AWARD NUMBER: W81XWH-20-1-0028

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

PRINCIPAL INVESTIGATOR: Zinan Zhou

CONTRACTING ORGANIZATION: Boston Children's Hospital

REPORT DATE: March 2021

TYPE OF REPORT: ANNUAL

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
March 2021	ANNUAL	1FEB2020 - 31JAN2021

4. TITLE AND SUBTITLE

Systematic Analysis of Genetic Mosaicism in FTD/ALS Brains

5a. CONTRACT NUMBER

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

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Zinan Zhou/ Junho Kim

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

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Boston Children's Hospital
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 have collected human postmortem brain tissues from several brain banks. We have also tested the targeted panel sequencing and single-cell L1 targeted sequencing approaches and gotten great performance from testing experiments. In the next funding period, we will apply these approaches to all collected human postmortem brain tissues to systematically examine the burden and pattern of various types of somatic mutations in FTD/ALS brains.

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
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	9	

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 recently started to extract gDNA from these samples and have completed about 20% of the samples.
Milestone at 10 months: Raw sequencing data
This has not been started yet. We expect to have all samples sequenced in the summer of 2021.

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
This has not been started yet since the HRPO forms were just approved recently. The downstream experiments are also pending. We will catch up the timeline in the second year of the this grant. The primary template-directed amplification (PTA) kits became commercially available in 2020 and showed a much better quality of whole-genome amplification of single cells in our hands. This new method for whole-genome amplification will result in single-cell genome with great quality and save us time in doing quality control experiments. Therefore, we should be able to catch up our timeline to complete this major task in the second year of this grant.

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

Milestone at 6 months: Raw data of L1 targeted sequencing
This has not been started yet since the HRPO forms were just approved recently. However, we have improved our proposed L1 targeted sequencing method and the data analysis pipeline while we were waiting for the HRPO approval. We should be able to complete this major task in the second year of this grant.

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
This has not been started yet since the analysis requires data generated in the Major Task 2. We should be able to complete this major task in the second year of this grant.

Major Task 5: Identify DSB using BLISS

Milestone at 6 months: Raw data of the BLISS sequencing
This has not been started yet since the HRPO forms were just approved recently.

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 human postmortem brain tissues from brain banks and gotten the HPRO forms approved. We have also extracted gDNA from about 20% of these samples. We have designed the MIPs panel targeting the exons and exon-intron junctions of 88 neurodegeneration related genes, which includes 30 ALS/FTD genes, 8 AD genes, 27 PD genes as well as genes associated with other rare dementia. We performed a spike-in experiment to test the performance of the MIPs panel 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%. We developed and applied a customized pipeline for calling somatic variants from the spike-in

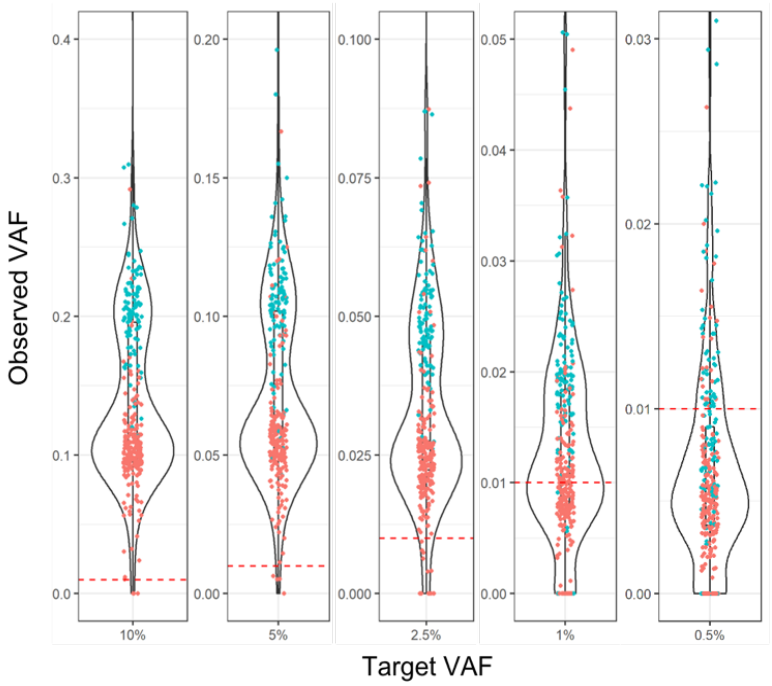

Figure 1. VAFs of somatic variants called in the spike-in MIPs sequencing data. *The gDNA of GIAB12878 was mixed into the gDNA of GIAB24695 at the listed target VAFs for heterozygous germline variants of GIAB12878. Heterozygous germline variants (red) are clustered at the expected VAFs and homozygous germline variants (green) are clustered at VAFs with doubled values.*

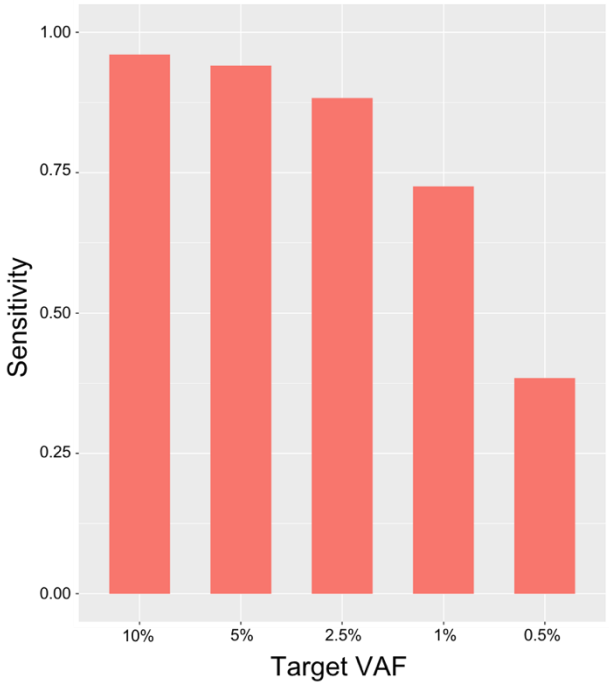

Figure 2. Sensitivity of the customized pipeline in calling somatic variants from the spike-in MIPs sequencing data. *The sensitivity off our customized pipeline range from ~40% to 95% at different target VAFs.*

test experiment with our MIPs panel. The observed VAFs of somatic variants were very well in line with the target VAFs at all five VAF ratios (Fig. 1). And the sensitivity of our customized pipeline for calling somatic variants from the spike-in MIPs data ranged from 40% to 95% (Fig. 2). The MIPs panel sequencing, together with our customized pipeline, allows us to confidently identify clonal somatic mutations even with a VAF of 0.5%. Our test sequencing result also suggested that this experimental set-up will lead to an average depth of 2000X and an average on-target rate of 85% by sequencing 48 MIPs libraries using one lane of the Illumina HiSeq X sequencer. Although we couldn't finish sequencing these brain

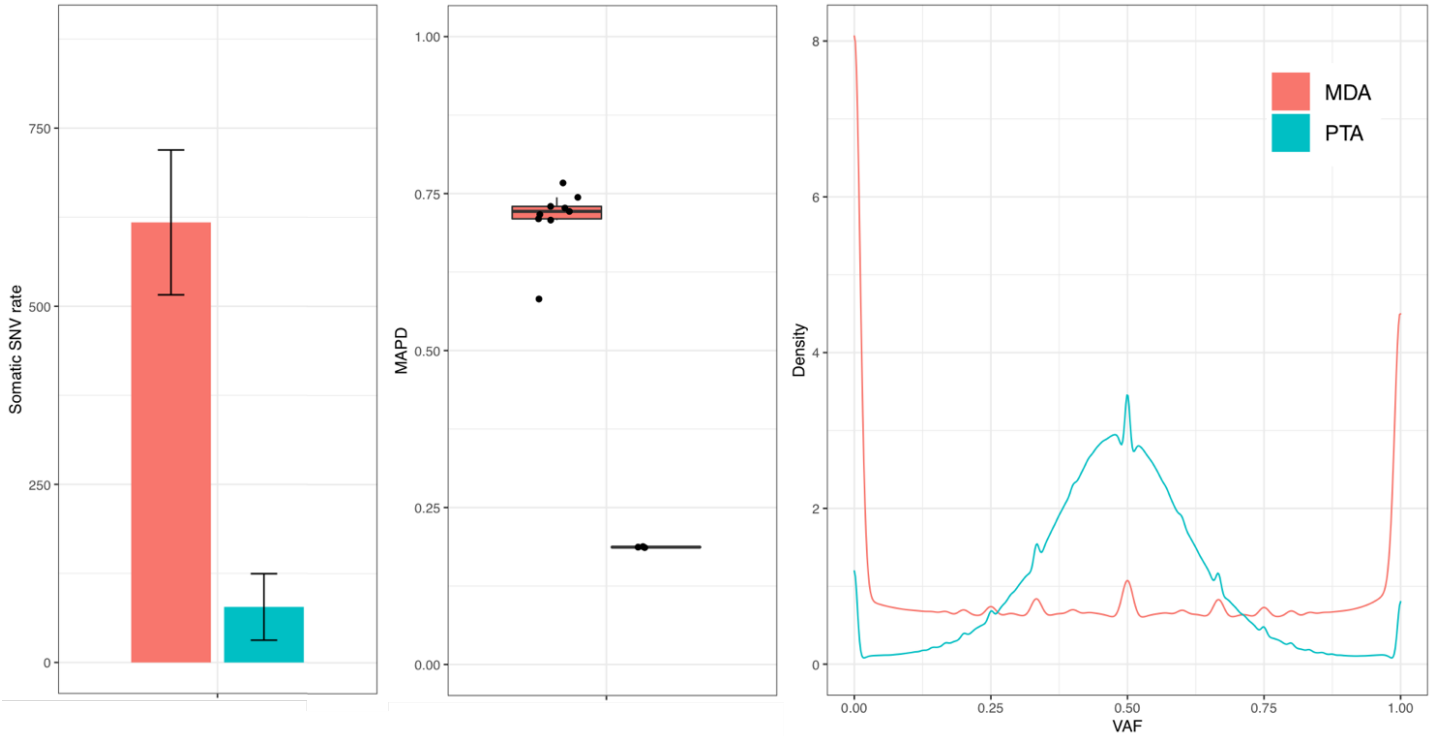

Figure 3. Comparison between PTA and MDA. *9 neurons amplified by MDA and 3 neurons amplified by PTA from the postmortem brain of a normal individual at the age of five months were compared. PTA-amplified neurons showed lower somatic SNV rate (left), lower variability in sequencing depth/better evenness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

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

We have not started to perform single-cell whole genome sequencing and single-cell L1 targeted sequencing yet as the HRPO was just approved recently. However, we tested the new primary template-directed amplification (PTA) kits for whole genome amplification (WGA) recently developed by BioSkryb. 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. 3). 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. This means we will be able to generate scWGS faster and catch up the timeline of our proposal in the second year of this grant.

We also tested the performance of our L1 targeted sequencing with DNA amplified from single cells with PTA. A test sequencing with the combination of PTA and L1 targeted sequencing shows 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 combination approach also has very subtle variations in the sensitivity across single cells and the bulk DNA sample (Fig. 4). 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 will apply this method to single cells of both FTD/ALS and control brains to compare the burden and pattern of L1 retrotransposon insertions.

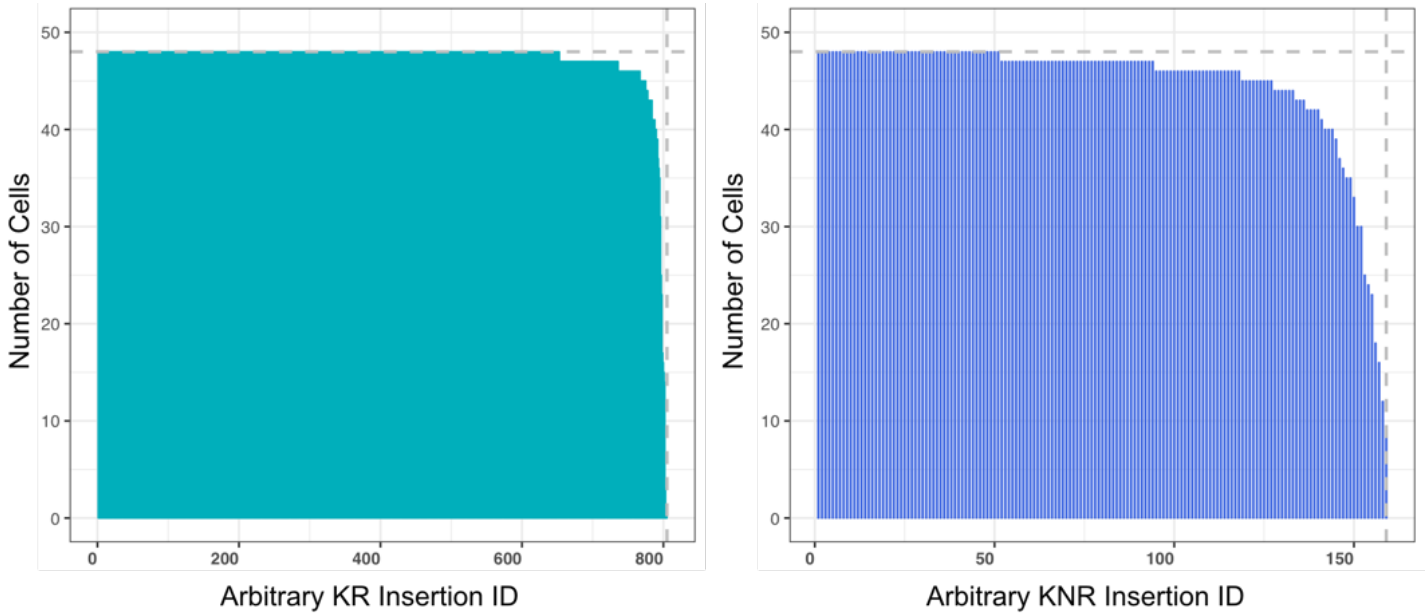

Figure 4. Recovery of germline KR and KNR insertions in single cells using PTA L1 targeted sequencing. *48 neurons of a normal human postmortem brain show great evenness of the recovery of germline KR and KNR insertions. Each vertical bar represents the number of neurons having each KR or KNR insertion was called. Since KR and KNR are germline insertions, all 48 single-neurons from the same donor are expected to have each insertion called.*

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

These experiments have not been started yet. Once we get the scWGS data generated in Specific Aim 2, we will apply our indel calling pipeline to the data and compare the indel between FTD/ALS and control cases. BLISS sequencing will also be done for the human brain tissues that we recently got.

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?

As we have gotten all human brain tissues for our studies, we will start to generate data as we originally proposed in the specific aims. We will extract gDNA from all these samples and make MIPs targeted sequencing libraries for them. The analysis of MIPs data should take less time since we have developed a custom pipeline for the data analysis and

tested it performance with reference gDNA. The good performance of our MIPs design and custom pipeline suggests that we should be able to get a set of candidate variants with a high validation rate. These candidates will then be screened for pathogenic variants and validated.

We will also generate scWGS data for both FTD/ALS and control brains using the newly developed PTA method. The PTA method amplifies DNA of single cells with high quality, which then allows us to prepare much less cells to get enough number of cells for our studies and skip the quality control step with low-coverage sequencing. This will save us time and allow us to catch up the timeline of our proposal in the second year of this grant. The scWGS generated with PTA will then be analyzed to identify somatic SNVs and indels. The burden and mutational signatures of somatic SNVs and indels between FTD/ALS and control brains will be compared.

DNA of single cells amplified by PTA will also be used for making L1 targeted sequencing libraries. As we have shown, the combination of PTA and L1 targeted sequencing represents a robust approach to identify somatic L1 retrotransposon insertions in single cells. We will use this approach to compare the burden and pattern of L1 retrotransposon insertions between FTD/ALS and control brains.

Finally, we will generate BLISS libraries to capture double-strand breaks in neurons. The burden of double-strand breaks will be compared between FTD/ALS and control brains.

4. Impact

What was the impact on the development of the principal discipline(s) of the project?
Nothing to Report
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 will perform 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, we will perform PTA for this project.

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 spent most of our time during the lockdown on computational works including the MIPs panel design and writing data analysis pipelines.

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: no change

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report

What other organizations were involved as partners?

Nothing to Report

8. Special Reporting Requirements

Appendices