.

8. Special Reporting Requirements

Appendices