Origin and Role of Nuclear Somatic Mutations in Frontotemporal Dementia

Sulekha Raveendran Nair
PhD candidate, Gonville and Caius College

Department of Clinical Neurosciences
University of Cambridge, UK

This thesis is submitted for the degree of Doctor in Philosophy
May 2022

Statement of Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. I further state that no substantial part of my thesis has already been submitted or being concurrently submitted, in part or whole, for any such degree, diploma, or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed limit (60,000 words) for the relevant Degree Committee (Faculties of Clinical Medicine and Veterinary Medicine).

Signed: **Sulekha R Nair**

Date: **30-May-2022**

Sulekha Raveendran Nair
PhD Student, Department of Clinical Neurosciences, Gonville and Caius College, University of Cambridge, UK.

Abstract

Origin and Role of Nuclear Somatic Mutations in Frontotemporal Dementia

– Sulekha Raveendran Nair

Ageing is a natural process that occurs in all cells of multicellular organisms, including non-dividing cells like neurons, and is associated with the accumulation of mutational burden and a loss of genomic maintenance over time. Recent advances in technology have brought attention to the accumulation of somatic mutations in neurons that may cause ageing and several specific mutational signatures have been associated with disease states. Frontotemporal Dementia (FTD) is a young-onset form of dementia displaying great heterogeneity with respect to clinical phenotype and pathology. Several genes have been implicated in FTD, but their true mutational prevalence has been poorly understood due to the limitations of accuracy and reproducibility of bulk and single-cell sequencing technologies at the sub-clonal levels, respectively.

Duplex Sequencing, the error-corrected next generation sequencing (NGS) technique that combines the strengths of bulk and single-cell sequencing, was applied to genomic DNA extracted from three different frozen brain regions of a patient with FTD-Tau pathology and an age-matched control. Two main approaches to DS were explored – the amplicon-based and the double capture-based methods – and two different protocols were investigated for each method. Optimizations of several parameters were required, particularly of the library amplification step, to obtain the optimal number of reads to make a duplex consensus of true mutations in each sample. Although the amplicon-based method was easier and less time-consuming, the double capture-based method that used a greater number of biotinylated probes proved able to remove more false-positives and detect true mutations more accurately.

Due to the selective degeneration of the frontal and temporal lobes in FTD, we hypothesized that the number of somatic mutations detected in the temporal sample of the FTD-Tau patient would be higher than the other brain regions of the patient which would in turn be greater than the age-matched control. While the overall number of mutations detected in the

FTD-Tau patient is higher than in the age-matched control, the overall number of mutations detected in the temporal lobe were comparable to those detected in the occipital lobe and the medulla brain regions of the patient. However, the per-bp mutation rate of the temporal lobe of the patient was the highest of the six brain regions analysed, indicating that this brain region may be more susceptible to mutation accumulation in the disease state. Despite being a sporadic case of FTD, the patient exhibited a statistically significant presence of early-arising somatic single nucleotide variations (SNVs) (variable allele frequency, VAF $\geq$ 0.05 and VAF < 0.25), ultralow frequency somatic SNVs (VAF < 0.01), and germline SNVs (VAF > 0.25 and VAF < 1.00), indicating a complex interplay of mutational accumulation being associated with the FTD-Tau disease state in the patient. Somatic SNVs (VAF < 0.25) particularly found within the intronic, exonic, ncRNA, and the UTR regions have a higher presence in the patient than the control, indicating a relationship between somatic mutational burden in genomic regions not directly involved in the transcription and translation of the tau gene and disease state in the patient.

This study is important in that it sheds some light on the origin and role of nuclear mutations, particularly that of ultralow frequency somatic SNVs, in the pathogenesis of the FTD-Tau disease state in a sporadic patient case *versus* an age-matched control by successfully applying the error-corrected NGS technique, duplex sequencing, on DNA extracted from different brain regions of both the individuals.

Acknowledgements

I have received a great deal of support and encouragement during the writing of this thesis.

I would first like to thank my supervisor, Professor Patrick Chinnery, for the opportunity to earn a doctorate degree at his lab and for his constant support and guidance throughout these years.

I am indebted to Dr Yu (Nell) Nie, my post-doctoral mentor, without whose supervision and assistance, this project would have been incomplete. Her enthusiasm and drive inspired me to work meticulously and tirelessly in the lab. Thank you, Nell, for teaching me all that I know on how to conduct experiments on the bench and for performing most of the data analyses described in this thesis.

A special thank you to Dr Claudia Calabrese for taking the time to help me with some of the coding and calculations to determine the mutation frequencies of the samples.

I would also like to acknowledge all my colleagues in my lab and department for their wonderful support and the warm friendships I have come to cherish over the years.

I am grateful to the Cambridge Trust, and in particular, the Cambridge India Ramanujan Scholarship, for not only funding my PhD but for also helping me build a network of fellow scholars and friends that allowed me to experience life at Cambridge to the fullest.

I am appreciative of the efforts of Dr Ruth Scurr, my Tutor at Gonville and Caius for a majority of my PhD, who was invaluable in providing College aids to better support my health.

A big thank you to Pagen, Chris, Shelly, and Deb for their efforts in making sure I was physically and mentally well enough to tackle another day of work and helped to greatly manage the pain levels in my joints and exhaustion in the past three years.

And lastly, there are not enough words to convey how thankful I am to my parents and my sister who have all taken turns to be my counsellors and cheerleaders irrespective of the time I called and what was happening in their lives. They are the pillars of my strength and inspiration, and their abiding love, support, and encouragement have meant the world to me.

– Sulekha R Nair

Table of Contents

Statement of Declaration ..... **i**

Abstract ..... **ii**

Acknowledgements ..... **iv**

Table of Contents ..... **v**

List of Figures ..... **xi**

List of Tables ..... **xiv**

List of Abbreviations ..... **xvi**

Chapter I: Introduction **1**

1.1 The human nervous system 1

1.2 Neurodegenerative Diseases 2

1.2.1 Pathology of neurodegeneration 2

1.2.2 Risk factors for neuropathogenesis 3

1.2.3 Cellular mechanisms of neuropathogenesis 6

1.2.3.1 Neuronal mechanisms 6

1.2.3.2 Non-neuronal mechanisms 7

1.2.4 Presentation of neuropathology: common neuronal disorders 8

1.3 Somatic Brain Mosaicism 14

1.3.1 Types of somatic mutations 16

1.3.2 Mechanisms for somatic mosaicism 17

1.3.3 Detection of somatic mosaicism 19

1.3.3.1 Relevance: Mutational Burden, Ageing, and Disease 19

1.3.3.2 Approaches to study mosaicism 21

1.4 Detection of somatic mutation using duplex sequencing in frontotemporal dementia brain samples 26

1.4.1 What is Frontotemporal dementia? 26

1.4.2 Prevalence of FTD 27

1.4.3 Pathogenesis of FTD 28

1.4.4 Genetics of FTD 31
1.4.4.1 Detection of FTD mutations using NGS techniques 31
1.4.4.2 Nuclear single nucleotide variants associated with FTD 35
1.4.4.3 Mitochondrial single nucleotide variants associated with FTD 40
1.4.4.4 Epimutations associated with FTD 41
1.4.5 FTD: Current Therapeutic Practices 43
Chapter II: Research Aims 45
Chapter III: Materials and Methods 46
3.1 Materials 46
3.1.1 FTD cases and controls 46
3.1.2 Reagents 47
3.1.3 Equipment 48
3.1.4 Software 49
3.2 Methods 50
3.2.1 Genomic DNA Extraction 50
3.2.2 Size selection and DNA Purification with SPRI 51
3.2.2.1 Principle of solid-phase reversible immobilization (SPRI) 51
3.2.2.3 Single-tail beads selection 52
3.2.2.4 Double-tail beads selection 52
3.2.3 Duplex library Preparation using KAPA protocol 53
3.2.3.1 Enzymatic Fragmentation 53
3.2.3.2 End-Repair and A-tailing (ER-AT) 53
3.2.3.3 Adapter-ligation 54
3.2.3.4 Post-ligation purification 55
3.2.3.5 Library Amplification 55
3.2.3.6 Post-amplification purification 56
3.2.3.7 Quantitative polymerase chain reaction (qPCR) (For optimizations) 56
a) qPCR primers 57
b) qPCR programme 57

3.2.4 Amplicon-based method of duplex library preparation 58
3.2.4.1 Pre-library amplification with polymerase chain reaction (PCR) 58
a) Primers 58
b) PCR programmes 58
3.2.4.2 Agarose gel preparation and Image Acquisition 59
3.2.4.3 Duplex library preparation 60
3.2.5 Double capture-based method of duplex library preparation – Protocol 1 (with in-house biotinylated baits) 60
3.2.5.1 Duplex Library Preparation 61
3.2.5.2 Biotinylated baits preparation 61
3.2.5.3 Buffers and solutions 61
3.2.5.4 Baits-beads mixture preparation 62
3.2.5.5 Hybridization Mixture 64
3.2.5.6 Capture 1 65
3.2.5.7 Post-capture 1 washing 65
3.2.5.8 Capture 2 66
3.2.6 Double capture-based method of duplex library preparation – Protocol 2 (with customized IDT probes) 67
3.2.6.1 Duplex library preparation 67
3.2.6.2 Probes 67
3.2.6.3 Hybridization reaction between duplex libraries and probes 68
3.2.6.4 Preparation of buffers 68
3.2.6.5 Streptavidin beads wash 69
3.2.6.6 Bead capture 69
3.2.6.7 Washes 69
3.2.6.8 Heated Washes 69
3.2.6.9 Room temperature washes 70
3.2.6.10 Post-capture 1 PCR amplification 70
3.2.6.11 Second round of capture 70

3.2.6.12 Post-capture 2 PCR amplification 71
3.3 Analyses 71
3.3.1 Duplex sequencing computational workflow 71
3.3.1.1 Preparing a reference genome for use 71
3.3.1.2 Parsing and filtering duplex tags 72
3.3.1.3 Bioinformatics processing 72
3.3.2 Statistical Analyses 73
3.4 Summary of the different approaches to DS explored in the thesis 74
Chapter IV: Optimization of Amplicon-based Duplex Sequencing 76
4.1 Introduction 76
4.2 Results 78
4.2.1 Preparations of MAPT amplicons 78
4.2.1.1 GXL DNA polymerase-based amplified products 78
4.2.1.2 SuperFi DNA polymerase-based amplified products 81
4.2.2 Library amplification optimizations 83
4.2.2.1 Amplification cycle number – Optimizations 84
4.2.2.2 DNA input amount for amplification – Optimizations 86
4.2.3 qPCR verification test 88
4.2.4 Computational analyses of sequencing data 90
4.3 Discussion 92
Chapter V: Optimization of Double capture-based Duplex Sequencing – Protocol 1 95
(With in-house biotinylated baits) 95
5.1 Introduction 95
5.2 Results 96
5.2.1 Duplex Library preparation 96
5.2.2 Biotinylated baits preparation 96
5.2.3 Capture protocol optimizations 97
5.2.4 Computational sequencing data analyses 102
5.2.4.1 Family size 102

5.2.4.2 On-target ratio 102
5.3 Discussion 103
Chapter VI: Optimization of Double capture-based Duplex Sequencing – Protocol 2 (With custom-designed IDT probes) 105
6.1 Introduction 105
6.2 Results 106
6.2.1 Library Preparation optimizations for Protocol 2 107
6.2.1.1 Duplex library preparation 107
6.2.1.2 Probes 107
6.2.1.3 Capture 1 107
6.2.1.4 Capture 2 108
6.2.2 Computational sequencing data analyses 110
6.2.2.1 Family Size 110
6.2.2.2 On-target ratio 112
6.2.2.3 Number of reads: Q30 vs DCS 112
6.2.2.4 Depth of Sequencing 115
6.2.2.5 Mutation detection 116
6.3 Discussion 116
Chapter VII: Optimized Duplex Sequencing on all FTD-Tau samples 118
7.1 Introduction 118
7.2 Results 119
7.2.1 Amplicon-based approach to duplex sequencing 119
7.2.1.1 Pre-library amplifications 119
7.2.1.2 Combining the three MAPT amplicons in equimolar proportions for each FTD sample 120
7.2.1.3 Duplex Library Preparations 120
7.2.1.4 Computational Sequencing Data Analyses 121
7.2.1.5 Summary of results for SuperFi-mediated amplicon-based DS 129

7.2.2 Double capture-based approach to duplex sequencing (with custom designed IDT probes for whole MAPT gene) 130
7.2.2.1 Duplex library preparation 130
7.2.2.2 Probes 130
7.2.2.3 Capture 1 130
7.2.2.4 Capture 2 131
7.2.3 Computational sequencing data analyses 135
7.2.4 Summary of results for Double capture Protocol 2 DS (with IDT probes for whole MAPT gene) 149
7.3 Discussion 149
Chapter VIII: Conclusions and Future Directions 151
8.1 Primary Findings and Discussions 152
8.1.1 Comparison between the different duplex sequencing approaches 152
8.1.2 Mutational burden and role of somatic and germline SNVs in the FTD-Tau patient 153
8.1.3 Origin of somatic mutations in FTD disease progression 154
8.2 Strengths of the research project 155
8.3 Limitations of the research project 157
8.4 Future Investigations 157
8.4.1 Applying optimized DS protocol to other FTD pathologies 157
8.4.2 Single-cell DNA sequencing 158
8.5 Summary 159
References 160
Appendix I 190
Appendix II 192
Appendix III 194
Appendix IV 199
Appendix V 201
Appendix VI 203
Appendix VII 241

List of Figures

Figure 1: Association of Mutational Burden with Ageing and Neurodegenerative Diseases. 4
Figure 2: Somatic brain mosaicism..... 15
Figure 3: The limit of detecting somatic mutations. 20
Figure 4: Principle of Duplex Sequencing..... 24
Figure 5: Frequency of each of the three genetic groups by geographic location in familial FTD. 33
Figure 6: Clinicopathological spectrum of FTD. 39
Figure 7: Principle of Ampure XP beads - DNA fragment size vs beads ratio. 51
Figure 8: Duplex adapter sequence and structure..... 54
Figure 9: Schematic of the basic computational workflow for duplex sequencing 72
Figure 10: Flowchart to demonstrate the different approaches to duplex sequencing explored in this thesis. 74
Figure 11: Flowchart of amplicon-based duplex library preparation..... 77
Figure 12: Fidelity rate of commonly-used DNA polymerases..... 77
Figure 13: GXL-based MAPT-A amplicon optimization 79
Figure 14: GXL-based MAPT-B amplicon optimizations 79
Figure 15: GXL-based MAPT-C amplicon optimization 80
Figure 16: SuperFi-based MAPT-A amplicon optimization 81
Figure 17: SuperFi-based MAPT-B amplicon optimization..... 82
Figure 18: SuperFi-based MAPT-C amplicon optimization 83
Figure 19: Amplicon-based duplex library preparation for library amplification optimizations. 84
Figure 20: Library amplification cycle number optimizations 85
Figure 21: DNA input amounts for PCR – Optimizations (set 1: 4, 8, 12, 16 amol) 87
Figure 22: DNA input amounts for PCR - Optimizations (set 2: 10, 1.67, 0.28 amol)..... 88
Figure 23: qPCR verification for the optimization of amplicon-based duplex libraries. 89
Figure 24: Amplicon-based DS - Family size influenced by library amplification parameters. 90
Figure 25: Amplicon-based duplex libraries – Family size calculations. 91
Figure 26: Amplicon-based DS: Mutation frequencies detected. 92
Figure 27: Double capture-based duplex library preparation - Protocol 1 (with in-house biotinylated baits)..... 96
Figure 28: In-house biotinylated baits preparation for double capture (Protocol 1). 97
Figure 29: Attempt 1: Library amplification optimizations post-both captures. 98
Figure 30: Attempt 2: Library amplification optimizations post-capture 1. 99
Figure 31: Attempt 2: Library amplification optimizations post-capture 2. 100
Figure 32: Attempt 3: Library amplification optimizations post-both captures. 101

Figure 33: Double-capture Protocol 1 – Family size calculations 102
Figure 34: Double-capture Protocol 1 – On-target ratio calculations. 103
Figure 35: Double capture-based duplex library preparation - Protocol 2 (with custom designed IDT probes) 105
Figure 36: Amplicon-based *vs* Double capture-based duplex library preparation Protocol 2 - Post-capture 1 library amplifications 108
Figure 37: Double capture-based duplex library preparation Protocol 2 - Post-capture 1 library amplifications..... 110
Figure 38: Double capture-based Protocol 2 – Family size calculations. 111
Figure 39: Double capture-based Protocol 2 – On-target ratio calculations..... 112
Figure 40: Double capture-based Protocol 2 – Number of reads. 114
Figure 41: Double capture-based Protocol 2 – Depth of sequencing..... 115
Figure 42: Double capture-based Protocol 2 – Detection of mutation frequencies..... 116
Figure 43: Pre-library amplifications of the three MAPT amplicons with the SuperFi polymerase. .. 119
Figure 44: Amplicon-based duplex library preparation (SuperFi-mediated)..... 121
Figure 45: SuperFi-mediated amplicon-based DS – Family size calculations..... 122
Figure 46: SuperFi-mediated amplicon-based DS – Depth of Sequencing. 123
Figure 47: SuperFi-mediated amplicon-based DS – Mutational burden detection. 124
Figure 48: SuperFi-mediated amplicon-based DS – Distributions of detected mutations..... 126
Figure 49: Double capture-based DS: Post-capture 1 library amplification..... 131
Figure 50: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 1 132
Figure 51: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 2 132
Figure 52: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 3..... 133
Figure 53: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimizations 4 and 5 133
Figure 54: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 6 134
Figure 55: HS D5000 results for final amplified duplex libraries for NovaSeq. 135
Figure 56: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Family size calculations. 136
Figure 57: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Check for the "spacer" sequence. 138

Figure 58: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – On-target ratio calculations. 139
Figure 59: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Depth of sequencing. 139
Figure 60: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Mutational burden detection. 141
Figure 61: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Somatic mutational burden within different genomic regions. 144
Figure 62: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Mutational frequency calculations. 147
Figure 63: Mutation detection on FTD-Tau patient and control with conventional NGS 156

List of Tables

Table 1: List of different known types of age-related somatic mutations. 17
Table 2: Mutational signatures seen in aged and diseased brain samples. 20
Table 3: The advantages and limitations of the different sequencing approaches 25
Table 4: Types of FTD based on clinical symptoms. 26
Table 5: Types of FTD based on molecular pathology. 27
Table 6: Pathology and clinical phenotypes of gene mutations associated with FTD. 38
Table 7: Detailed description of FTD disease cases and controls. 46
Table 8: Enzymatic fragmentation reaction contents 53
Table 9: Enzymatic fragmentation incubation conditions. 53
Table 10: End-repair and A-tailing (ER-AT) reaction contents 54
Table 11: End-repair and A-tailing (ER-AT) incubation conditions 54
Table 12: Adapter-ligation reaction contents 54
Table 13: Adapter-ligation incubation conditions 55
Table 14: PCR tube contents for one reaction of duplex library amplification 55
Table 15: PCR programme for duplex library amplification optimization – varying DNA input 56
Table 16: PCR programme for duplex library amplification optimization – varying PCR cycle number 56
Table 17: Oligonucleotide sequences for MAPT amplicons qPCR primers and probes 57
Table 18: Assays for qPCR verification of adapter-ligated duplex library 57
Table 19: qPCR programme for verification of adapter-ligated duplex library 58
Table 20: Primers designed for MAPT amplicons of interest. 58
Table 21: PCR tube contents for one reaction of pre-library amplification with SuperFi and GXL DNA polymerases 59
Table 22: PCR programmes for pre-library amplification for the MAPT amplicons 59
Table 23: Description of primers used to make baits DNA. 61
Table 24: Protocol for the preparation of buffers and solutions for double-capture protocol 1 62
Table 25: Calculations for combining all six MAPT in-house baits in equimolar proportions. 63
Table 26: List of blocking oligos used in the hybridization mixture for capture. 64
Table 27: Contents of Hybridization Mixture I per sample 65
Table 28: Thermocycler programmes for IDT capture protocol 68
Table 29: PCR tube contents for one reaction of duplex library amplification post-capture 2. 70
Table 30: Advantages and disadvantages of each sequencing approach to DS employed in this thesis. 75
Table 31: Primers used to amplify the amplicon with the PCR programme used. 84

Table 32: Calculations for combining the three MAPT amplicons in equimolar proportions. 120
Table 33: SuperFi-mediated amplicon-based DS – Student's t-test to calculate significance of mutational burden in FTD-disease *vs* control. 125
Table 34: SuperFi-mediated amplicon-based DS – Kolmogorov-Smirnov test to calculate significance of mutational burden in FTD-disease *vs* control. 128
Table 35: Overall per-bp mutation rate calculations per sample. 241
Table 36: Overall per-bp mutation rate per genomic region per sample calculations. 244

List of Abbreviations

Acronym	Definition
A	Adenine
ac	Acetylation
ACMG	American College of Medical Genetics and Genomics
AD	Alzheimer's Disease
aFTLD	Atypical Frontotemporal Lobar Degeneration
AGD	Argyrophilic Grain Disease
AL	Adapter-ligation (Third step of KAPA HyperPlus Protocol for duplex library preparation)
ALS	Amyotrophic Lateral Sclerosis
Amp.	amplified
APOE	Apolipoprotein E
APOJ	Apolipoprotein J
APP	Amyloid-β Precursor Protein
ARX	Gene encoding the Aristaless Related Homeobox protein
BAM	Binary Alignment Map
BBB	Blood-Brain Barrier
BIBD	Basophilic Inclusion Body Disease
BO (1-6)	Blocking Oligonucleotides 1-6
bvFTD	Behavioural-variant Frontotemporal Dementia
BWA	Burrows-Wheeler Aligner
BWT	Bind-Wash-Tween buffer
C	Cytosine
cap	Capture
cat. no.	Catalogue Number
CBD/S	Corticobasal Disease/Syndrome
CDC	Centre for Disease Control, US
CDKL5	Cyclin-dependent kinase-like 5
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain containing 10
ChIP-seq	Chromatin Immunoprecipitation and massively parallel sequencing
CHMP2B	Chromatin-Modifying Protein 2B
CLN	Genes encoding Ceroid Lipofuscinosis, Neuronal proteins
CMA	Chromosomal Microarray Analysis
CNS	Central Nervous System
CNTN-1	Contactin-1
CNVs	Copy Number Variations
CpG islands	Cytosine-phosphate-Guanine (short stretches of palindromic DNA) islands
CS	Cockayne Syndrome
CTE	Chronic Traumatic Encephalopathy
C9ORF72	Chromosome 9 Open Reading Frame 72
dATPs	Deoxyadenosine Triphosphates
DCTN1	Dynactin 1
DCS	Duplex Consensus Sequences
DCVs	DNA Content Variations

DDR	DNA Damage Repair
DLB	Dementia with Lewy Bodies
DN	Dystrophic Neurites
DNA	Deoxyribonucleic acid
DNMT	DNA Methyl Transferase
dNTPs	Deoxynucleotide Triphosphates
DS	Duplex Sequencing
EDTA	Ethylenediamine tetra acetic acid
ER-AT	End-Repair and A-Tailing (Second step of KAPA HyperPlus Protocol for duplex library preparation)
EtOH	Ethanol
EWS	Ewing Sarcoma protein
EWSR1	Ewing Sarcoma RNA Binding protein 1
F	Female
FB	Forward-biotin labelled primer
F-primer	Forward primer
FET	Protein family consisting of FUS, EWSR1, and TAF15 proteins
FISH	Fluorescence *in situ* Hybridization
FTD	Frontotemporal Dementia
FTD-ALS	Frontotemporal Dementia associated with Amyotrophic Lateral Sclerosis
FTD-FUS	Frontotemporal Dementia with FET-positive inclusions
FTD-MND	Frontotemporal Dementia associated with Motor Neuron Disease
FTD-Tau	Frontotemporal Dementia associated with Tau protein-positive inclusions
FTD-TDP17T	Frontotemporal Dementia and parkinsonism linked to chromosome 17
FTD-TDP43	Frontotemporal Dementia associated with TDP-43-positive inclusions
FTD-UPS	Frontotemporal Dementia with inclusions positive for ubiquitin proteasome system markers
FTD-3	Frontotemporal Dementia with Type 3 phenotype
FTLD	Frontotemporal Lobar Dementia
FUS	Fused in Sarcoma protein
FUS/TLS	Fused in Sarcoma/translocated in liposarcoma protein
G	Guanine
GATK	Genome Analysis Tool Kit
GBA	β-Glucocerebrosidase
gDNA	Genomic Deoxyribonucleic acid
GGT	Globular Glial Tauopathy
GRN	Progranulin gene encoding granulin protein
GS	Goldman Score
GWAS	Genome-wide Association Studies
G_4C_2	‘GGGGCC’ repeat
HD	Huntington’s Disease
HDAC	Histone Deacetylase
HS	High Sensitivity (Qubit DNA Assay and ScreenTapes)
HSP	Heat Shock Protein
HTT	Gene encoding for Huntingtin protein (associated with HD)
H3	Histone protein 3

H4	Histone protein 4
IDT probes	Custom-designed probes by Integrated DNA Technologies, US
IHC	Immunohistochemistry
Indels	Insertions and Deletions
iPSCs	Induced Pluripotent Stem Cells
ITM2B	Integral Membrane Protein 2B
K	Lysine (amino acid)
LINE1	Long Interspersed Nuclear Elements 1
lncRNAs	Long non-coding RNAs
LR-PCR	Long-Range Polymerase Chain Reaction
LRRK2	Leucine-Rich Repeat Kinase 2
M	Male
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MAO-B	Monoamine Oxidase B
MAPT	Microtubule-Associated Protein Tau gene
me	Methylation
Med	Medulla
MEG3	Maternally Expressed Gene 3
miRNAs	Micro RNAs
MND	Motor Neuron Disease
mRNA	Messenger Ribonucleic Acid
MS	Multiple Sclerosis
MSA	Multiple System Atrophy
MST	Multiple System Tauopathy with Dementia
mtDNA	Mitochondrial DNA
mtSNVs	Mitochondrial DNA SNVs
NaCl	Sodium salt
NADH	Nicotinamide Adenine Dinucleotide (NAD) + Hydrogen (H)
NaOH	Sodium hydroxide
nfvPPA	Non-fluent Variant Primary Progressive Aphasia
NCBI	National Centre for Biotechnology Information, US
NCI	Neuronal Cytoplasmic Inclusions
NCLs	Neuronal Ceroid Lipofuscinosis
ncRNA	Non-coding RNA
NDD	Neurodegenerative Diseases
ND1	Mitochondrially encoded NADH Dehydrogenase 1
NEAT1	Nuclear Enriched Abundant Transcript 1
NFT	Neurofibrillary Tangles
NGS	Next Generation Sequencing
NIBD	Neurofilament Inclusion Body Disease
NIFID	Neuronal Intermediate Filament Inclusion Disease
NII	Neuronal Intranuclear Inclusions
NORD	National Organization for Rare Disorders
Occ	Occipital lobe
OPTN	Optineurin
OXPHOS	Oxidative Phosphorylation
PAFAH1B1/ LIS1	Platelet-activating factor Acetylhydrolase 1B subunit alpha/ Lissencephaly-1 gene

PARK3	Parkinson disease 3 protein
PCD	Programmed Cell Death
PCR	Polymerase Chain Reaction
PCR-RFLP	Polymerase Chain Reaction–Restriction Fragment Length Polymorphism
PD	Parkinson's Disease
PEG	Polyethylene Glycol
PINK1	PTEN-induced kinase 1
piRNAs	PIWI-associated RNAs
PIWI	P-element induced Wimpy testis protein
PLS	Primary Lateral Sclerosis
PNFA	Primary Non-fluent Aphasia
PNS	Peripheral Nervous System
PRKN	Parkin
PSEN1	Presenilin 1
PSEN2	Presenilin 2
PTEN	Phosphatase and TENsin homolog deleted on chromosome 10 (tumour-suppressor gene)
PSP	Progressive Supranuclear Palsy
qPCR	Quantitative Polymerase Chain Reaction
RB	Reverse-biotin labelled primer
R-primer	Reverse Primer
RNA	Ribonucleic acid
RNA seq	RNA Sequencing
ROS	Reactive Oxygen Species
RT-PCR	Real Time–Polymerase Chain Reaction
S-phase	Synthesis phase (of the cell cycle)
SAM	Sequence Alignment Map
SCARB2	Scavenger Receptor Class B member 2
scATAC-seq	Single-cell Assay for Transposase-Accessible Chromatin Sequencing
scDNA seq	Single-cell DNA Sequencing
scRNA seq	Single-cell RNA Sequencing
sc-seq	Single-cell Sequencing
SD	Semantic Dementia
seq	Sequencing
sRNAs	Small RNAs
siRNAs	Silencing RNAs
SIRT	Sirtuin family proteins
SMN	Gene encoding Survivor Motor Neuron
SMRT-seq	Single-Molecule Real-Time Sequencing
SNCA	Gene encoding α-synuclein
SNPs	Single Nucleotide Polymorphism
SNVs	Single nucleotide variants
SOD1	Superoxide Dismutase 1 [Cu-Zn]
SOLiD SEQ	Small Oligonucleotide Ligation and Detection System of Sequencing
SPRI	Solid-Phase Reversible Immobilization
SPTBN1	Spectrin β chain 1
SQSTM1	Sequestosome 1

SSCS	Single-Strand Consensus Sequences
svPPA	Semantic Variant Primary Progressive Aphasia (also known as Semantic Dementia)
T	Thymine
TAE	Tris-acetate EDTA
TAF15	Gene encoding TATA-binding protein-associated Factor 15
TANK	Gene encoding TRAF family member-Associated NF-Kappa-B activator
TARDBP	Gene encoding TDP-43
TBP	TATA Box Binding Protein
TBK1	TANK-Binding Kinase 1
TDP-43	Transactive Response DNA Binding Protein 43kDa
TE	Transposable Element
Temp	Temporal lobe
TET	Wash-Tween buffer
TMEM106B	Transmembrane protein 106B
TNF	Tumour Necrosis Factor protein
TRAF	TNF Receptor-Associated Factor proteins
TREM2	Triggering receptor expressed on myeloid cells 2
TSC1	Tuberous sclerosis complex 1
TSC2	Tuberous sclerosis complex 2
UBQLN2	Ubiquilin 2
UCHL1	Ubiquitin carboxy-terminal hydrolase L1
UK	United Kingdom
UPS	Ubiquitin Proteasome System
US	United States
USD	United States Dollars
UTRs	Untranslated regions
VCP	Valosin Containing Protein
V(D)J	Variability, Diversity, and Joining recombination (Mechanism of somatic recombination in developing lymphocytes)
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WRN	Gene encoding for Werner syndrome ATP-dependent helicase
WS	Werner's Syndrome
XP	Xeroderma Pigmentosum

Symbol	**Units of Measurement**
~	approximately
∞	Infinite hold
amol	Attomole (1×10^{-18} moles)
bp	base pairs
C_q	Quantitation threshold cycle number (for qPCR)
Da	Daltons
^0C	degrees Celsius
dC_q	Difference between two threshold cycle numbers
ddC_q	base-2 logarithmic expression fold change
fmol	Femtomole (1×10^{-15} moles)

× *g*	Relative centrifugal force
h/hrs	Hour(s)
kb	kilobases
kDa	Kilo Daltons
Mb	megabases
µg	microgram
µm	micrometre
µM	micromolar
µl	microlitres
mg	milligram
ml	millilitres
mM	millimolar
mmol	millimoles
min	minutes
M	molar
ng	nanogram
nM	nanomolar
nmol	nanomoles
pg	picogram
rpm	Rotations per minute
sec	seconds
(v/v)	Volume-by-volume

Chapter I: Introduction

1.1 The human nervous system

The human nervous system broadly consists of two parts – the central nervous system (CNS) and the peripheral nervous system (PNS).

The CNS is a well-organized structure of the brain and spinal cord, the former of which is in the cranial skull, and is broadly divided into the cerebrum, the brainstem, and the cerebellum that are made by ~80 billion neurons (Jan H Lui, David V Hansen 2011), which are the primary functional units of the nervous system. The spinal cord extends from the brainstem into the vertebral column, and mainly transmits signals between the brain and the body through the sensory and motor neurons. The PNS consists of nerves enclosed in bundles of long neuronal processes called the axons and is responsible for connecting the CNS to all the other parts of the body.

Two main types of cells compose the nervous system – the neurons and the glial cells.

Neurons are the primary units of function in that they process and transmit information across the nervous system via electric/electrochemical signals. Heterogeneous and highly interconnected to form complex networks, the typical neuron is made up of a soma, dendrite(s), and an axon, and may be classified as an *afferent* or *sensory* neuron, *efferent* or *motor* neuron, or an *interneuron* based on its functions. The neurons communicate either through *excitatory* or *inhibitory* synapses, in which the action potential increases or decreases in the post-synaptic cell during transmission respectively (Purves D, Augustine G J 2001). This excitatory/inhibitory nature can strengthen or weaken over time in response to an increase/decrease in synaptic activity, a property known as *synaptic plasticity* (Byrne J H 1997), which is essential to form the vastly inter-connected and dynamic neural circuit of the brain.

Glial cells, on the other hand, are non-neuronal cells of two main sub-types: *macroglia* – of which astrocytes, oligodendrocytes, ependymal cells, and radial glia are found in the CNS that provide support, nutrition, and electrical insulation of neurons (via the myelin sheath) – and the *microglia* which are heavily involved in providing immunity within the system (Allen and Barres 2009).

This wide array of cells in the human brain work in tandem to perform many complex tasks such as sensation, perception, cognition, memory, movement, communication, and homeostasis.

1.2 Neurodegenerative Diseases

Neurodegenerative diseases (NDD) arise due to the progressive loss of structure and/or function of the nervous system associated with age that ultimately leads to cell death, resulting in common symptoms such as memory loss, anxiety, changes in mood, aggression, disinhibition, agitation, etc. (Beart et al. 2017). Since these diseases involve nearly irreversible neurodegeneration of the nervous system at multiple levels of neuronal circuitry, ranging from molecular to systemic, they are also considered to be incurable at present. The estimate for the global number of patients with NDD is in the millions and this number is increasing by the rate of 10 million each year (World Health Organization 2008), thus building up the demand for treatment, rehabilitation, and other support services, costing governments in trillions of USD in medical healthcare worldwide (T. Miller 2017). The rising frequency of cases, the costs involved, and the lack of cures or an ability to successfully ameliorate disease progression has made it imperative to understand the origin and mechanisms of neurodegenerative pathology.

1.2.1 Pathology of neurodegeneration

NDD are characterized by the progressive loss of selectively vulnerable neuronal populations, also referred to as **neurodegeneration**, and can be broadly classified with respect to (Dugger and Dickson 2017):

1. The *primary clinical features*, e.g., dementia, parkinsonism, motor neuron disease (MND), etc.,
2. *Anatomic distribution of neurodegeneration*, e.g., frontotemporal degenerations, extrapyramidal disorders, spinocerebellar degenerations, etc., or,
3. *Principal molecular abnormality* resulting in tauopathies, α-synucleinopathies, TDP-43 proteopathies, etc. as determined at the time of autopsy.

In addition to these specific protein accumulations and anatomical vulnerabilities, NDD share many fundamental processes associated with progressive neuronal dysfunction and death, such as proteotoxic stress due to abnormalities in the ubiquitin-proteasomal as well as autophagosomal/lysosomal systems, oxidative stress, programmed cell death, and neuroinflammation (Dugger and Dickson 2017). There is also growing evidence that abnormal

protein conformers could spread from cell to cell along anatomically connected pathways via "prion-like" mechanisms (Mudher et al. 2017; Jaunmuktane and Brandner 2020; Brás and Outeiro 2021), which shows that elucidating the mechanism of neuropathology is critical to our understanding of NDD.

1.2.2 Risk factors for neuropathogenesis

Accumulation of stochastic DNA damage due to defective repair mechanisms and the loss of genomic maintenance with time causally contributes to **ageing**, a major risk factor for NDD (Maynard et al. 2015). NDD usually present with late-onset symptoms, although the average age of onset differs with respect to the neuropathology. The two major contributing factors that could influence the onset and severity of NDD are: (1) *oxidative stress* – the main systemic condition that aggravates neurodegeneration through impaired modulation of the functioning of various biomolecules brought about by the presence of increased reactive oxygen species (ROS) and subsequent oxidation/reduction imbalance that can induce mutations capable of disrupting or altering genes such that one or more cell death processes are activated, eventually leading to neurodegeneration **(Fig. 1)** (Hardy 2005; Singh et al. 2019) and (2) *inflammation* – the impaired and prolonged activation of macrophages and microglia with age that further induces ROS production, often accompanied by impaired phagocytosis resulting in increased toxic protein production and formation of inclusion bodies within neurons that is commonly associated with neurodegeneration (Stephenson et al. 2018).

It must be noted that while protein aggregation into insoluble inclusions is a pathogenic feature in most NDD, the process of protein aggregation is not always pathogenic and in fact, there is growing evidence that the disruption of intracellular protein homeostasis occurs naturally with age leading to widespread protein insolubility and aggregation with no negative consequences and a marked absence of a disease state, possibly since these aggregates are non-toxic and are more easily cleared by intracellular mechanisms than their disease-state counterparts [reviewed in (David 2012)]. The current prevalent theory for the pathogenesis of NDD is the inability of brain cells to clear these toxic aggregates in response to a breakdown of protein degradation mechanisms due to genetic or epigenetic mutations with a particular "disease" signature that is also unexplainably transmissible by some type of "cell-to-cell" spread in older patients (Davis and Stroud 2013).

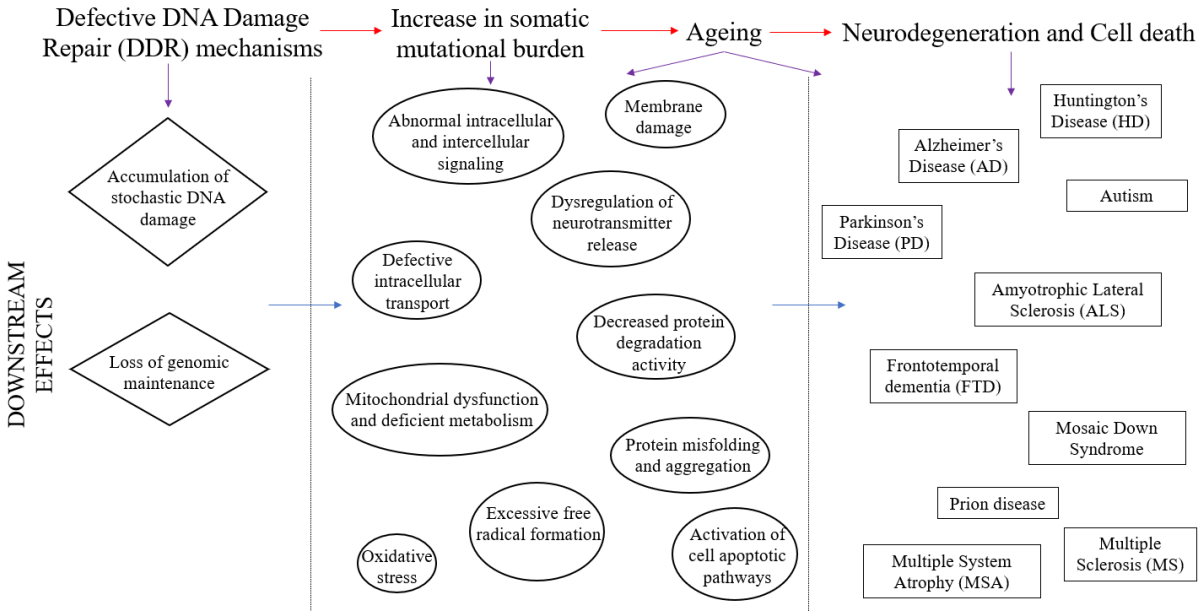


Figure 1: Association of Mutational Burden with Ageing and Neurodegenerative Diseases. Defective DNA damage repair mechanisms lead to an increase in somatic mutational burden which could lead to ageing and sometimes, neurodegeneration and cell death, often resulting in the presentation of clinical symptoms of a neurodegenerative disease.

On the other hand, NDD in children is very rare wherein the symptoms escalate quickly, the disease is often misdiagnosed, and typically leads to an early death. Paediatric NDD are mostly familial with an autosomal-recessive pattern of inheritance or sporadic (Radke, Stenzel, and Goebel 2018). While the underlying mechanisms of neurodegeneration in childhood is not fully understood, multiple causes have been proposed, including genetic, metabolic, viral, immunopathic, environmental, and epileptogenic (Dyken and Krawiecki 1983). Like adult-onset NDD, paediatric NDD may be brought about due to impairments in the cell clearance pathways such as lysosomal degradation pathways, but also other factors such as defects in the biogenesis of peroxisomes, single peroxisomal enzyme (transporter) deficiencies, congenital disorders of glycosylation, leukodystrophies, leukoencephalopathies, etc. (Radke, Stenzel, and Goebel 2018)

Some of the most common paediatric NDD in terms of prevalence are (Dyken and Krawiecki 1983):

1. Subacute Sclerosing Panencephalitis: Slow, but persistent, degradation of the CNS in children and young adults with a very high mortality rate due to viral infections caused by defective measles virus and no known cure at present (Rocke and Belyayeva 2022);
2. Neuronal Ceroid Lipofuscinosis (NCLs): A group of autosomal-recessive lysosomal storage neurometabolic disorders associated with lysosomal enzyme defects/lysosomal

soluble protein defects due to mutations in most commonly in CLN (ceroid lipofuscinosis, neuronal) genes that leads to intracellular accumulation of lipopigments and progressive neurodegeneration (Radke, Stenzel, and Goebel 2018);

3. Tuberous Sclerosis with Degeneration: An autosomal-dominant disorder resulting in epilepsy, developmental delay, and autism due to inactivating mutations with incomplete penetrance in genes encoding the proteins hamartin (Tuberous sclerosis complex 1, TSC1) or tuberin (TSC2) leading to dysfunctional TSC heterodimer complex; life expectancy is normal for many (Reith et al. 2011);
4. West Disease: A type of Early Epileptic Encephalopathy, West disease is characterized by severe drug-resistant epileptic disorders, with an early-life onset accompanied by cerebral dysfunction progression caused by mutations in several genes including, but not limited to, Aristaless Related Homeobox (ARX, Xp21.3), Cyclin-dependent kinase-like 5 (CDKL5, Xp22.13), Platelet-activating factor Acetylhydrolase 1B subunit alpha/ Lissencephaly-1 (PAFAH1B1/LIS1, 17p13.3) etc. (Pavone et al. 2020);
5. Werdnig-Hoffmann Disease: A spinal muscular atrophy wherein the anterior horn of the spinal cord experiences cellular degeneration *in utero* as well as after birth with progressive loss of motor neurons in the spinal cord and brainstem caused by autosomal-recessive mutations in survivor motor neuron (SMN) gene; no known cure at present (Emmady and Bodle 2022); and,
6. Hereditary Spastic Paraplegia (HSP): A group of inherited diseases (autosomal-dominant, autosomal-recessive, or X-linked recessive) due to mutations in more than seventy genes presenting with progressive stiffness, or spasticity, and contraction in the lower limbs; HSP may set in at any age from infancy to adulthood with more than 25% of patients being asymptomatic (McDermott et al. 2000).

Therefore, the process of neurodegeneration often starts before there is functional evidence of disease and the characteristics of this process can be distinct in several neurodegenerative biochemical pathways which makes the development and validation of specific biomarkers critical for NDD diagnosis and management.

1.2.3 Cellular mechanisms of neuropathogenesis

1.2.3.1 Neuronal mechanisms

Breakdown of the biochemical processes in neurons often leads to the systemic failure associated with many NDD. The origins of this breakdown can be attributed to:

1. **Genetics**

Genetic mutations, in both commonly associated genes as well as risk-modifying variants, can lead to NDD.

For instance, mutations in the MAPT gene that codes for the protein Tau can lead to NDD such as Alzheimer's disease (AD), behavioural-variant Frontotemporal Dementia (bvFTD), etc., while in NDD such as Huntington's Disease (HD) and spinocerebellar ataxias, a trinucleotide repeat of CAG results in a polyglutamine tract that is often toxic to the neuron and causes neurodegeneration with a dominant pathogenesis are commonly reported (Thompson 2008).

Moreover, genome-wide association studies (GWAS) have identified certain genetic markers, or loci, within genetic regions that increase risk of developing a particular phenotype with small to moderate effect size, i.e., risk-modifying variants, with most reported associations being intronic or intergenic which are most likely to affect the DNA structure and gene expression rather than the sequence of affected proteins (Manolio et al. 2009), e.g., two single nucleotide polymorphisms (SNPs) encompassing the transmembrane protein 106B (TMEM106B) gene (7p21.3) – rs1020004 and rs1990622 – are known risk factors that increase the susceptibility of GRN mutation carriers with FTD with TDP-43 immunoreactive pathology (FTD-TDP43) by behaving as GRN-modifiers via the modulation of progranulin protein levels in some familial and sporadic cases of FTD (Deerlin et al. 2010; Raffaele, Claudia, and John 2019).

Epigenetic variations that can cause the dysregulation of critical genes can also lead to NDD as in the cases of histone hypoacetylation in Amyotrophic Lateral Sclerosis (ALS) (Rouaux et al. 2003), global DNA hypomethylation and gene-specific hypermethylation seen in some AD patients (Bakulski et al. 2012), and decreased

histone acetylation and increased histone methylation due to CAG repeat expansion in HD (Ryu et al. 2006).

2. **Intracellular mechanisms**

NDD are frequently attributed to the inability of a cell to remove troublesome proteins or organelles due to malfunction(s) in the ubiquitin-proteasome and/or autophagy-lysosome pathways, resulting in the aggregation of misfolded proteins. Hence, NDD are also very often forms of *proteopathies*; some common examples include α-synuclein aggregation seen in Parkinson's Disease (PD), Dementia with Lewy bodies (DLB), and Multiple System Atrophy (MSA); Tau aggregation seen in AD and bvFTD; and, amyloid-β plaques seen in AD (Rubinsztein 2006).

Moreover, neurons typically have higher mitochondrial copy numbers and are therefore, more susceptible to *mitochondrial dysfunction* and oxidative stress that have been known to cause *DNA damage* that contribute to the pathogenesis of NDD such as AD, PD, HD, and ALS (M. T. Lin and Beal 2006).

3. **Programmed cell death (PCD)**

PCD could be activated by apoptotic, autophagic, or cytoplasmic pathways in several NDD like ***PD***, likely non-pathogenic activation of caspases 3, 8, and 9 in experimental models; ***ALS***, activation of PCD machinery as an indirect effect of mutant Superoxide dismutase 1 (SOD1) in inherited patient cases; ***AD***, activation of caspases 3, 8, and 9 in the hippocampus of patients; and, ***HD***, activation of caspases 1 and 3 in experimental models as a directly pathogenic or injury-related response (Vila and Przedborski 2003).

1.2.3.2 Non-neuronal mechanisms

Recent advances have shown that glial cells, specifically the microglia and the astroglia, contribute to NDD progression (Stevenson et al. 2020).

- **AD** is characterized by the build-up of amyloid-β plaques and tau deposits in neurons; however, some of the genes associated with these build-ups such as Apolipoprotein E (APOE), Apolipoprotein J (APOJ), etc., are mainly expressed in astrocytes, microglia, and oligodendrocytes (Arranz and De Strooper 2019). The failed clearance of amyloid-

β from the brain is also a prominent hypothesis in disease pathogenesis and several glial cells are implicated in the breakdown of the blood-brain barrier (BBB) (Tarasoff-Conway et al. 2015).

- Malfunctioning astrocytes have been hypothesised to play a critical role in amplifying **PD** pathology by accumulating and spreading α-synuclein to neighbouring dopaminergic neurons and hastening the neurodegeneration process (di Domenico et al. 2019). Loss-of-function mutations in astrocytes and microglia have also been proposed to impact disease progression by failing to clear α-synuclein inclusions in the brain (Filippini, Gennarelli, and Russo 2019).
- Dysfunctional glial cells have been demonstrated to play an important role in **HD** progression: mutant huntingtin protein (HTT) expressed in human progenitor cells engrafted in healthy mice showed a development of disease phenotypes while introduction of healthy human glial cells in transgenic HD mice increased survival rates and reduced motor deterioration (Benraiss et al. 2016).
- The role of glial cells in **FTD** is not fully understood; however, a meta-analysis of FTD cases revealed that astrocytic apoptosis increased with disease progression while neuronal apoptosis remained constant (Broe, Kril, and Halliday 2004). Moreover, autopsy samples of patients revealed high levels of reactive astrocytes with signs of degradation that would be likely to lead to neuronal apoptosis (Su et al. 2000).
- Impairments in astrocytes are thought to contribute to **ALS** progression via a loss-of-function mechanism wherein abnormal functioning of glutamate transporters and extracellular potassium ion concentration leads to the presence of extracellular glutamate and excitotoxicity that contribute to hyperexcitability in patients (Do-Ha, Buskila, and Ooi 2018).

Very often, more than one of the above cellular dysfunctions are triggered to occur simultaneously in a multitude of ways with the involvement of neurons and glial cells in complex feedback loops that perpetuate dysfunction and contribute to NDD progression (Gleichman and Carmichael 2020), which makes it very difficult to pin-point the appropriate target(s) to halt or reverse the neurodegeneration process in several brain cells.

1.2.4 Presentation of neuropathology: common neuronal disorders

The breakdown of neuronal biochemical processes can lead to specific pathologies that contribute to hallmark clinical phenotypes of certain NDD. Since symptoms are only seen when

the neurodegeneration is well-established within a cell as a result of the failure of multiple intracellular events, an overlap of clinical phenotypes of some common neurodegeneration disorders can also be seen.

A brief overview of the characteristics of the five most common and well-known NDD sheds further light into this phenomenon:

a) Alzheimer's disease (AD)

AD is characterized by the presence of amyloid plaques and neurofibrillary tangles resulting in the loss of neurons and synapses in the cerebral cortex that leads to a well-documented progressive pattern of cognition, particularly the progressive loss of episodic memory, and functional impairment until eventual death (Vinicius et al. 2019).

Prevalence: AD is the most common NDD, affecting up to 50 million people worldwide (Breijyeh and Karaman 2020). Familial (early-onset) AD is expressed as a dominant trait and accounts for less than 5% of AD cases while the remaining are sporadic (late-onset) AD cases (Barber 2012).

Pathogenesis: The exact mechanism of AD proteopathy has not yet been elucidated, although mutations present in the genes of amyloid-β precursor protein (APP), presenilin proteins (PSEN1 and PSEN2), or microtubule-associated protein tau (MAPT); reduced acetylcholine neurotransmitters; neuroinflammation; and, cell-signalling disruptions are the most common hypotheses, indicating that the disease is multifactorial.

Reports of DNA demethylation in the promoter region of the APP gene that lead to amyloid-β deposits in some AD cases; loss of function epimutations in of all three classes of histone deacetylases (HDACs) – I (HDAC 2 and 3), II (HDAC 4 and 6), and III (sirtuins, SIRT1-7) – can lead to lowered synaptic plasticity and memory formation (due to epimutations in Classes I & II) and increased tau aggregation (due to epimutations in Class III); and, the dysregulation of non-coding RNA such as circular RNA, micro RNA (miRNA), and messenger RNA (mRNA) in the hippocampus of some AD patients all suggest that epigenetic alterations may also have a role to play in the development of AD (X. Liu, Jiao, and Shen 2018).

Current therapeutic practices: Several therapeutic techniques have been in use, ranging from targeting acetylcholine esterase inhibitors as well as lowering amyloid-β and tau protein levels to attempting antioxidant treatments to varying degrees of success; however, multidisciplinary approaches with more effective links between basic and clinical research as well as pharmaceutical institutes may show greater promise in either halting or reversing the course of AD progression (Huang and Mucke 2012).

b) Parkinson's disease (PD)

PD is characterized by the presence of symptoms such as bradykinesia, rigidity, resting tremor, and posture instability (Lau and Breteler 2006). It is one of the primary forms of α-synucleinopathy syndromes, in addition to Dementia with Lewy Bodies (DLB) and Multiple System Atrophy (MSA).

Prevalence: PD is the second-most common NDD that affects the motor nerves of the CNS affecting about ~7-10 million patients world-wide (Ray Dorsey et al. 2018; Jankovic and Tan 2020). Most of the cases of PD are sporadic, but ~10-15% are familial cases (Papapetropoulos et al. 2007).

Pathogenesis: The disease occurs as a result of a dopamine deficit due to the death of neurons in the substantia nigra, although the cause of this neuronal death is still unknown – inherited and environmental factors have been proposed to play a role in PD development, with the age of onset normally being over the age of 60 (Jankovic and Tan 2020). Mutations in over nineteen genes have been implicated in PD, including, but not limited to, α-synuclein (SNCA), Parkinson disease 3 protein (PARK3), Ubiquitin carboxy-terminal hydrolase L1 (UCHL1), Parkin (PRKN), PTEN-induced kinase 1 (PINK1), and a 22q11 deletion (Deng, Wang, and Jankovic 2018).

A recent study of ~500 PD patients *versus* controls found an association of PD onset with increased epigenetic mutation load in that five genes most-associated with the disease were found to be enriched with methylation dysregulation epimutations that may also be associated with the progression of motor symptoms as well as time of death in PD patients, further highlighting the complexity of the disease (Chen et al. 2022).

Current therapeutic practices: Initial treatments include medications with levodopa, Monoamine oxidase-B (MAO-B) inhibitors, or dopamine agonists, while the more advanced interventions include surgery for deep brain stimulations (Armstrong and Okun 2020). There is no known cure for PD.

c) Dementia with Lewy Bodies (DLB)

DLB is a common NDD associated with the abnormal cytoplasmic aggregation of α-synuclein protein, or Lewy bodies, that can affect cognition, movement, behaviour, and mood with symptoms including, but not limited to, hallucinations, confusion, impairments in understanding and judgment, confusion, sleep disorders, unsteadiness, slow movements, fainting spells, etc. (McKeith 2004). It is another one of the primary forms of α-synucleinopathy syndromes, in addition to PD and MSA.

Prevalence: The global prevalence of DLB in routine clinical practice is unclear; however, it is estimated that 10-15% of dementia cases are that of DLB (McKeith 2004). DLB usually occurs sporadically in patients over the age of fifty, although familial inheritance has been reported in ~10% of the cases (Heerema 2020).

Pathogenesis: The exact cause of pathogenesis is still unknown and a "trigger" for the aggregation of α-synuclein in the CNS remains unidentified (Linard et al. 2022), although there is an overlap of risk factors amongst four NDD – DLB, AD, PD, and Parkinson's disease dementia – that are exhibited as mutations in the genes such as APOE, β-Glucocerebrosidase (GBA), α-synuclein (SNCA), and Leucine-rich Repeat Kinase 2 (LRRK2), and more rarely, in Scavenger Receptor Class B member 2 (SCARB2), Spectrin β chain 1 (SPTBN1), Contactin-1 (CNTN1), Presenilin proteins (PSEN1 & 2), Microtubule-associated protein Tau (MAPT), and Granulin (GRN) (Hansen et al. 2019), all of which can complicate diagnosis.

There are also a few studies that have reported hypomethylation of the SNCA CpG islands in DLB patients compared to controls as well as evidence of α-synuclein protein masking histone protein 3 (H3) acetylation sites in association with SIRT-3 proteins; LRRK2 promoting the deacetylation of H4 and H5 via HDAC3 dysregulation; etc., that suggest that several complex pathogenic mechanisms may be involved in the development of DLB [reviewed in (Urbizu and Beyer 2020)].

Current therapeutic practices: There is no known cure for DLB and current treatments include interventions to control the cognitive, psychiatric, and motor dysfunctions associated with the NDD (Capouch, Farlow, and Brosch 2018). Acetylcholinesterase inhibitors such as Donepezil and Rivastigmine mainly treat cognitive impairments, but also may help ameliorate the severity of psychiatric and motor dysfunctions to a certain extent (Hershey and Coleman-Jackson 2019) while anticonvulsants such as Zonisamide can significantly improve motor symptoms without worsening neuropsychiatric symptoms unlike Levodopa (Watts et al. 2022). Future pharmacologics to specifically treat DLB include clinical trials to test novel drugs serving as serotonin agonists, dopamine receptor potentiators, phosphodiesterase inhibitors, etc, as well as research into whether stem cell therapy could serve as a treatment modality for this NDD (Capouch, Farlow, and Brosch 2018) .

d) Huntington's disease (HD)

HD is a rare autosomal dominant NDD with an average age of onset between 30-50 years with severe uncoordinated involuntary movements, difficulty in speech, and dementia becoming apparent as the disease advances (R. A. C. Roos 2010).

Prevalence: HD is more common in those of European descent (4-15 per 100,000) than in Asian populations while the rate of occurrence in African populations is unknown (S. Frank 2014). HD is predominantly inherited whereas only ~8-15% of the cases appear to be sporadic (Mariani et al. 2016).

Pathogenesis: Mutations in the huntingtin gene (HTT) as well as CAG trinucleotide repeat expansion are typical causes for HD – huntingtin is normally an antiapoptotic agent, and in its mutated form, often activates caspases for apoptosis (Cattaneo, Zuccato, and Tartari 2005) whereas the CAG repeat expansion produces a polyglutamine tract that is highly toxic to the neuron (Illarioshkin et al. 2018).

Furthermore, transcriptional dysregulation via hypermethylation of the HTT gene and several transcriptional regulators as well as the identification of altered patterns of several chromatin marks at the genome-wide level in human HD tissues and various models of HD, including DNA methylation, histone H3K9 and H3K14 acetylation (H3K9ac and H3K14ac), H3K27ac, H3K4me3, H3K9me3, and H3K27me3, indicating that mutant HTT appears to modulate chromatin structure via several epimutations;

however, the mechanisms linking mutant HTT and altered epigenetic modifications are still poorly understood [reviewed in (Hyeon, Kim, and Yano 2022)].

Current therapeutic practices: Current therapeutic interventions have been limited to palliative care via physical therapy, speech therapy, and medications such as Tetrabenazine to reduce uncoordinated movements (chorea) or antipsychotics to treat chorea and control cognitive problems, with no known treatments to halt or reverse disease progression; however, dosage reduction strategies to modify disease progression such as CAG repeat expansion antisense oligonucleotides and molecules directed at transcription dysregulation have shown some encouraging results (R. A. C. Roos 2010; Dickey and Spada 2018).

e) Amyotrophic Lateral Sclerosis (ALS)

ALS is an NDD in which upper and lower motor neurons that control voluntary muscles show selective neurodegeneration (Min et al. 2020). Patients with ALS often exhibit early symptoms such as stiff muscles, muscle twitching, weakness and muscle wasting that eventually progress to loss of the ability to eat, speak, move, and breathe, leading to paralysis and early death (within 2-4 years after age of onset on average) (Masrori and Van Damme 2020).

Prevalence: ALS affects 2-3 patients per 100,000 of European descent (Hardiman et al. 2017). 90-95% of ALS cases are sporadic with both, familial and sporadic cases associated with the degeneration of cortical and spinal motor neurons (Zarei et al. 2015).

Pathogenesis: Several genes have been implicated in ALS, including, but not limited to, Superoxide dismutase 1 (SOD1), RNA-binding protein (FUS) gene, TAR DNA-binding protein 43 (TARDBP), and the Chromosome 9 open reading frame 72 (C9orf72) repeat expansion (Corcia et al. 2017). Most of these genes have also been implicated in frontotemporal dementia (FTD) which could explain how ~15% of ALS patients also develop FTD symptoms as well as how ~12.5% of FTD patients also develop ALS (~40% develop a motor neuron disease) – both are now considered to be a part of the same disease spectrum (Couratier et al. 2017). SOD1 is normally an antioxidant enzyme which is mutated in ALS and could be a potential source for DNA damage while mutations in TARDBP, FUS, and C9orf72 repeats are known to cause

protein aggregation and toxicity; however, the pathogenic mechanisms have not yet been solved (Min et al. 2020). There is also evidence that SOD1 mutations in astrocytes can release factors selectively toxic to motor neurons, a significant finding in that it shows non-neuronal cells could also play a role in neurodegeneration (Nagai et al. 2007).

Moreover, a recent blood-based epigenome-wide study meta-analysis identified 45 differentially methylated positions in 6,763 patients compared to controls that were associated with metabolism, cholesterol biosynthesis and immunity, and when this data was combined with their GWAS, the analyses revealed that cholesterol biosynthesis was potentially causally related to ALS (Hop et al. 2022).

Current therapeutic practices: There is no known cure for ALS at present, and current treatments only work towards improving symptoms and increasing the lifespan by 2-3 months such as the use of medications like Riluzole, or non-invasive and/or mechanical ventilation (R. G. Miller et al. 2007; Hobson and McDermott 2016).

Evidently, the neuropathogenesis of even the most of the common NDD have not yet been resolved. Examining the association of DNA sequence alterations (genetic biomarkers) and disease to discern the underlying mechanisms for disease progression could provide insights to creating novel and effective drugs which, in conjunction with early diagnosis using well-defined biofluid and molecular imaging markers, could help mitigate the symptoms entirely (Beart et al. 2017). Moreover, in some cases, a major player involved in neuropathogenesis are the same – e.g., tau protein gene in some diseases such as AD and FTD or α-synuclein in PD and DLB – or they may be different – e.g., the involvement of CLN genes in NCL or the SMN genes in Werdnig-Hoffman Disease – which affect the same/different pathways in same/different parts of the cell that determines the rate of disease development and its severity. This heterogeneity in the pathogenesis and severity of neuronal dysfunction and degeneration exhibited in familial as well as sporadic forms of a disease may be explained by further exploring heterogeneity in the central nervous system.

1.3 Somatic Brain Mosaicism

Brain cells display great heterogeneity with respect to their morphology, function, connectivity, as well as electrophysiological activity (in neurons). Recent advances in technology have made it possible to explore this diversity at the molecular level to an

unprecedented level – single-cell sequencing and expression profile studies show that there are several subtypes within the same brain regions that not only differ at the physical and cellular levels but also differ at the genomic level with respect to DNA sequence (Lake et al. 2016; Manno et al. 2016; Verheijen, Vermulst, and van Leeuwen 2018). This is thought to occur due to errors in the fidelity of DNA replication and repair during neurodevelopment, which may then bring about somatic mutations in neuronal stem cells or progenitor cells that are subsequently propagated throughout their lineages, resulting in a patchwork of neurons with different genetic make-up, a phenomenon called *somatic brain mosaicism* (Verheijen, Vermulst, and van Leeuwen 2018) (**Fig. 2**).

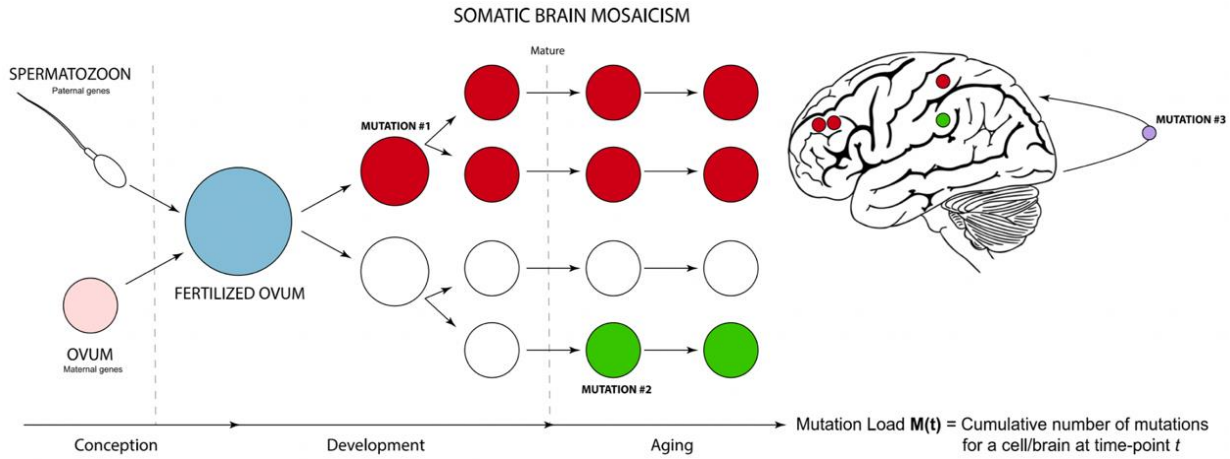


Figure 2: Somatic brain mosaicism.

Somatic mutations may be early-arising, in which case more cells across multiple organs are affected, or they may be late-arising, in which case only a few cells within certain organs are affected, during the lifetime of an organism. In the brain, somatic mutations may arise during neurodevelopment as post-zygotic somatic mosaicism which affects more areas of the brain (Mutation #1); may arise in post-mitotic neurons which affects a smaller area of the brain (Mutation #2); or, may arise during adult neurogenesis (Mutation #3) as a result of various stresses such as genomic alterations, impaired nuclear integrity, aberrant nucleocytoplasmic transport, mitochondrial defects, etc., that contribute to neuronal damage leading to somatic brain mosaicism. Adapted and modified from (Verheijen, Vermulst, and van Leeuwen 2018).

These somatic mutations could affect genomic integrity and result in gene expression changes during development that impact neuronal structure and function, altering neural circuit development and plasticity which can cause abnormal information processing leading to sporadic NDD (Nicolas and Veltman 2018).

Somatic mosaicism can present as various symptoms at varying degrees of severity depending on when they arise during development – mosaicism that occurs earlier along neurodevelopment in neuronal progenitors is propagated throughout the lineage and affects larger areas of the nervous system, while mosaicism in post-mitotic neurons results in very fine changes in a much smaller region of nervous system, thus presenting in a range from severe disease state to asymptomatic, respectively.

1.3.1 Types of somatic mutations

Depending on when they occur during the lifetime of a neuron, several types of somatic mutations can arise over time. The characteristics of each type of mutation is listed in (**Table 1**).

Type of mutation	Definition	Size and avg. prevalence within brain	Relevant details and References
DNA Content variations (DCVs)	Any changes in the amount of DNA across the genome, ranging from alterations in DNA sequence to chromosomal structure	Average increase ~ 250 Mb	Can vary with neuroanatomical location (Rohrback et al. 2018)
Aneuploidy	Gain and/or loss of chromosomes from a euploid complement		Neuronal hypoploidy is six times more common than hyperploidy due to a negative pressure during differentiation that prefers chromosomal loss (Rohrback et al. 2018)
Aneusomy	Copy number change in a single, defined chromosome without complete knowledge of the state of the remaining chromosomes		Can survive to adulthood and contribute to functional diversity of the neural circuitry (Rohrback et al. 2018)
Copy Number Variations (CNVs)	Variable number of repeats of genome sections of an individual	Range varies from 9% to all neurons; 2-10 Mb	Deletions usually outnumber amplifications (Rohrback et al. 2018)
Transposable/ retrotransposable elements	Transposons (TE): DNA sequence that can change its position within a genome, possibly creating or reversing mutations, often as duplication of the same genetic material, and altering the genetic identity of a cell and genome size without any RNA intermediates Retrotransposition: DNA sequences are first transcribed to RNA and then reverse transcribed to DNA that are re-inserted into the genome at target sites	< 0.6 to ~14 per genome de novo presence	Ubiquitously found in all prokaryotic and eukaryotic organisms, TEs account for the majority of cis-regulatory DNA in humans and are instrumental in the process of evolution. They are also involved in epigenetic regulation of neuronal differentiation. Mosaic Long Interspersed Nuclear Elements (LINE1) contribute to neuronal diversity through somatic retrotransposition. (Pray 2008; Bourque et al. 2018; Mustafin and Khusnutdinova 2020; Wells and Feshotte 2020)

Single nucleotide variants (SNVs)	Synonymous/non-synonymous variations in a single nucleotide that occurs at a specific position in an individual's genome	In the range of thousands	SNVs appear to accumulate with age; a healthy human brain is estimated to contain ~80% of C > T SNVs out of all mutations on average (Rohrback et al. 2018)
Insertions and deletions (Indels)	Single/multiple base insertions or deletions to the genome, often resulting in frameshift mutations (unless a multiple of 3) that produce truncated/altered proteins when present in coding/regulatory regions of the genome	In the range of hundreds-thousands	Frameshift mutations accumulate with age and are major biomarkers for various NDD (Evans, Burbach, and Leeuwen 1995)
Epimutations	Changes to functional protein due to non-genetic factors such as errors during transcription, splicing (in eukaryotes), translation, or post-translational modifications; or, due to changes in the epigenetic markers of the genome that could result in changes in gene expression, thus affecting neuronal function	In the range of hundreds-thousands	(Saxowsky and Doetsch 2006; Hugo, Krijger, and Laat 2016; Verheijen, Vermulst, and van Leeuwen 2018)
Proteins that can initiate mutations	A hypothetical model called "seeding" describes the prion-like spreading of aberrant pathogenic proteins from one neuron to another induced by single DNA mutations in post-mitotic neurons as a possible explanation for the widespread pathology of AD		(Frost B. 2010; Verheijen, Vermulst, and van Leeuwen 2018)

Table 1: List of different known types of age-related somatic mutations.

1.3.2 Mechanisms for somatic mosaicism

Mechanisms for the creation of somatic mutations are not yet fully understood: replication and DNA-proof-reading errors occurring in the S-phase are the best-known sources of somatic mutations during neurodevelopment, while other exogenous and endogenous factors are thought to cause their accumulation in postmitotic neurons over time (Verheijen, Vermulst, and van Leeuwen 2018).

Exogenous factors are any environmental factors that induce mutations – electromagnetic radiations; mutagenic chemicals; and, viruses and other immune reactions that can result in inflammation (Verheijen, Vermulst, and van Leeuwen 2018); whereas **endogenous factors** are

spontaneous events that give rise to somatic mutations over time via multiple mechanisms – DNA damage, defective DNA damage repair (DDR), errors in DNA replication, transcription, and metabolic processes (Evans, Burbach, and Leeuwen 1995; Verheijen, Vermulst, and van Leeuwen 2018), along with those that arise due to unexpected events: mutagenesis due to the action of proteins belonging to the V(D)J site-specific recombination machinery of the immune system (Chun et al. 1991), changes in any protein synthesis machinery, etc. (Verheijen, Vermulst, and van Leeuwen 2018).

While the exact mechanisms through which these somatic mutations arise are still unknown, a few theories have been proposed:

i. *Replication slippage* occurs during DNA replication involving the denaturation, polymerase pausing, disassociation of polymerase from the DNA, and subsequent displacement of strands resulting in the mispairing of the complementary bases that can lead to a trinucleotide or dinucleotide expansion, or contraction (Hartl and Ruvolo 2012). Slippage occurs within repeat sequences of both prokaryotic and eurkaryotic genomes, and is thought to explain microsatellite instability (Viguera, Canceill, and Ehrlich 2001);
ii. *Deamination of cytosine and 5-methylcytosine*, to generate uracil and thymine respectively, is believed to be responsible for spontaneous mutations and is recognized as a frequent cause of SNVs that can contribute to human disease (Lewis et al. 2016); and,
iii. *Polymerase stuttering* is the process by which a polymerase transcribes a nucleotide several times without progressing further on the mRNA chain, involving the formation of secondary structures such as hairpins with ‘pseudoknots’ or ‘choke points’ on both sides of the sequence being repeatedly transcribed. This process is common in viral mRNA capping and editing which is mechanistically similar to polyadenylation, which could theoretically lead to small frameshift mutations on slippery nucleotides on the mRNA strand in prokaryotes and eukaryotes (Hausmann et al. 1999; Verheijen, Vermulst, and van Leeuwen 2018).

Further elucidating these mechanisms could help in understanding the process of ageing and the origin of disease pathology, and perhaps could contribute to identifying target molecules to inhibit, or even reverse, these processes to mitigate their effects in the future.

1.3.3 Detection of somatic mosaicism

1.3.3.1 Relevance: Mutational Burden, Ageing, and Disease

Accumulation of somatic mutations has been associated with ageing for several years, for e.g., there is an ageing-dependent increase in mutation frequency in several cells such as human and mouse lymphocytes, human renal epithelial cells, and transgenic reporter mouse strains [reviewed in (Kennedy, Loeb, and Herr 2012)]. This accumulation can vary at a single locus of a gene depending on the tissue within the same organism, explaining the diverse tissue-specific susceptibility in different age-related diseases (S. A. Frank 2010; Kennedy, Loeb, and Herr 2012). Meanwhile, rare progeroid syndromes, i.e., the disease state of showing signs of accelerated ageing, are often caused by defects in genes involved in DDR (Navarro, Cau, and Lévy 2006). An important example is Werner's syndrome (WS) which is caused by mutations in the Werner syndrome ATP-dependent helicase (WRN) gene that is crucial for DNA repair, homologous gene recombination, and telomere maintenance. On the other hand, in Cockayne syndrome (CS) and Xeroderma Pigmentosum (XP), the mutations abolish transcription-coupled nucleotide excision repair (National Organization for Rare Disorders (NORD) 2019).

These early findings of the accumulation of age-associated chromosomal alterations were documented with a variety of cytogenic approaches including chromosome painting, single nucleotide polymorphism (SNP) arrays, etc. (Ramsey et al. 1995; Forsberg et al. 2012). The emergence of Next Generation Sequencing (NGS) technologies in the last two decades has increased the resolution of mutation detection down to <1% for single-base substitutions. The extent of somatic mutations in healthy tissues can be thought of as an iceberg, such that the true prevalence of these mutations is only now being recognized as technologies have further improved **(Fig. 3)** (Risques and Kennedy 2018).

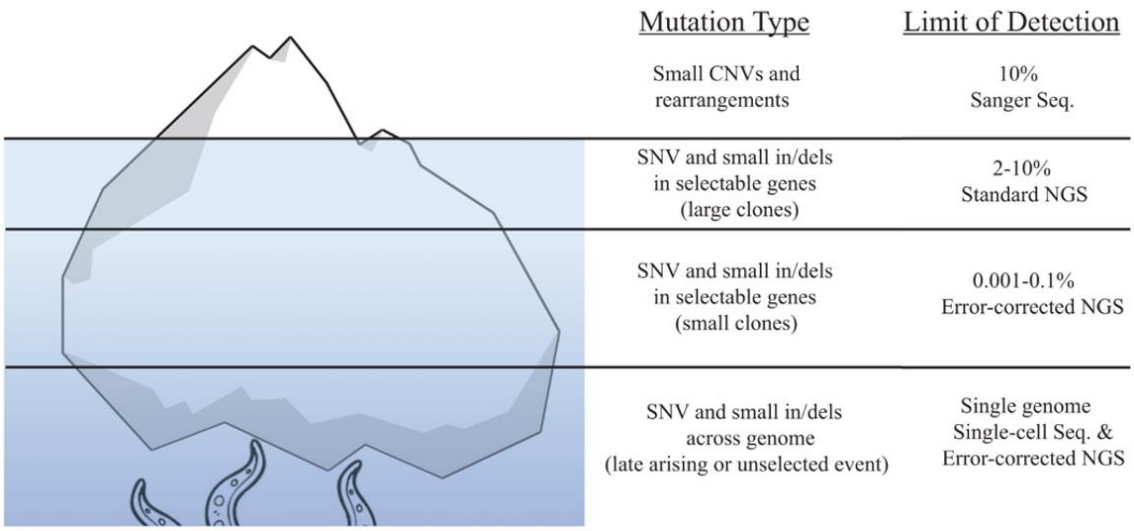


Figure 3: The limit of detecting somatic mutations.

The degree of somatic mutations present in a healthy sample is often hidden due to the limit of detection of a given technology, defined as the ability of the technology to identify a certain mutation in a biopsy/sample. This can be thought of like an iceberg, with an increasing ability to discern true mutation rates as we go deeper past the limits of detection. Adapted and modified from (Risques and Kennedy 2018)**.** CNV: Copy number variations; In/dels: Insertions and deletions; NGS: Next generation sequencing; Seq.: sequencing; SNVs: Single nucleotide variants.

One such example that highlights the importance of ultra-resolution of somatic mosaicism is the study which described the use of single-cell DNA sequencing to determine the results of ageing or defective DDR by quantifying somatic SNVs in post-mitotic neurons and found that molecular patterns of SNVs evolve with age, resulting in three distinct signatures that drive mutational spectra (Lodato et al. 2018) (**Table 2**).

Signature	Name	Most common mutation seen	Significance
A	Clock-like	C > T and T > C	Increases with age (regardless of brain region or disease state); is independent of DNA replication
B	Neurodevelopment	C > T	Does not correlate with age; has mutational mechanisms associated with neurogenesis
C	Disease	C > A	Closely associated with oxidative DNA damage and disease states (neurons with defective DDR)

Table 2: Mutational signatures seen in aged and diseased brain samples.

A: Adenine; C: Cytosine; DDR: DNA damage repair; DNA: Deoxyribonucleic acid; G: Guanine; T: Thymine. Adapted from (Lodato et al. 2018)

Signature A shows an association of accumulation of mutations with ageing in normal human brains (*genosenium*) that is even faster in progeroid diseases, while signature C shows that accumulation of mutations can give rise to diseased state in neurons (Lodato et al. 2018), findings that could only be detected at the highest levels of genomic resolution.

1.3.3.2 Approaches to study mosaicism

As mentioned earlier, cytogenetic approaches were the first to detect chromosomal anomalies in healthy as well as aged populations and techniques such as karyotyping, fluorescence *in situ* hybridization (FISH), and chromosomal microarray analysis (CMA) continue to be used as important diagnostic tools in the fields of oncology, haematology, and foetal and genetic medicine (Ozkan and Lacerda 2021). However, the development of various NGS technologies over the years have provided approaches to also analyse somatic mutations at varying limits of detection – from ~1% in older technologies to < 0.01% in newer ones – and are briefly outlined below:

A. Bulk Sequencing

Bulk sequencing refers to the high throughput sequencing methods such as "short read" NGS as well as "long read" third generation sequencing techniques with applications in exome sequencing, genome sequencing, genome resequencing, transcriptome profiling (RNA-seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization (de Magalhães, Finch, and Janssens 2010).

The current crop of *next/second generation sequencing* approaches are characterized by their massive-scale and low-cost that can produce 10^3-10^6 reads concurrently by fragmenting the template DNA into smaller sequences anywhere between 50-700 bp (depending on the platform) for "short reads", amplifying them after random sampling, and then sequencing them at once in sequencers with high confidence to give highly scalable results which can then be mapped to an appropriate reference genome and analysed for *de novo* DNA sequencing (Ansorge 2009; de Magalhães, Finch, and Janssens 2010). The commonly-used commercial sequencing platforms employ one of the following methods: chain termination (Sanger) sequencing, 454 pyrosequencing, sequencing by ligation (SOLiD sequencing), sequencing by synthesis (Illumina), ion semiconductor (Ion Torrent) sequencing, combinatorial probe anchor synthesis, and GenapSys sequencing which provide a wide range of possibilities in terms of the cost, time per run, read length, number of reads per run, and applications, with each having an accuracy of ~99.9% (Phred30) [reviewed in detail (Ansorge 2009; L. Liu et al. 2012; Quail et al. 2012)].

The more recently-developed *third generation sequencing* techniques, on the other hand, have "long read" lengths of 30,000-100,000 bp or more with applications in genome assembly,

transcript reconstruction, metagenomics, and epigenetics studies (Bleidorn 2016). The two main technologies are single-molecule real-time (SMRT) sequencing (Pacific Biosciences) which has ~87% raw read accuracy (Chin et al. 2013) and Oxford Nanopore Sequencing which has a single-read accuracy of ~95% (Niedringhaus et al. 2011). With the advantages of portability and high sequencing speed, this sequencing generation is currently undergoing active development to further reduce the higher error-rates compared to the second generation NGS and to improve molecular resolution (Gupta 2008).

The principal advantages to these bulk sequencing technologies that there are no constraints on the DNA input as the libraries are prepared from bulk cells/tissue and the data is reproducible with the possibility to perform cross-platform validations, if necessary. There are several developments in the sequencing technologies such that DNA can be sequenced at greater depths with high sensitivity and accuracy via user-friendly library preparation protocols. However, it is not possible to determine the source cell of the DNA used which makes it difficult to identify sub-clonal variants or to understand cell lineage relationships with good resolution and accuracy. These drawbacks can be overcome by using single-cell sequencing.

B. Single-cell sequencing

Single-cell sequencing (sc-seq) refers to the use of optimized NGS technologies on individual cells to extract sequencing information that provides a higher resolution of cellular differences and a better understanding of individual cellular function in its microenvironment (Eberwine et al. 2014). The elucidation of heterogeneous samples, rare cell types, cell lineage relationships, somatic mosaicism, non-culturable microbial systems analyses, and disease evolution can be performed with sc-seq (Nawy 2014).

Sc-seq is usually performed by isolating a single cell, extracting the nucleic acid and its amplification, preparing the sequencing library, Sanger or Illumina sequencing, and bioinformatic analyses (Eberwine et al. 2014). Depending on the nucleic acid extracted, the sc-seq may *be single-cell DNA sequencing* (scDNA seq) typically used for the analyses of genomic mosaicism or intra-tumour heterogeneity in cancer diagnosis (Gawad, Koh, and Quake 2016); *single-cell RNA sequencing* (scRNA seq) for examining gene expression profiles and trancriptome analyses (Tammela and Sage 2020); *single-cell DNA methylome sequencing* to quantify DNA methylation for studying epigenetic dynamics at the sub-cellular levels (Guo

et al. 2013); or, *single-cell assay for transposase-accessible chromatin sequencing* (scATAC-seq) to map chromatin accessibility across the genome of an individual cell (Stein 2019).

The advantages of single-cell sequencing are that the exact cell being sequenced is known and sub-clonal variations can be resolved more accurately than with bulk sequencing methods; however, the template is in picolitres which makes amplification with tedious protocols necessary before library preparation, a step that can introduce errors to the system. Additionally, the accuracy of sc-seq is much lower than that of bulk sequencing and the individual cell is often destroyed during the preparatory process which makes reproducibility difficult.

Due to the inherent nature of the approach used and the input of DNA required by the protocol, mutations arising earlier during cellular development are more likely to be detected by high-fidelity bulk sequencing techniques while *de novo* mutations arising in the later stages of development are more easily discernible via sc-seq approaches.

C. Error-corrected NGS: Duplex Sequencing

Error-corrected NGS techniques like Duplex Sequencing (DS) offer a middle ground between the previously described two limits of detection in that it overcomes some of the limitations of sc-seq such as reproducibility and the need for amplification before library preparation while also retaining the advantages of bulk sequencing – no constraints on DNA input and sequencing with higher sensitivity and accuracy as compared to bulk techniques, e.g., DS is capable of detecting mutations with frequencies as low as (5×10^{-8}) while the highest mutational frequency detected by conventional bulk techniques is 10^{-5} (Kennedy et al. 2014) – and that of sc-seq – resolving sub-clonal mutations, thus combining the advantages of both bulk and sc-seq approaches. Since libraries made from both the strands of the template DNA are sequenced, strand-specific mutations can be detected quite easily with the appropriate bioinformatic pipelines. However, the cost is a major drawback for DS as it requires ~40 times higher depth than conventional bulk sequencing for a single sample.

The principle of DS is essentially based on the concept of molecular tagging and the fact that double-stranded DNA is complementary to achieve the high degree of accuracy mentioned earlier. In this technique, special adapters with a randomized base duplex tag sequence are synthesized **(Fig. 4A)** and appended to either end of DNA fragments of interest so that when these are amplified with flow-cell primers to create duplex libraries, each molecule of the DNA

of interest will have two copies with the adapters ligated to each end in either an αβ or a βα orientation **(Fig. 4B)**. Since double-stranded DNA is complementary, the idea is that with appropriate bioinformatic pipelines, all the sequences amplified from both strands of the original molecule can be grouped and all those errors not shared by all the members of each family could be eliminated to get single-strand consensus (SSCS), and then the sequences from both strand-families could be compared to eliminate any errors not complementary in both strands to obtain duplex consensus (DCS) of true mutations **(Fig. 4C)**.

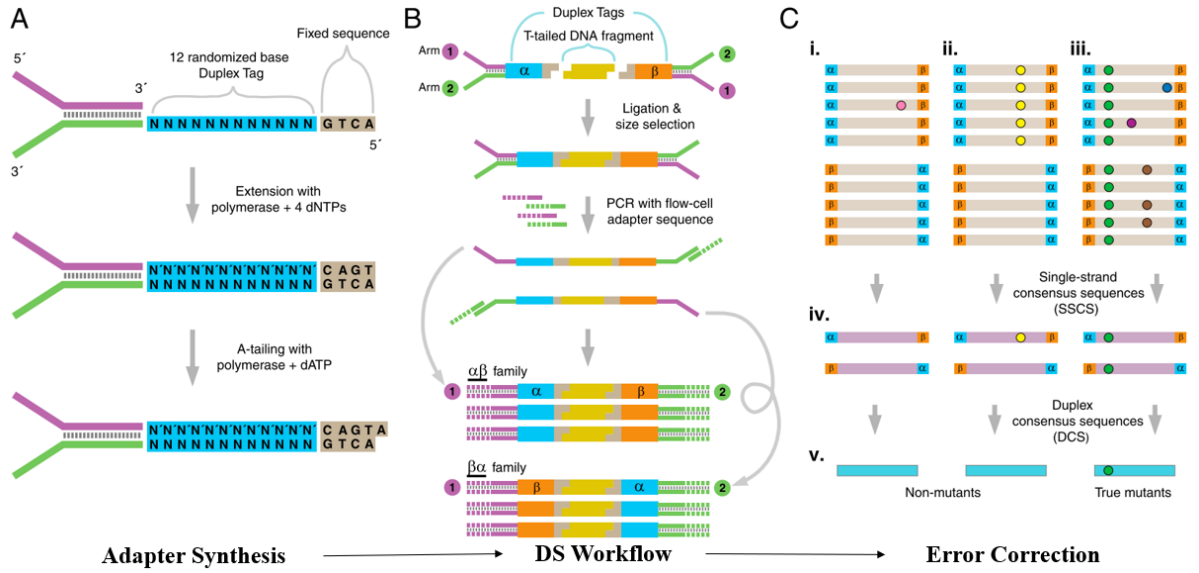


Figure 4: Principle of Duplex Sequencing.

(a) *Adapter Synthesis:* Schematic of the synthesis of duplex sequencing adapter resulting in a duplex adapter with the random double-stranded tag and the invariant spacer sequence.

(b) *DS Workflow*: Ligation of the adapters to the sample DNA results in a unique 12-nt tag sequence on both ends of the molecule. PCR amplification of each strand of a DNA duplex results in two distinct, but related, PCR products.

(c) *Error-correction with an apt bioinformatic pipeline*: Reads sharing unique α and β tag sequences are grouped together into tag families of form αβ or βα, and an SSCS is created for each tag family. Mutations are of three different types: sequencing mistakes (blue or purple dots); first-round PCR errors (brown dots); true mutations (green dots). Formation of the SSCS removes the first type of error, but not first-round PCR errors. Comparing SSCSs from the paired families with tags αβ and βα generates a DCS, which eliminates these first-round PCR errors. True mutations are scored if and only if they are present at the same position in both strands of the DNA.

DCS: Duplex strand consensus sequences; DNA: Deoxyribonucleic acid; PCR: Polymerase chain reaction; SSCS: Single strand consensus sequences.

Adapted from (Schmitt et al. 2012).

Table 3 summarises a comparison of the three main sequencing strategies described above:

Bulk sequencing	Single-cell sequencing	Error-corrected sequencing (DS)
+ No constraints on DNA input + Reproducible data + Cross-platform validations possible + Short-read and long-read sequencing is better established + Possible to achieve great depths of sequencing + Good for studying SNVs and small indels in selectable genes (large clones) + Limit of Detection: standard error rate is 10^{-2} to 10^{-3} per base call	+ The exact cell/cell-type being sequenced is known + Good resolution of sub-clonal variants is possible which makes it easier to study somatic mosaicism, delineating cell maps, etc. + SNVs and small indels across the genome (including late arising or unselected events) can be identified more easily	+ No constraints on DNA input + Reproducible data + Cross-platform validations possible + Good resolution of sub-clonal variants is possible which makes it easier to study somatic mosaicism, cell lineage relationship analysis, etc. + Good for studying SNVs and small indels in selectable genes (large clones) as well as SNVs and small indels across the genome (including late arising or unselected events) + Strand-specific mutations can be identified as well + Limit of Detection: standard error rate is ~10^{-8} per base call
– Almost impossible to identify sub-clonal variants with good resolution and accuracy – Difficult to determine cell lineage relationships for disease evolution analysis – *De novo* mutations arising in the later stages of development are unlikely to be identified	– Lower DNA input quantities (in picolitres range) which makes it difficult to handle – Accuracy of sequencing is lower than that of bulk seq – Irreproducible as the cell is destroyed during preparatory process – Amplification is usually necessary before library construction which can introduce errors into the system – Tedious protocols as compared to bulk sequencing	– Labour-intensive and longer library preparation process – Greater depths of sequencing required – Cannot sequence many samples in the same run – Extremely expensive

Table 3: The advantages and limitations of the different sequencing approaches.
A summary of the advantages (first row) and disadvantages (last row) of bulk sequencing (first column, single-cell sequencing (second column), and error-corrected sequencing like duplex sequencing (DS) (third column).
DNA: Deoxyribonucleic acid; SNVs: Single nucleotide variants.

In our present study, duplex sequencing was performed on patient samples diagnosed with frontotemporal dementia.

1.4 Detection of somatic mutation using duplex sequencing in frontotemporal dementia brain samples

1.4.1 What is Frontotemporal dementia?

Frontotemporal Dementia (FTD) is a neurodegenerative disorder that occurs due to the relatively selective degeneration of the frontal and temporal lobes of the brain. It is second to AD in terms of global prevalence of dementia and accounts for ~20% of young-onset dementia cases (Snowden, Neary, and Mann 2002) with symptoms typically manifesting between the ages of 45-65 years [average age of onset = 56.33 years] (Moore et al. 2020). Common expressions of FTD include, but are not limited to, significant changes in social and personal behaviour, apathy, blunting of emotions, and deficits in expressive as well as receptive language – these have been identified as belonging to three broad classifications of FTD (**Table 4**):

Type of FTD	Symptoms
Behavioural-variant FTD (bvFTD)	• Changes in social behaviour and conduct, • Loss of social awareness, • Poor impulse control; and, • Obsessive-compulsive disorder
Semantic Dementia (SD)	• Semantic understanding is inhibited, i.e., there is little to no word comprehension; however, • Capability to speak with good fluency and grammar is retained
Primary Non-fluent Aphasia (PNFA)	Difficulties in speech production
Frontotemporal dementia–motor neuron disease (FTD-MND)	In rare cases, FTD can also be associated with motor neuron diseases (MND), especially ALS

Table 4: Types of FTD based on clinical symptoms.

ALS: Amyotrophic Lateral Sclerosis, bvFTD: Behavioural-variant Frontotemporal Dementia; FTD: Frontotemporal Dementia; FTD-MND: Frontotemporal Dementia associated with Motor Neuron Disease; MND: Motor Neuron Diseases; PNFA: Primary non-fluent Aphasia; SD: Semantic Dementia. Adapted from (Cardarelli, Kertesz, and Knebl 2010)

Based on the type of protein involved in aggregation in the frontal and temporal lobes of post-mortem brains, FTD can also be grouped into four main categories with several clinico-neuropathological correlations (**Table 5**):

Type of FTD	Protein aggregation of	Commonly associated with
FTD-Tau	Tau	• *Dementia-dominant sub-type:* Mostly associated with bvFTD, sometimes also accompanied by poor semantic understanding • *Parkinsonism-dominant sub-type:* More commonly, Corticobasal syndrome (CBS); more rarely Progressive Supranuclear palsy (PSP)
FTD-TDP43	TDP-43	Associated with bvFTD and PNFA
FTD-FUS [Old nomenclature: FTD-FET]	FET family proteins [FUS, EWS, TAF15]	More commonly associated with FTD-MND/ALS
FTD-3 [Old nomenclature: FTD-UPS]	Ubiquitin	Dementia-only phenotype of bvFTD

Table 5: Types of FTD based on molecular pathology.

Different types of FTD pathologies based on the protein aggregation seen as well as common clinical phenotypes expressed.

bvFTD: Behavioural-variant Frontotemporal Dementia, FET family: Protein family consisting of Fused in Sarcoma protein (FUS), Ewing Sarcoma RNA Binding protein 1 (EWSR1), and TATA-binding protein-associated Factor 15 (TAF15) proteins, FTD-3: Frontotemporal Dementia with Type 3 phenotype, FTD: Frontotemporal Dementia, FTD-FUS: Frontotemporal Dementia with FET-positive inclusions, FTD-MND/ALS: Frontotemporal Dementia associated with Motor Neuron Disease/ Amyotrophic Lateral Sclerosis, FTD-Tau: Frontotemporal Dementia associated with Tau protein-positive inclusions, FTD-TDP43: Frontotemporal Dementia associated with TDP-43-positive inclusions, PNFA: Primary Non-fluent Aphasia, TDP-43: Transactive Response DNA Binding Protein 43kDa.

Adapted from (Mackenzie et al. 2011; Benussi, Padovani, and Borroni 2015)

Thus, successful diagnosis and prognosis of FTD can be challenging, particularly when the symptoms seen overlap with common phenotypes associated with other diseases as well – for example, AD [social and behavioural changes, problems in short term memory and performing spatio-visual tasks], DLB [fluctuating coarse hallucinations, Parkinsonism], vascular dementia [step-wise progression pseudobulbar palsy], and HD [chorea and dementia] (Tsuji 2012) – making the identification of specific biomarkers for each clinico-pathological phenotype necessary for effective detection of FTD and its sub-type.

1.4.2 Prevalence of FTD

Several studies have estimated that FTD accounts for approximately 8% of patients with dementia at the rate of 28 cases per 100,000 individuals aged 60-70 years in studies conducted in several European countries, including Sweden, England, and the Netherlands (Kirshner and Chawla 2018) while another study from the UK estimated the prevalence of FTD at 10.8/100,000 (Coyle-Gilchrist et al. 2016). These cases can be defined as **familial** when inherited in a family, or **sporadic** when the manifestation occurs spontaneously in a patient without any previous occurrence in the family tree. Behavioural variant-FTD accounts for 60-70% of FTD cases (Snowden, Neary, and Mann 2002; Pan and Chen 2013), SD accounts for

20-30% of FTD cases (Johnson et al. 2005; Landin-Romero et al. 2016), and PNFA accounts for ~20-25% of FTD cases (Johnson et al. 2005; Pan and Chen 2013); whether these rates of incidence of FTD (and its subtypes) remain true in other countries is yet to be examined (Landin-Romero et al. 2016). Moreover, the difficulties in correctly diagnosing FTD, especially the subtype based on clinical phenotype and pathology, as well as the limited access to quality healthcare and technology in several communities makes it difficult to estimate the true geographical distribution of FTD and its various subtypes across the world.

1.4.3 Pathogenesis of FTD

As mentioned earlier, like with many other NDD, FTD pathology is associated with the abnormal intracellular accumulation of a specific protein usually determined by techniques such as immunohistochemistry (IHC) and this specific protein accumulation has been often used to subdivide FTD pathology in the belief that the most characteristic defect would be reflective of the underlying pathogenic process (Mackenzie et al. 2011).

Given below is a brief summary of the theories proposed concerning the pathogenesis of the four main molecular subtypes of FTD:

1. **FTD-Tau**

   Tau is a microtubule-associated protein crucial for maintaining neuronal structural integrity and axoplasmic transport and is predominantly present in axons as well as in glial cells at low levels (G. Lee and Leugers 2012). Abnormal intracellular accumulation of hyperphosphorylated tau is characteristic to several tauopathies, including those that present as FTD. In each of these cases, the tau aggregates may differ with respect to the degree of phosphorylation as well as the content of the different tau isoforms that may be attributed to defective alternative splicing of exons 2, 3, and 10 of the MAPT gene encoding for tau protein (V. M. Y. Lee, Goedert, and Trojanowski 2001).

   FTD-Tau represents ~40% of all FTD cases and includes various subtypes such as (Mackenzie and Neumann 2016):

   a. *Pick's disease* – Aggregates composed of Pick bodies that are predominantly made of 3R isoforms of tau that result in the manifestation

of bvFTD or non-fluent variant primary progressive aphasia (nfvPPA) clinical phenotype with no motor deficits.

b. *Progressive Supranuclear Palsy* (PSP) – Aggregates composed of 'globose' neurofibrillary tangles (NFT) that are predominantly made of 4R isoforms of tau that result in the manifestation of movement dysfunctions such as early postural instability, axial rigidity, bradykinesia, and ophthalmoplegia; some cases of PSP are accompanied by dementia with features of bvFTD or nfvPPA.

c. *Corticobasal Degeneration* (CBD) – Aggregates composed of NFT as well as tau-immunoreactive glial and neuronal cytoplasmic inclusions (NCI) that result in the manifestation of movement dysfunctions such as rigidity, bradykinesia, dystonia, apraxia, cortical sensory signs, and alien limb phenomenon as well as dementia with features of bvFTD or nfvPPA.

d. *Globular Glial Tauopathy* (GGT) – Aggregates composed of 4R isoforms of tau with distinctive globular morphology in oligodendrocytes and astrocytes that result in the manifestation of bvFTD (with or without extrapyramidal features), primary lateral sclerosis (PLS), or a combination of both.

e. *Argyrophilic Grain Disease* (AGD) – Aggregates composed of 4R isoforms of tau with argyrophilic, tau-immunoreactive grain-like morphology that result in late-onset, progressive, mild amnestic dementia.

f. *FTD and parkinsonism linked to chromosome 17* (FTDP-17T) – Aggregates composed of hyperphosphorylated tau in neuronal and glial cells of subcortical brain. Pathogenic mutations affecting alternative splicing of exon 10 of the MAPT gene encoding tau leading to the relative increase of the 4R isoform is associated with the parkinsonism phenotype whereas mutations in exons 9, 11, 12, and 13 lead to predominance of neuronal inclusions of Pick bodies or NFT associated with dementia as the clinical phenotype.

2. FTD-TDP

The transactive response DNA-binding protein M, 43 kD (TDP-43) is a highly conserved, ubiquitously expressed DNA/RNA-binding protein involved in multiple aspects of RNA processing including transcription, splicing, transport, and stabilization

(Buratti and Baralle 2010). Although its precise functions in the brain are still unknown, TDP-43 has been implicated in neuronal development, axon guidance, and synaptic activity (Polymenidou et al. 2011).

~50% of the FTD cases have an accumulation of tau-negative, ubiquitin-positive inclusions (old nomenclature: FTD-U) with abnormal modifications that include hyperphosphorylation, ubiquitination, and N-terminal truncation resulting in FTD-TDP43 pathology (Neumann et al. 2006; Mackenzie and Neumann 2016). Differences in the morphology and cortical laminar distribution of the aggregates allows for FTD-TDP pathology to be divided into four main pathological subtypes (Mackenzie et al. 2011):

a. *Type A* is characterised by abundant short dystrophic neurites (DN) and compact oval/crescent-like NCI and is typically manifested as bvFTD or nfvPPA;
b. *Type B* is characterised by moderate numbers of compact/granular NCI in superficial and deep cortical layers with relatively few DN and neuronal intranuclear inclusions (NII) and is usually manifested as FTD-ALS;
c. *Type C* is characterised by the abundant presence of long tortuous neurites, predominantly in the superficial cortical laminae, with few/no NCI and is most common in cases of semantic-variant primary progressive aphasia (svPPA); and,
d. *Type D* is characterized by a high presence of lentiform NII and short DN in the neocortex with only rare NCI and is only found in cases with mutations in the valosin-containing protein gene (VCP).

3. FTD-FET/FUS

The FET family proteins, composed of Fused in sarcoma (FUS), Ewing's sarcoma (EWS), and TATA-binding protein-associated factor 15 (TAF15), are also homologous, ubiquitously expressed DNA/RNA-binding proteins crucial in various aspects of DNA and RNA metabolism such as RNA transcription, processing and transport, microRNA processing, and DNA repair (Sama, Ward, and Bosco 2014).

~5-10% of all FTD cases found to have a tau/TDP-negative pathology – the characteristic pathology of several diseases such as atypical FTD with ubiquitin-

positive inclusions (aFTD-U), neuronal intermediate filament inclusion disease (NIFID), and basophilic inclusion body disease (BIBD) – labelled positive for antibodies of FUS, EWS, and TAF15 proteins respectively, and these diseases were grouped together under the broad division of FTD with FET-positive inclusions (FTD-FET, old nomenclature: FTD-FUS), and the FTD clinical phenotype most commonly associated with this molecular pathology is FTD-ALS (Mackenzie et al. 2011; Mackenzie and Neumann 2016).

4. **FTD-UPS/3**

There are very few cases characterised by granular NCI that are immunoreactive for ubiquitin and p62, i.e., for the ubiquitin proteosome (UPS), but also negative for tau, TDP-43, and FUS (Holm, Isaacs, and MacKenzie 2009). Referred to as the familial FTD linked to chromosome 3 (FTD-3, old nomenclature: FTD-UPS), this condition is caused by mutations in the chromatin-modifying protein 2B (CHMP2B) gene that encodes a protein belonging to the chromatin-modifying protein/charged multivesicular body protein family (Isaacs et al. 2011). The downstream effects of the general dysfunctional endosomal trafficking and lysosomal protein degradation brought about by these pathogenic mutations is unknown and the disease-specific protein aggregation associated with this pathology remains to be discovered (Mackenzie and Neumann 2016).

1.4.4 Genetics of FTD

1.4.4.1 Detection of FTD mutations using NGS techniques

Traditionally, targeted Sanger sequencing was performed in rare cases when genetic testing was needed to clarify diagnosis wherein single gene testing or disease specific sequencing panels were performed sequentially based on probable diagnosis and family history. This gold standard of sequencing one gene at a time for diagnosis was relatively slow, labour-intensive, and highly expensive (Goldman and Deerlin 2018). The advent of NGS technologies allowed for simultaneous analysis of several genes at lower costs in shorter amounts of time which in turn enabled the study of patients within larger cohorts to better understand the genes involved – their pathogenicity, penetrance, and inheritance patterns – in familial and sporadic FTD.

Very little was known about the genetics of FTD up to the 1990s as affected families were mostly described in isolation and many different names were used to describe the mutations (Lynch et al. 1994). Although the earliest gene found to be associated with what was then termed 'frontotemporal dementia and parkinsonism linked to chromosome 17' or the microtubule-associated protein tau (MAPT), was done using traditional sequencing methods in the late 1990s, NGS technologies were widely in use to identify several candidate genes in individual families that were then reviewed for pathogenicity according to the guidelines set by the American College of Medical Genetics and Genomics (ACMG), for e.g., valosin-containing protein gene identified in 2004, chromatin-modifying 2B (CHMP2B) gene identified in 2005, progranulin gene identified in 2006, etc. (Richards et al. 2015; Diana A. Olszewska et al. 2016). These mutations were largely germline SNVs and indels, and developments in sequencing technologies made it possible to also identify other types of somatic variants, the most important one being the hexanucleotide repeat expansion on chromosome 9 (C9ORF72) confirmed to be pathogenically associated with FTD in 2011 (DeJesus-Hernandez et al. 2011).

The lowering costs of sequencing and relative ease of use along with the lesser time required to perform NGS as well as GWAS have expanded the application of this technology to larger groups and cohorts of patients with FTD which have helped reveal the implications of common molecular pathways in FTD by identifying variants as "common" or "rare" [reviewed in (Ciani et al. 2019)].

Globally, 30-40% of FTD cases have a strong genetic component manifesting with an autosomal dominant pattern of inheritance in families (Greaves and Rohrer 2019; Moore et al. 2020). Most of the heritability of FTD has been accounted for by mutations in three genes: microtubule-associated protein tau (MAPT), progranulin (GRN), and chromosome 9 open reading frame 72 (C9orf72) (Diana A. Olszewska et al. 2016). Most studies that have explored the genetic forms of FTD have focused on individual cases and families and have shown that the MAPT mutations are prevalent in ~ 10-20% of familial FTD cases, GRN mutations in 5-20% of familial cases, while the C9orf72 repeat expansion accounted for the largest mutational burden in 25-50% of familial FTD (Baker et al. 2006; Nguyen et al. 2013; Zee and Broeckhoven 2014; Benussi, Padovani, and Borroni 2015; Diana A. Olszewska et al. 2016).

However, a recent international study aiming to analyse the phenotypic forms of genetic FTD from 3403 individuals of 1492 families from 29 countries showed that while ~42%

individuals carried the C9orf72 expansion, ~35% had GRN mutations, and the MAPT mutations comprised the least common group at ~23% amongst these 3403 patients, there is also geographical variability seen in the distribution of these mutations with a different spread of prevalence among the three genetic groups in different regions, e.g., patients had predominantly GRN mutations in Italy and Spain compared to other genetic groups whereas MAPT mutations were more common in patients in Netherlands and the US west coast (Moore et al. 2020) (**Fig. 5**).

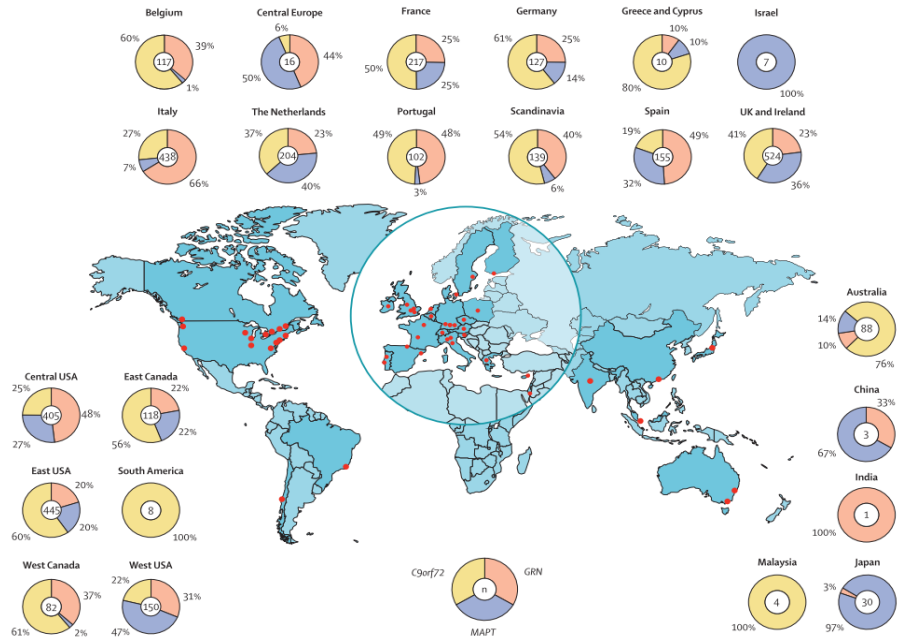


Figure 5: Frequency of each of the three genetic groups by geographic location in familial FTD. Countries with data included in the study are shown in dark blue. Individual centres are shown as red dots on the map. Pie charts show relative frequency of each of the three genetic groups within a geographical area – C9orf72 (yellow), GRN (orange), and MAPT (violet) – with the number in the centre representing the number of cases included within that area.
C9orf72: Chromosome 9 open reading frame 72 repeat, GRN: Progranulin gene, MAPT: Microtubule-associated protein tau gene. Adapted from (Moore et al. 2020).

There are several studies that focus on specific geographical locations to analyse variants in the most implicated genes as seen in the 2021 Chinese cohort study wherein they identified "pathogenic" and "likely pathogenic" variants present in a cohort of about 100 FTD patients that appeared to be single nucleotide transitions and transversions, single nucleotide insertions and deletions, and duplications as well as CAG and G_4C_2 repeat expansions in HTT and C9orf72 respectively identified using targeted sequencing on the Illumina NovaSeq 6000 platform (Jiao et al. 2021; Jiang et al. 2021). Whole exome sequencing in a well-characterized cohort that explored variants in genes most implicated in autosomal-dominant neurodegenerative dementias in about 40 FTD patients of Spanish ethnicity showed the presence of CNVs, missense, nonsense, splicing, and frameshift variants, and also highlighted

the importance of an adequate sample size (Ramos-campoy et al. 2020). Smaller studies, such as the Northern Italy cohort study, performed targeted sequencing for MAPT, GRN, and C9orf72 on ~400 frontotemporal lobar degeneration (FTLD) pedigrees in an attempt to validate previously proposed pedigree classification criteria and genetic markers found within their own populations for a more customized genetic screening for early-onset patients with unknown family history (Fostinelli et al. 2018).

Targeted exome sequencing is valuable in identifying novel genetic variants and this is more clearly reflected in studies with larger cohort sizes. For instance, the study of familial FTD in 198 patients in Netherlands by (Mol et al. 2021) used whole genome sequencing (WGS) and whole exome sequencing (WES) to identify SNVs, CNVs, indels, duplications, and microdeletions of which 74 were pathogenic variants (of which 4 were novel variants). Similar results were yielded in the first genetic screen study in a Greek cohort by (Marisa et al. 2019) which explored about 30 genes in ~120 unrelated FTD patients with targeted WES and found novel truncating and missense mutations more commonly seen in familial than sporadic cases. The Turkish cohort study with 175 FTD patients performed by (Artan et al. 2021) also showed the presence of novel pathogenic variants that were either SNVs, deletions, missense, splicing, or nonsynonymous variants.

Studies to evaluate longitudinal changes in familial and sporadic FTD patients helps give greater support to an unbiased sequencing screen irrespective of clinical presentation or family history as seen in the genetic characterization of a large FTD series in a North American cohort with 692 patients (Ramos et al. 2020). This study yielded the presence of pathogenic and likely-pathogenic variants in common and rare FTD-associated genes of familial and sporadic cases and helped identify 207 carriers in 146 families. Cross-referencing the presence of gene mutations and clinical presentation also helped identify molecular isoforms in MAPT and GRN that are more likely to cause FTD. Moreover, annual evaluations of symptomatic and asymptomatic patients with known mutations in the three main genetic groups of FTD with a standardized battery of measures also allowed investigators to observe how clinical phenotype manifests in real-time in early-familial FTD.

However, the largest cohort study in the UK by (Koriath et al. 2020) with a total of ~3250 dementia syndrome patients of which ~800 were FTD cases showed the presence of pathogenic and benign variants in 17 genes and the phenotypes associated with deleterious and non-deleterious mutations were characterized. 11% of the deleterious variants were found to be

novel and cohort stratification with a Goldman Score (GS) revealed risk/penetrance. Here, the Goldman Score refers to the quantification of a patient's family history wherein GS1 corresponds to at least three affected family members over two generations linked by a first degree relative; GS2 relates to a patient from a family with three cases but not fulfilling the criteria for GS1; GS3 relates to one relative with early-onset dementia, or GS3.5 for one relative with late-onset dementia; cases with a known negative family history were called GS4; and lastly, cases with a censored or unknown family history were categorised as GS4.5. The results from this study helped formulate a suggested decision-making chart for dementia gene-panel testing which is critical for quick and accurate diagnosis of FTD and related dementias.

While great progress has been made in the efforts to understand how FTD pathogenesis occurs by sequencing the DNA of patients from different parts of the world, difficulties associated with the correct and timely diagnosis of the disease and the lack of adequate access to good and affordable sequencing technologies in less-developed communities are some of the main reasons that still account for the gaps in our collective knowledge of the nature of this disease.

1.4.4.2 Nuclear single nucleotide variants associated with FTD

Nuclear SNVs – inherited or acquired – are the most common variations seen in FTD and are most easily detected by targeted sequencing, which requires an understanding of the varying types of FTD associated with different molecular pathologies and clinical phenotypes.

About 70% of all FTD cases are sporadic with approximately equal numbers of TDP-43 proteinopathies and tauopathies, and a smaller number of FUSopathies. About 30% is genetic with TDP-43 proteinopathies being the commonest cause brought about by mutations in *C9orf72* (usually TDP-43 types A or B), *GRN* (type A), *TBK1* (types A or B), *VCP* (type D), *SQSTM1* (FTD-3), and *TARDBP* (type B), then tauopathies (mutations in *MAPT*), FUSopathies (mutations in *FUS*), and other proteinopathies (mutations in *CHMP2B*) (Greaves and Rohrer 2019). The mutations observed in each gene are mostly SNVs, CNVs, indels, duplications, and microdeletions, and the characteristics of each gene are further described in (**Table 6**).

Genes with mutations	FTD pathology	Clinical phenotype seen	Global Prevalence	Penetrance	Age of disease onset	References
MAPT 3R Tau	FTD-Tau	bvFTD, SD, PNFA	10-20% familial, < 3% sporadic FTD cases	~100% after age 65	45-65 years	(Diana A. Olszewska et al. 2016; Goldman and Deerlin 2018; Moore et al. 2020)
MAPT 4R Tau (Microtubule-associated protein tau)		bvFTD, PNFA				
GRN (Progranulin)	FTD-TDP Type A	bvFTD, PNFA, FTD-MND	5-20% familial, 1-5% sporadic FTD cases	~100% after age 80	35-89 years	(Diana A. Olszewska et al. 2016; Goldman and Deerlin 2018; Moore et al. 2020)
C9orf72 (Chromosome 9 open reading frame 72)	FTD-TDP Types A, B and U	bvFTD, PNFA, FTD-MND	~25% of FTD cases	Age-related and incomplete penetrance	27-83 years	(Diana A. Olszewska et al. 2016; Murphy et al. 2017; Moore et al. 2020)
TARDBP (TAR-DNA binding protein 43)	FTD-TDP Type B	bvFTD, SD, FTD-MND	< 20 known FTD cases	Reduced penetrance	29-77 years	(Diana A. Olszewska et al. 2016; Blauwendraat et al. 2018; MedlinePlus Genetics 2020)
Unknown gene(s)	FTD-TDP Type C	SD; rarely bvFTD	Rare FTD cases	N/A	Not documented	(Mackenzie et al. 2011; Sieben et al. 2012)
VCP (Valosin containing protein)	FTD-TDP Type D	bvFTD	< 2.0% of FTD cases	20% penetrance	40-60 years	(Smith et al. 2011; Pottier et al. 2016; Diana A. Olszewska et al. 2016; Blauwendraat et al. 2018)
FUS (Fused in sarcoma)	FTD-FUS	bvFTD, FTD-MND	~10% of FTD cases	Varied penetrance	40-50 years	(Huey et al. 2012; Benussi, Padovani, and Borroni 2015; Takada 2015; Diana A. Olszewska et al. 2016)

CHMP2B (Chromatin-modifying protein 2B)	FTD-3	bvFTD	Rare FTD cases	Age-related and almost complete penetrance	46-65 years	(Benussi, Padovani, and Borroni 2015; Pottier et al. 2016; P. Roos et al. 2020)
TBK1 (TANK-binding kinase 1)	FTD-TDP	bvFTD, PNFA, FTD-MND	~ 1.0% of FTD cases	Incomplete penetrance	48-80 years	(Pottier et al. 2016; Van Mossevelde et al. 2016; Blauwendraat et al. 2018)
UBQLN2 (Ubiquilin 2)	FTD-3 (X-linked)	FTD-MND (mostly ALS and FTD-ALS)	Very rare FTD cases	High yet incomplete penetrance	Median: Men = 33 and women = 44.5 years	(Takada 2015; Pottier et al. 2016)
SQSTM1 (Sequestosome 1)	FTD-3	bvFTD, FTD-MND	0.9 – 3.0% of FTD cases	Low penetrance	49-72 years	(Ber et al. 2013; Diana A. Olszewska et al. 2016; Pottier et al. 2016)
PSEN1 and **PSEN2** (Presenilins 1 and 2)	Unknown proteinopathy	bvFTD	Very rare FTD cases	Incomplete penetrance	Not enough cases to obtain range	(Larner and Doran 2009; Takada 2015; Blauwendraat et al. 2018)
ITM2B (Integral membrane protein 2B)	Unknown proteinopathy	bvFTD	Extremely rare FTD cases	Incomplete penetrance	Not enough cases to obtain range	(Vldal et al. 1999; Diana A. Olszewska et al. 2016)
OPTN (Optineurin)	FTD-TDP Type A	bvFTD, PNFA, FTD-MND	Rare FTD cases; more common in ALS and FTD-ALS	Incomplete penetrance	Not enough cases to obtain range	(Pottier et al. 2016, 2018)
TBP (TATA box binding protein)	Unknown proteinopathy	bvFTD	<2.0% of FTD cases	Reduced penetrance	3-75 years	(Bruni et al. 2004; Diana A. Olszewska et al. 2016; Diana Angelika Olszewska et al. 2019)
DCTN1 (Dynactin 1)	FTD-3	FTD-MND	Extremely rare FTD cases	Low penetrance	Not enough cases to obtain range	(Benussi, Padovani, and Borroni 2015; Takada 2015; Pottier et al. 2016)

TREM2 (Triggering receptor expressed on myeloid cells 2)	Unknown proteinopathy	bvFTD	Rare FTD cases	Incomplete penetrance	20-50 years	(Borroni et al. 2014; Benussi, Padovani, and Borroni 2015)
CHCHD10 (Coiled-coil-helix-coiled-coil-helix domain containing 10)	FTD-TDP, Potential involvement of mitochondrial pathways	bvFTD, FTD-MND	< 1.0% of FTD cases	Incomplete penetrance	50-65 years	(Bannwarth et al. 2014; Pottier et al. 2016)
TMEM106B (Transmembrane protein 106B)	FTD-TDP Type A	bvFTD, PNFA, FTD-MND	-	Risk factor and a susceptibility allele for TDP43 pathology	-	(Pottier et al. 2016)

Table 6: Pathology and clinical phenotypes of gene mutations associated with FTD.
ALS: Amyotrophic Lateral Sclerosis, bvFTD: behavioural variant FTD, FET family: Protein family consisting of Fused in Sarcoma protein (FUS), Ewing Sarcoma RNA Binding protein 1 (EWSR1), and TATA-binding protein-associated Factor 15 (TAF15) proteins, FTD: Frontotemporal Dementia, FTD-3: Frontotemporal Dementia with Type 3 phenotype, FTD-ALS: FTD associated with ALS, FTD-FUS: Frontotemporal Dementia with FET-positive inclusions, FTD-MND: FTD associated with motor neuron disease, FTD-Tau: Frontotemporal Dementia associated with Tau protein-positive inclusions, FTD-TDP: Frontotemporal Dementia associated with TDP-positive inclusions, PNFA: Primary Non-fluent Aphasia, SD: Semantic Dementia, TDP: Transactive Response DNA Binding Protein.

Patients with monogenic inherited FTD often present with an extensive phenotypic variability, even amongst related individual with identical pathogenic mutations (Benussi, Padovani, and Borroni 2015). For instance: the frequency of *MAPT gene mutations* varies from 0-3% in sporadic FTD to 5-20% in familial cases, and heterogeneous clinical phenotypes with high penetrance have been reported within families with the same MAPT mutation (Van Herpen et al. 2003; Boeve et al. 2005); patients with same *GRN mutations* associated with FTD within a family often show phenotypic variability with age-dependent and incomplete penetrance (Cruts and Broeckhoven 2008; Gabryelewicz et al. 2010); although more than 30 *C9ORF72 repeats* are known to be present somatically in the brain with normal phenotype in older individuals (Xi et al. 2012; Gijselinck, Cruts, and Van Broeckhoven 2018), there is also evidence of somatic instability due to the pathogenic presence of several hundred copies of the C9ORF72 repeats, resulting in a range of mutations sizes in particular tissues with variations between tissues in the same patient, e.g., intermediate expansions in blood-sample-derived DNA that may be associated with massive expansions in the brain (Dejesus-hernandez et al. 2011; Rohrer et al. 2015; Gijselinck, Cruts, and Van Broeckhoven 2018); nearly 80% of the *VCP mutation-carrying patients* have a positive family history with varying phenotypes of FTD or FTD-MND within the same family (Benussi, Padovani, and Borroni 2015); and, reports

that FTD-MND caused by *TARDBP mutations* present within rare families usually express great phenotypic heterogeneity (Chiò et al. 2010) are all suggestive of the role somatic mutations could play in the severity of phenotype expressed in FTD patients.

These mutations can result in different types of protein aggregation and pathology that go on to manifest in varying clinical symptoms with complex correlations such that the patient can fall anywhere on the clinicopathologic spectrum of FTD, thus emphasizing the vast heterogeneity in the pathology and phenotype and the mutations that are associated with each that makes diagnosis and elucidating pathogenic mechanisms difficult (**Fig. 6**).

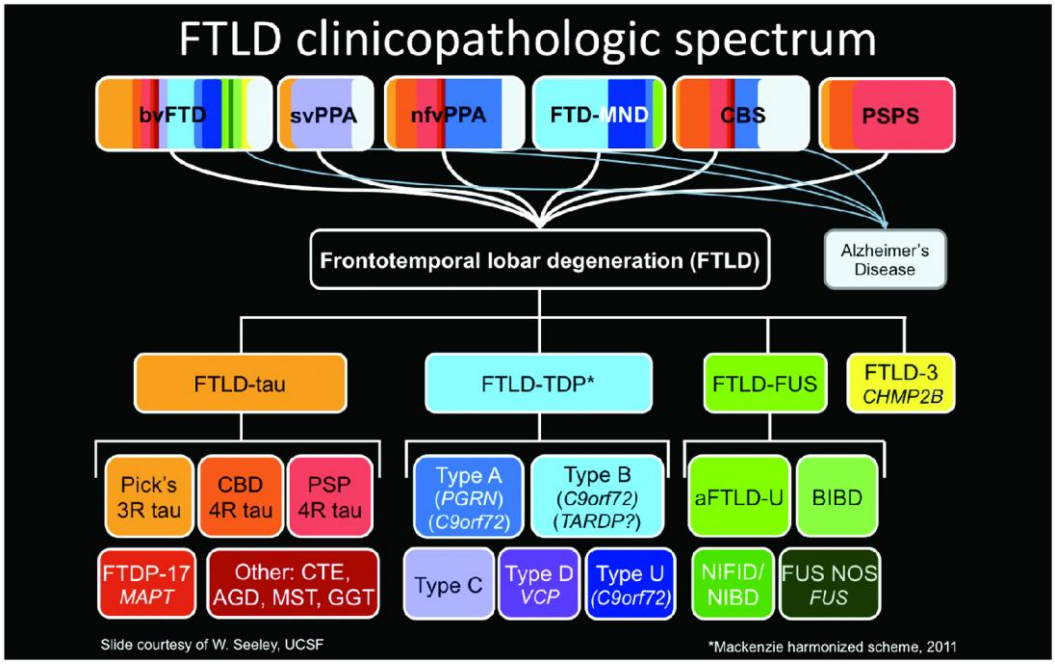


Figure 6: Clinicopathological spectrum of FTD.

The major clinical syndromes along the FTD spectrum are on the top row while the pathological conditions are on the bottom. The width of the coloured bar in each clinical syndrome represents the approximate prevalence of that syndrome (as a percentage) that has been associated with each color-coded pathology below. A portion of most of the clinical syndromes may also be attributed to Alzheimer's disease pathology (shown in white).

AGD: Argyrophilic grain disease, aFTLD-U: Atypical FTLD with ubiquitin-positive inclusions, BIBD: Basophilic inclusion body disease, bvFTD: behavioural variant FTD, CBD/S: Corticobasal disease/syndrome, CHMP2B: Chromatin-modifying protein 2B, CTE: Chronic traumatic encephalopathy, FET family: Protein family consisting of Fused in Sarcoma protein (FUS), Ewing Sarcoma RNA Binding protein 1 (EWSR1), and TATA-binding protein-associated Factor 15 (TAF15) proteins, FTD: Frontotemporal Dementia, FTD-3: Frontotemporal Dementia with Type 3 phenotype, FTD-ALS: FTD associated with ALS, FTD-FUS: Frontotemporal Dementia with FET-positive inclusions, FTD-MND: FTD associated with motor neuron disease, FTLD: Frontotemporal lobar degeneration, FTLD-tau: FTLD associated with tau protein inclusions, FTLD-TDP: FTLD associated with TDP-positive inclusions, GGT: Globular glial tauopathy, MST: Multiple system tauopathy with dementia, nfvPPA: Non-fluent variant primary progressive aphasia, NIBD: Neurofilament inclusion body disease, NIFID: Neuronal intermediate filament inclusion disease, svPPA: Semantic variant primary progressive aphasia (also known as semantic dementia), PSP: Progressive supranuclear palsy.

Adapted from (Mackenzie et al. 2011; Dickerson, Ducharme, and Onyike 2016).

1.4.4.3 Mitochondrial single nucleotide variants associated with FTD

The role of mitochondrial mutations in FTD pathogenesis has not yet been fully understood. While analyses of the 100,000 Genome Project shows that SNVs in the mitochondrial DNA are unlikely to play a role in the pathogenesis of neurodegenerative diseases (Wei et al. 2017), there are several corroborations that mitochondrial deletions, transitions, missense mutations, etc., are likely to be associated with FTD pathogenesis.

One of the first evidence indicating the presence of mitochondrial variants in FTD pathology was a report wherein Polymerase Chain Reaction–Restriction Fragment Length Polymorphism (PCR–RFLP) analysis was done to identify two mitochondrial transitions within NADH Dehydrogenase subunit 1 (ND1) sequence of a FTD patient (Grazina et al. 2004). Since then, more and more studies have linked mitochondrial dysfunction as a characteristic feature of several subtypes of FTD including, but not limited to: the likely association of mitochondrial deletions in some cases of FTD-ALS brought about by the overexpression of a mitochondrial intermembrane space protein, CHCHD10, in one single large family (Bannwarth et al. 2014); UBQLN2 mutations resulting in loss-of-function in mitochondrial protein import and subsequent decline in oxidative phosphorylation (OXPHOS) and network dynamics in some FTD/ALS patients (B. C. Lin et al. 2021); missense mutations in TBK1 are known to differentially disrupt mitophagy in patients with likely bvFTD, PNFA, and FTD-MND phenotypes, particularly FTD-ALS (Harding et al. 2021); and, mutations in SQSTM1 (resulting in the deficiency of the p62 protein it encodes for in fibroblasts of some FTD patients) was found to be associated with mitochondrial complex I inhibition and the subsequent decrease in mitochondrial membrane potential was accompanied by an increase in cytosolic reactive oxygen species (ROS) production (Bartolome et al. 2017). Recently, an analysis of the tau interactome modified by mutations that cause FTD-Tau pathology was studied in human induced pluripotent stem cell (iPSC)-derived neurons which showed that the mutations impair bioenergetics with a significant decrease in tau protein interactions with mitochondrial proteins (Tracy et al. 2022). Moreover, transgenic models with mutant/knockout TDP43, FUS, C9ORF72 repeat expansions, etc. also appear to cause mitochondrial deficits such as mitochondrial fragmentation and age-dependent inner mitochondria impairment [reviewed in (Anoar, Woodling, and Niccoli 2021)].

On a special note, results of the dual sequencing performed at high depths to identify high-fidelity mitochondrial SNVs (mtSNVs) and mtDNA rearrangements in the post-mortem brain

samples (temporal lobe, occipital lobe, and medulla) of FTD patients and age-matched controls investigated in the present thesis (described in **Table 7**) showed that both were elevated in the temporal lobe of the FTD patients, with a higher mutational burden seen in ribosomal gene variants predicted to affect intra-mitochondrial protein synthesis as well as a higher proportion of missense variants seen in genes coding for respiratory chain subunits, thus indicating a higher presence of mitochondrial variants affecting OXPHOS within the temporal lobe that may contribute to the regional vulnerability seen in FTD pathogenesis (Nie et al. 2022).

1.4.4.4 Epimutations associated with FTD

There are an increasing number of studies that have begun investigating the role of epigenetic modifications, including DNA methylation, histone modifications, and non-coding RNAs (ncRNA), and their association with ageing and NDD, especially FTD, in the recent years, especially since they can be reversed and could, thus, play a major role as potential therapeutic targets to at least modulate the severity, if not completely reverse, the onset of symptoms of NDD [reviewed in (Fenoglio, Scarpini, and Galimberti 2018)].

a. Methylation:

DNA methylation is most well-characterized epigenetic modification wherein DNA hypermethylation is associated with gene repression whereas DNA hypomethylation is associated with gene upregulation. DNA methylation is regulated by DNA methyltransferase (DNMT) enzymes.

In some cases of FTD, hypermethylation of the GRN promoter accompanied by reduced GRN mRNA levels has been reported (Banzhaf-Strathmann et al. 2013), and a specific methylation signature involving the hyper-and hypomethylation of certain non-MAPT mediated gene loci that is also pathologically associated with FTD-Tau suggested that this signature could be considered a risk factor for neurodegeneration (Y. Li et al. 2014). Hypermethylation of the 5’ CpG island located in the promoter region of the C9ORF72 (likely influenced by the length of the repeats) leading to reduced C9ORF72 expression levels has been reported in in ~10-30% of FTD-ALS patients (E. Y. Liu et al. 2014; Xi et al. 2015).

b. Histone Modifications:

Histone proteins can be modified via several post-translational changes such as acetylation, methylation, phosphorylation, ubiquitination, and citrullination, of which histone acetylation is the best understood, wherein acetylated histones are associated with gene expression upregulation and deacetylated histones relate to transcriptional silencing. Histone acetylation is mediated by histone acetyltransferases (addition of acetate molecule) and histone deacetylases (HDACs) (removal of acetate molecules).

While current research in histone acetylation in FTD mostly deals with transgenic models and iPSCs (Fenoglio, Scarpini, and Galimberti 2018), deacetylated histones at the C9ORF72 locus have been reported to reduce gene expression in FTD-ALS patients with the hexanucleotide repeat expansion (Belzil et al. 2013), and multi-HDAC-targeting inhibitors (e.g., vorinostat) were identified as enhancers of progranulin expression in a FTD patient with a nonsense GRN mutation (She et al. 2017), thus, highlighting the need for further research into histone modifier-targeted drug therapies.

c. ncRNAs:

Previously described as 'junk, non-functional DNA,' this genomic region is now known to be transcribed into non-coding RNA (ncRNA) that regulates epigenetic modelling by recruiting chromatin-modifying complexes and moderates the nuclear structure. ncRNAs mainly operate through repressive control with the potential to also act as gene activators and are comprised of: small RNAs (sRNAs) [< 200 nt length, *subtypes:* micro RNAs (miRNAs), short interfering RNAs (siRNAs), and P-element induced Wimpy testis (PIWI) protein-associated RNAs (piRNAs)] and long non-coding RNAs (lncRNAs) [> 200 nt length, *subtypes*: categorized with respect to their direction and position of transcription as antisense (AS), intergenic, exonic, and overlapping] (Derrien et al. 2012).

It is well established that altered RNA processing could contribute to NDD, including FTD (V.V., T.F., and L. 2013), and the evidence of both types of ncRNAs being altered in FTD have been reported: miR-659, miR-107, miR-29b are some of the miRNAs directly involved in the regulation of progranulin levels, along with ncRNAs mir-335-5p and let-7e that were found in higher quantities in the cerebrospinal fluid of FTD patients than controls, whereas the miR-132/212 cluster was found to repress the TMEM106B gene that increases progranulin levels post-transcriptionally (Fenoglio, Scarpini, and Galimberti 2018).

Meanwhile, lncRNAs, known to be capable of binding TDP-43 and Fused in Sarcoma/translocated in liposarcoma (FUS/TLS), are seen to be dysregulated upon the reduction of functional TDP-43 or FUS/TLS in different modality and disease status as found for Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1), Nuclear Enriched Abundant Transcript 1 (NEAT1), and Maternally Expressed Gene 3 (MEG3) in some FTD-TDP43 post-mortem samples (Tollervey et al. 2011).

1.4.5 FTD: Current Therapeutic Practices

Like most NDD, none of the subtypes of FTD have any known treatments or cure beyond management of symptoms. Psychoactive medications such as selective serotonin receptor inhibitors – sertraline, paroxetine, or fluoxetine – that can decrease disinhibition-impulsivity, repetitive behaviours, and eating disorders (Swartz et al. 1997); and, atypical antipsychotics – low doses of trazodone or aripriprazole – that can help in the management of significantly disturbed or agitated behaviour, are routinely used as pharmacological interventions for FTD (Mendez 2009). Open-label studies of anticholinesterase medicines and memantine have been negative, with behavioural symptoms exacerbated in some cases (Liscic et al. 2020).

Going forward, therapeutic approaches that target genetic factors such as engineering protein "disaggregases", e.g., heat shock protein 104 (HSP104) that can free and reactivate TDP-43 trapped in inclusions (Torrente et al. 2016; Tariq et al. 2018); engineering antibodies against TDP-43 RNA Recognition Motif 1 (RRM1) (Pozzi et al. 2019) or against the G-quadruplex structure of C9ORF72 repeats (Simone et al. 2018); engineering proteins that inhibit stress granule formations to supress TDP-43 mediated toxicity (Khalfallah et al. 2018); molecules to promote autophagy enhancement to successfully reduce cytoplasmic FUS levels (Marrone et al. 2019); gene therapy aimed at restoring proper GRN expression levels (Arrant et al. 2018); and, epigenetic modulators to reverse the severity of symptoms [reviewed in (Fenoglio, Scarpini, and Galimberti 2018)] are some of the promising interventions undergoing study in transgenic models and clinical trials at present (Liscic et al. 2020).

Thus, the high degree of variability in frequency and geographical distribution in the top three common genetic groups of FTD, the difference in the spectrum of sporadic *vs* genetic FTD, and the clinicopathological heterogeneity displayed in FTD populations, especially between individuals of the same family, further indicate that the principles of Mendelian inheritance are not enough to explain FTD pathology and, in fact, could be also due to somatic mutations and epimutations, the former of which can be analysed with higher accuracy using

the error-corrected NGS, duplex sequencing (DS), with a targeted gene sequencing approach. Furthermore, the current avenues of research targeting genetic factors to discover new pharmacological interventions emphasize the need to fully understand all sources of pathogenesis, including inherited and acquired mutations, in the hopes to successfully mitigate symptoms of FTD.

Chapter II: Research Aims

Ageing is a natural process of life in multicellular organisms that is commonly associated with an increasing mutational burden and a loss of genomic maintenance over time. Sometimes, these mutations can lead to disruptions in biochemical processes that lead to cellular dysfunction, degeneration and/or cell death. Neurodegenerative diseases (NDD) often lead to patients exhibiting debilitating symptoms with no cures that affect their quality of life and whose long-term care and management can be very expensive, thus necessitating the need to understand the origin of pathogenesis to help discern the disease progression better.

The overall aim of this research project is to establish duplex sequencing protocols to detect ultra-low frequency somatic mutations to understand their association with ageing of the human brain and NDD pathogenesis, particularly that of frontotemporal dementia. This will involve the application of targeted gene sequencing of tissues from several brain regions to understand the origin and role of somatic mutations in brain cells over their lifetime:

1. Optimize and collate duplex sequencing protocols to detect ultra-low frequency somatic mutations in specific regions of a nuclear gene;
2. To compare the occurrence of these low-frequency somatic mutations in the DNA extracted from different parts of the brain of the FTD patient and the corresponding age-matched control to determine true mutation rates; and,
3. To estimate when these mutations arise to help determine their origin during disease progression and the lifetime of the person.

Chapter III: Materials and Methods

3.1 Materials

3.1.1 FTD cases and controls

Frozen brain samples from the temporal lobe, occipital lobe, and the medulla from sixteen FTD disease cases and their respective age-matched controls that can be used in our experiments have been described in **(Table 7)**.

FTD Molecular Pathology	Case No.	Sex	Age (years)	Age-Matched Control	Sex	Age (years)
TDP43	PT145	M	60	C557	M	61
TDP43	PT154	M	72	NP16.161	M	72
TDP43	PT164	M	70	NP16.164	M	70
TDP43	PT217	F	67	NP16.184	F	69
UPS	JR31	F	46	NP16.258	F	49
TDP43	JR52	F	71	NP16.284	F	70
TDP43	JR53	M	77	NP16.267	M	77
TDP43	JR58	F	77	NP16.293	F	75
TDP43	JR62	M	70	C553	M	72
TDP43	JR64	M	64	NP17.006	M	64
TDP43	JR71	F	79	PT163	F	78
UPS	NP16.86	M	50	NP16.154	M	35
TDP43	NP16.256	M	90	NP16.119	M	95
FTLD -ALS	NP16.56	F	65	NP16.141	F	61
FTLD-ALS	NP16.61	M	61	C557	M	61
Tau/PICKS	NP16.163	F	70	C576	F	69

Table 7: Detailed description of FTD disease cases and controls.

There were sixteen cases of autopsy-confirmed FTD in this tissue bank [GREF: G7662; NRES: IRAS 136697 Rec: 13/YH/0310]. This included thirteen cases of FTD-TDP43 pathology, two cases of FTD-UPS (with ubiquitin and p62-positive inclusions, negative for tau, beta-amyloid and alpha-synuclein) pathology, two cases with FTLD-ALS pathology, and a case of FTD-tau with classical Pick's disease. Correspondingly, there were also fifteen age- and sex-matched controls with no clinical or neuropathological evidence of neurodegenerative disease for these FTD patients, with two FTD patients sharing one age- and sex-matched control.

Highlighted in yellow are the two individuals – FTD-Tau/Pick's disease patient (NP16.163) and corresponding age-matched control (C576) – whose DNA from the three brain regions underwent duplex sequencing.

Post-mortem delay for the NP16.163 patient is **56 hours**; *post-mortem delay* for the C576 individual is **40 hours**.

Clinical genetic testing revealed that the FTD-Tau/Pick's patient (NP16.163) was a **sporadic** case.

FTD: Frontotemporal dementia; FTD-ALS: FTD associated with Amyotrophic Lateral Sclerosis (ALS); FTD-TDP43: FTD with TAR DNA binding protein 43 (TDP43) inclusions; FTD-UPS: FTD with inclusions positive for ubiquitin proteosome system markers; F: Female, M: Male.

These frozen samples were obtained from the Human Research Tissue Bank, Addenbrooke's Hospital, Cambridge [GREF: G7662; NRES: IRAS 136697 Rec: 13/YH/0310] (The Human Research Tissue Bank is supported by the NIHR Cambridge Biomedical Research Centre).

3.1.2 Reagents

- 128J cell line DNA used as 'test DNA' for some optimization experiments.
- 2X HI-RPM Hybridization Buffer (cat. no.: 5188-6417, lot no.: 0006485566, Agilent Technologies, US) used in the hybridization mixture for Double-capture (with in-house biotinylated baits).
- 10X aCGH Lyophilized Blocking Agent (cat. no.: 5188-6416, lot no.: 0006408029, Agilent Technologies, US) used in the hybridization mixture for Double-capture (with in-house biotinylated baits).
- Agarose (lot no.: W13371891, Thermo Scientific, US) used for agarose gel preparations.
- Beckman Coulter Agencourt Ampure XP (A63881, Beckman Coulter, Fisher Scientific, UK) for beads-based DNA purifications.
- Blocking oligos (lab-designed, IDT technologies) used for Double-capture (with in-house biotinylated baits) to block the adapter and Illumina primer sequences so that only duplex libraries containing desired genome sequences are captured.
- Duplex adapters (50 µM, in-house stock prepared as per (Kennedy et al. 2014))
- Dynabeads® M-270 Streptavidin (cat. no.: 65305, Thermo Fisher Scientific, US) for beads-based capture reactions.
- Ethanol (lot no.: BCCD1481, Sigma-Aldrich, US)
- KAPA HiFi HotStart ReadyMix (2X, 50 reactions) (cat. no. KK2601, Roche Sequencing, US) used for post-capture library amplifications in Double-capture (with custom-designed IDT probes).
- KAPA HyperPlus kit (code: KK8512, cat. no. 07962410110, Roche Sequencing, US) used in duplex library preparation.
- Nuclease-free water (AM9937, Ambion, US).
- P5 and P7 index primers (Illumina, US) to construct duplex libraries.
- PCR and qPCR primers as well as in-house biotinylated baits and custom-designed probes for MAPT (Integrated DNA Technologies, US).

- Platinum SuperFi II DNA polymerase (cat. no. 12361010, Thermo Fisher Scientific, US) used for amplicon-based duplex library preparation.
- PrimeSTAR GXL DNA polymerase (R050A, Takara Bio Inc., Japan) used for amplicon-based duplex library preparation.
- QIAamp Fast DNA Tissue kit (cat. no. 51404, QIAGEN, US) used to extract DNA from FTD patient and control frozen brain samples
- Quick-Load 1 kb-extend DNA ladder (N3239S, New England Biolabs, UK)
- Quick-Load Purple 1 kb DNA ladder (N0552S, New England Biolabs, UK)
- Qubit dsDNA HS Assay kit (500 reactions) (cat. no. Q32854, Thermo Fisher Scientific, US)
- Sodium chloride (lot no.: SZB40460V, Sigma-Aldrich, US) used to make wash buffers for Double-capture (with in-house biotinylated baits)
- Sodium hydroxide pellets (product no.: 28244.262, batch: 15C120018, VWR Chemicals, US) used to make wash buffers for Double-capture (with in-house biotinylated baits)
- SYBR-Safe DNA Gel Stain (lot no.: 2399638A, Invitrogen, Life Technologies, Thermo Fisher Scientific, US) used for DNA staining in agarose gel preparations.
- Tris (1 M, pH 8.0) (lot no.: 00712558, Invitrogen, Life Technologies, Thermo Fisher Scientific, US) used to make wash buffers for Double-capture (with in-house biotinylated baits)
- Tris-base Acetate EDTA (TAE) buffer (pH 8.3, 1X) (diluted from 20X stock bought at the departmental lab store) for agarose gel preparation.
- Tween20 (lot no.: SLBR6201V, PC code: 1002406511, Sigma-Aldrich, US) used to make wash buffers for Double-capture (with in-house biotinylated baits)
- xGen Universal Blockers-TS-Mix (16 reactions) (cat. no.: 1075474, Integrated DNA Technologies, US) used for Double-capture (with custom-designed IDT probes)
- xGen Hybridization and Wash kit (cat. no. 1080577, Integrated DNA Technologies, US) used for Double-capture (with custom-designed IDT probes)

3.1.3 Equipment

- DynaMag-2 magnet (Ref. 12321D, Invitrogen, US) to separate beads and supernatant

- High Sensitivity D1000 ScreenTape (code: 5067-5584, Agilent Technologies, US)
- High Sensitivity D5000 ScreenTape (code: 5067-5592, Agilent Technologies, US)
- Lo-Bind microcentrifuge tubes (1.5 ml) (cat. no. F170109Q, Eppendorf, Germany)
- Micropipettes and pipette tips (1 µl, 20 µl, 200 µl, 1000 µl) (Pipettman, Star Lab, UK)
- Mini Dry Bath (Flowgen Bioscience Limited, UK)
- Mini Spinner (Star Lab, UK)
- MiSeq 2500 (Illumina, US)
- Molecular Imager ChemiDoc^{TM} XRS^{+} Imaging System with Image Lab software (Bio-Rad Laboratories, US)
- NovaSeq 6000 (Illumina, US)
- PCR consumables (0.2 ml 8-strip PCR tubes with individually attached flat caps) (Star Lab, UK)
- PCR machine: Veriti 96-well thermocycler (Applied Biosystems, Life Technologies, US)
- qPCR machine: C1000 Touch Thermocycler (Bio-Rad Laboratories, US)
- Qubit 2.0 Fluorometer (Invitrogen, Life Technologies, US)
- Savant SPD1Z1P SpeedVac Concentrator (Thermo Scientific, US)
- TapeStation Bioanalyzer 2200 (Agilent Technologies, US)
- Thermocycler and shaker (ThermoMixer F1.5, Eppendorf, Germany)
- Vortex–Genie 2 machine (Scientific Industries Inc., US)

3.1.4 Software

- Bio-Rad iQ5™ optical software v2.0 (Bio-Rad Laboratories, US) for quantitative polymerase chain reaction (qPCR) data collection
- Image Lab (Bio-Rad Laboratories, US) for agarose gel image acquisition and analysis
- Microsoft Office Excel 2010 (Microsoft Systems, US) for qPCR data analyses
- For sequencing analyses (as recommended in (Kennedy et al. 2014)):
    - Burrows-Wheeler Aligner (BWA) (H. Li and Durbin 2009)
    - Sequence Alignment Map (SAM) (H. Li et al. 2009)
    - Genome Analysis Tool Kit (GATK) (McKenna et al. 2010)
    - FASTQC (Andrews 2010)

- Python 3 (van Rossum and Drake 2009)
- RStudio (RStudio Team 2020).

3.2 Methods

3.2.1 Genomic DNA Extraction

DNA was extracted with the QIAamp Fast DNA Tissue Kit (cat. no. 51404) using a combination of mechanical, chemical, and enzymatic lysis (QIAamp® Fast DNA Tissue Kit Handbook 2015; QIAamp ® Fast DNA Tissue Kit Safety Data Sheet 2021).

Suitably-sized (~20 mg) tissue samples were placed in the supplied Tissue Disruption Tube. A suitably scaled-up master digestion buffer mix (as multiple samples were processed) was made with the following buffers and enzymes: 200 µl AVE (buffer containing sodium-azide to prevent microbial contamination; provided in the kit), 40 µl VXL (lysis buffer containing guanidine hydrochloride for protein denaturation; provided in the kit), 1 µl DX reagent (comprises of stearyl alcohol and tridecanol to reduce foaming, provided in the kit), 20 µl Proteinase K (proteinase for protein digestion and nuclease inactivation; provided in the kit), and 4 µl RNase A (ribonuclease A for RNA degradation; provided in the kit). This digestion buffer mix was added to the tissue disruption tube and the lid tightly capped. The samples were homogenized using the Vortex-Genie2 at full speed for 5 min each. The samples were then incubated in a thermomixer at 1000 rpm for 10 min at 56 ^{0}C. If there was residual tissue left after this incubation, then homogenization and incubation steps were repeated once. Then, 265 µl Buffer MVL (lysis buffer containing guanidine thiocyanate for cell lysis and denaturation of various DNase and RNase enzymes; provided in the kit) was added to each sample and mixed by pipetting ten times. This mixture was added to the supplied QIAamp Mini spin column and centrifuged for 1 min at 4 ^{0}C, 20,000 *x g*. Then, the spin column was placed in a clean 2 ml collection tube (provided) and the filtrate discarded. 500 µl Buffer AW1 (Wash buffer 1 comprising of guanidine hydrochloride; provided in the kit) was to the spin column and centrifuged for 1 min at 4 ^{0}C, 20,000 *x g*. Then, the spin column was placed in a new 2 ml collection tube (provided) and the filtrate discarded. 500 µl Buffer AW2 (Wash buffer 2; provided in the kit) was to the spin column and centrifuged for 1 min at 4 ^{0}C, 20,000 *x g*. Then, the spin column was placed in a new 2 ml collection tube (provided) and the filtrate discarded. The spin column was centrifuged for 2 min at 4 ^{0}C, 20,000 *x g* and placed in a clean 1.5 ml microcentrifuge tube (provided). 100 µl ATE (Buffer components: 10 mM Tris-Cl pH 8.3, 0.1

mM EDTA, and 0.04% sodium-azide; provided in the kit) was added directly onto the spin column membrane, incubated at room temperature for 1 min, and then centrifuged for 1 min at 4 ^{0}C, 20,000 *x g*. This last step with 100 µl ATE addition, subsequent incubation and centrifugation was repeated for increased yield before discarding the spin column. The extracted DNA was now ready for further experiments.

3.2.2 Size selection and DNA Purification with SPRI

3.2.2.1 Principle of solid-phase reversible immobilization (SPRI)

Ampure XP beads (A63881, Beckman Coulter) are used to select DNA of interest and to discard the shorter fragments resulting from PCR. The beads are paramagnetic due to a layer of magnetite surrounding their polystyrene core, which is in turn coated with carboxyl molecules that reversibly bind the negatively-charged DNA in the presence of the crowding agent – a mix of polyethylene glycol (PEG) and 2.5 M NaCl, or sodium chloride (Deangelis, Wang, and Hawkins 1995) **(Fig. 7a)**. This immobilization is dependent on the concentration of the crowding agent; therefore, the volumetric ratio of beads to DNA is critical – lower the ratio of beads:DNA, larger are the fragments that bind to the beads due to their larger total charge that promotes their electrostatic interactions with the beads, thus displacing smaller DNA fragments (**Fig. 7b**).

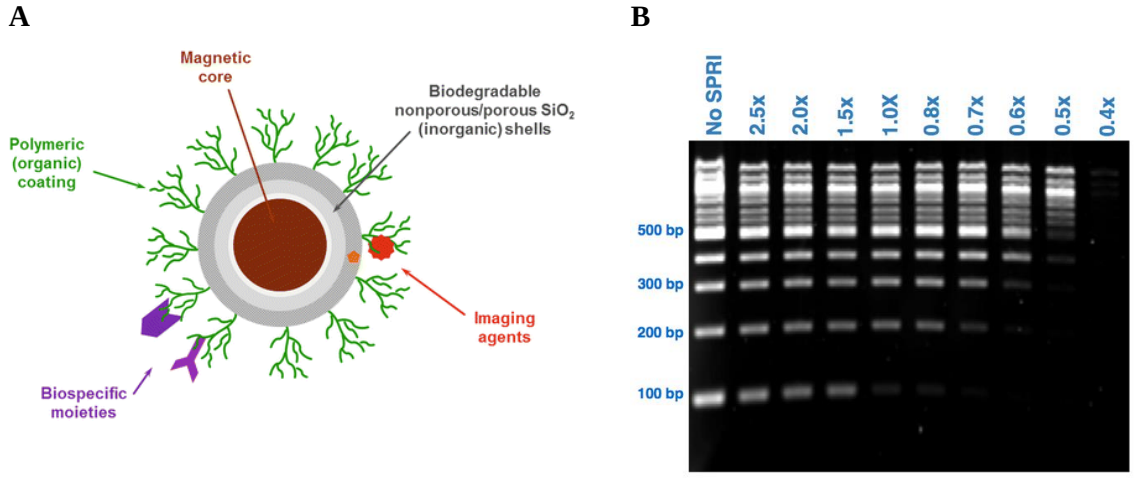


Figure 7: Principle of Ampure XP beads - DNA fragment size vs beads ratio.

(A) Diagrammatic representation of DNA fragments binding to Ampure XP beads. Adapted from (Kovačević 2016).

(B): Gel results of fragment sizes obtained with varying beads:DNA ratios. Adapted from (CoreGenomics 2012).

3.2.2.3 Single-tail beads selection

For our purposes, we need to select for fragments ≥ 300 bp. The DNA and 0.7X ratio of Ampure XP beads were mixed in Lo-Bind microcentrifuge tubes (cat. no. F170109Q, Eppendorf). The components were mixed gently by pipetting ten times, followed by a brief spin and incubation at 37 ^{0}C, and 5 minutes in a thermo-shaker (ThermoMixer F1.5, Eppendorf) at 550 rpm. The Lo-Bind tubes were incubated on the DynaMag-2 magnet (Ref. 12321D, Invitrogen) (magnetic rack) for 5 minutes at room temperature (until supernatant cleared). Still on the magnetic rack, the supernatant was aspirated and discarded, while the beads were washed with 200 µl 80% ethanol (v/v) for 30 seconds. The ethanol wash was repeated twice, followed by air-drying the beads for 3 minutes (care taken to not over-dry the beads). The Lo-Bind tubes were removed from the magnetic rack and the beads suspended in 25 µl nuclease-free water (AM9937, Ambion). This was incubated at 37 ^{0}C for 5 minutes in the thermo-shaker (ThermoMixer F1.5, Eppendorf) at 550 rpm. The Lo-Bind tubes were placed on the magnetic rack at room temperature for 5 minutes, and then the supernatant was aspirated into fresh eppendorfs. High Sensitivity D1000 ScreenTape analysis (code 5067-5584, Agilent) with TapeStation 2200 (Agilent) [fast, automated, and flexible electrophoresis for sample quality control routinely used in NGS workflows (Agilent Technologies 2015)] and Qubit dsDNA HS assay were done to check for quality and quantity of the DNA respectively.

3.2.2.4 Double-tail beads selection

Two rounds of bead selections were done when the DNA fragments to be selected were 300-400 bp. For this, we first performed 0.64X Ampure XP beads selection (ratio of bead volume to DNA volume), incubated at 37 ^{0}C for 5 minutes in a thermo-shaker (ThermoMixer F1.5, Eppendorf) at 550 rpm. The Lo-Bind tubes were then placed on the DynaMag-2 magnet (Ref. 12321D, Invitrogen) (magnetic rack) for 5 minutes at room temperature (till supernatant cleared). Still on the magnetic rack, the supernatant was aspirated and transferred to a fresh tube. 1.0X ratio of Ampure XP beads was added to the DNA in the fresh tube and the beads selection protocol was performed as described in Chapter 3, section 3.2.2.2. The volume of beads required for double-selection was calculated with the following formula: if,

a = first round bead:DNA ratio
b = second round bead:DNA ratio (always greater than a), then,
Volume of beads in first selection = *a* * initial DNA volume
Volume of beads in second selection = [(*b* * initial DNA volume) - volume used in first selection]

In our experiments, we performed a 0.64X followed by a 1X double selection for 50 µl PCR end products, beads volume for 1st selection = (0.64 x 50) = 32 µl; beads volume for 2nd selection = [(1 x 50) – 32] = 18 µl.

3.2.3 Duplex library Preparation using KAPA protocol

The duplex DNA libraries were prepared with the (KAPA Biosystems 2017) protocol modified to suit our purposes and KAPA HyperPlus Kit (Code: KK8512, cat. No. 07962410110).

3.2.3.1 Enzymatic Fragmentation

The required amount of template DNA was taken in a PCR tube and nuclease-free water added to make the volume 35 µl and then the enzymatic fragmentation reagents added as described in (**Table 8**):

Contents	Volume (µl)
DNA (50 ng input)	35
Frag Buffer	5
Frag Enzyme	10
Total	**50**

Table 8: Enzymatic fragmentation reaction contents.

The tube was vortexed and gently spun before incubation. Since the amplicon fragment size of interest is ~150 bp, the reaction tube was incubated in a thermocycler **(Table 9)**:

Step	Temperature (^{0}C)	Time (min)
Pre-cool block	4	5
Fragmentation	37	30
Hold	4	∞

Table 9: Enzymatic fragmentation incubation conditions.

Fragmentation temperature set as per kit recommendations to the standard fragmentation temperature of 37 ^{0}C for high-quality genomic DNA; fragmentation time set as 30 minutes as per kit recommendations to achieve final mode fragment length of 150 bp.

3.2.3.2 End-Repair and A-tailing (ER-AT)

The reagents as described in (**Table 10**) were added to the tube from the end of (Chapter 3, section 2.3.1) before vortexing and a gentle spin followed by incubation as described in (**Table 11**):

Contents	Volume (µl)
Fragmented DNA	50
ER-AT Buffer	7
ER-AT Enzyme	3
Total	**60**

Table 10: End-repair and A-tailing (ER-AT) reaction contents.

Step	Temperature (^{0}C)	Time (min)
ER-AT	65	30
Hold	4	∞

Table 11: End-repair and A-tailing (ER-AT) incubation conditions.

3.2.3.3 Adapter-ligation

The number of adapters (sequence and structure described in **Fig. 8**) in the system needs to be 20x that of the ER-AT end-product in the system. Calculations indicated that 5 µl of 3 mM adapters were optimal for our purposes.

Figure 8: Duplex adapter sequence and structure.
(a) Oligonucleotide sequences used for duplex adapter preparation (b) Diagram of adapter design and binding to DNA of interest (N = any nucleotide; R = nucleotides in DNA of interest).
MWS51 = P5 primer, MWS55 = Test primer in all further experiments.

The reagents as described in (**Table 12**) were added to the tube from the end of (Chapter 3, section 2.3.2) before vortexing and a gentle spin followed by incubation as described in (**Table 13**):

Contents	Volume (µl)
ER-AT DNA	60
Ligation Buffer	30
Ligase	10
Adapters (3 µM)	5
Nuclease-free water	5
Total	**110**

Table 12: Adapter-ligation (AL) reaction contents.

Step	Temperature (^{0}C)	Time (min)
ER-AT	20	30
Hold	4	∞

Table 13: Adapter-ligation (AL) incubation conditions.

3.2.3.4 Post-ligation purification

The DNA product from the end of Chapter 3, section 3.2.3.3 was purified as in Chapter 3, section 3.2.2.2 (single-tailed beads selection).

3.2.3.5 Library Amplification

The number of DNA fragments used in the PCR, along with the fraction of a sequencing lane dedicated to a particular sample, are the primary adjustable variables that dictate the number of sequencing reads that share the same tag sequence which strongly influences the final number of duplex consensus read sequences (DCSs) formed (Kennedy et al. 2014). Thus, optimizing the PCR input and the number of amplification cycles is crucial to maximizing the efficiency of the duplex sequencing (DS) technique.

1. Optimization – Varying DNA input

Previous work in the lab showed that >10 amol input DNA gave large tag family size (reads with duplex consensus) and inefficient DS. Two sets of experiments were performed to study library amplification when the DNA input for amplification were below 20 amol and 2 amol respectively: Set I – (i) 4 amol, (ii) 8 amol, (iii) 12 amol, and (iv) 16 amol; Set II: 60 amol purified adapter-ligated DNA was serially diluted (dilution factor = 6) – (i) 10 amol, (ii) 1.67 amol, and (iii) 0.28 amol – and the reagents described in (**Table 14**) were added and set to the PCR programme described in (**Table 15**):

Contents (For one reaction)	GXL DNA Polymerase Volume (μl)
Buffer	5.0
dNTP Mix	2.0
DNA Polymerase	0.5
P5 (10 μM)	0.5
Test (10 μM)	0.5
DNA	Variable
Nuclease-free water	16.5 – DNA vol.
Total	**25**

Table 14: PCR tube contents for one reaction of duplex library amplification.

	Temperature (^{0}C)	Time (min)	Repeat
Initial Denaturation	98	0:45	x1
Denaturation	98	0:15	x20 cycles
Annealing	60	0:30	
Extension	60	0:30	
Final Extension	72	1:00	x1
Hold	4	∞	

Table 15: PCR programme for duplex library amplification optimization – varying DNA input.

2. Optimization – Varying PCR cycle numbers

10 amol of purified adapter-ligated DNA were mixed with the reagents described in (**Table 14**) and set to the PCR programme described in (**Table 16**):

Step	Temperature (^{0}C)	Time (min)	Repeat				
Initial Denaturation	98	0:45	x1				
Denaturation	98	0:15	x16	x18	x20	x22	x24
Annealing	60	0:30					
Extension	60	0:30					
Final Extension	72	1:00	x1				
Hold	4	∞	-				

Table 16: PCR programme for duplex library amplification optimization – varying PCR cycle number.

3. Optimized Library amplification for amplicon-based duplex library

We observed that < 2 amol DNA for PCR input gave no readings on the Qubit or HS D5000 ScreenTape, and that cycle numbers > 20 gave higher molecular weight non-specific fragment formation (see chapter 4, section 4.2.2). Hence, for all the amplicon-based duplex library preparations, we put in 4 amol of purified adapter-ligated DNA for 20 cycles.

3.2.3.6 Post-amplification purification

The DNA product from the end of Chapter 3, section 3.2.3.5 was purified as in Chapter 3, section 3.2.2.2 (single-tailed beads selection).

3.2.3.7 Quantitative polymerase chain reaction (qPCR) (For optimizations)

qPCR monitors the amplification of targeted DNA during the PCR via fluorescent dye-labelled DNA probes that are complementary to the target sequence, and the relative initial amount of DNA in the sample can be determined from quantitation cycle (C_q) and expression fold change (2^{-dCq}) values.

a) qPCR primers

qPCR was used to verify the presence of amplified adapter-ligated DNA species in the duplex library system by using unique and specific primer–DNA probe mixes to quantify the adapter-ligated amplicon and the gene of interest (**Table 17**). The fluorescent dyes conjugated with the 5' DNA probes were 'HEX' and 'FAM', both of whose 3' quencher was 'BHQ_1'.

qPCR	Amplicons	Sequence (5' … 3')	Nucleotide Position Start	Nucleotide Position End
Primers	MAPT-A F (10 µM)	GCTTCACTCCCTTCCTTCCTTC	106911	106932
	MAPT-A R (10 µM)	GACTTGACATTCTTCAGGTCTGGC	107130	107153
	MAPT-B F (10 µM)	CTCATCGAAAGTGGAGGCGTC	120764	120784
	MAPT-B R (10 µM)	CATGGGACGTGTGAAGGTACTC	120943	120964
	MAPT-C F (10 µM)	CACTTTGCCCTGTAGACTGCAG	129057	129078
	MAPT-C R (10 µM)	CGTGGGTGATATTGTCCAGGG	129225	129245
	RNaseP F (10 µM)	AGATTTGGACCTGCGAGCG	CDC	
	RNaseP R (10 µM)	GAGCGGCTGTCTCCACAAGT	CDC	
Probes	MAPT-A_Hex (10 µM)	CAAAATCAGGGGATCGCAGCGGC	106956	106978
	MAPT-B_HEX (10 µM)	GAAGCTGGATCTTAGCAACGTCC	1208865	1208887
	MAPT-C_HEX (10 µM)	GAGGTGGCCAGGTGGAAGTAAAATC	129156	1291180
	RNaseP_FAM (10 µM)	TTCTGACCTGAAGGCTCTGCGCG	CDC	

Table 17: Oligonucleotide sequences for MAPT amplicons qPCR primers and probes.

*(*http://biotools.nubic.northwestern.edu/OligoCalc.html *to calculate melting temperatures and GC content of primers designed);* (CDC protocol of real-time PCR (RT-PCR) for influenza A (H1N1): http://www.who.int/csr/resources/publications/swineflu/en/)

The DNA samples were diluted to 0.1 ng/ml to the following assays (**Table 18**):

Assay 1 Contents	Assay 2 Contents	Assay 3 Contents	Vol. (µl)
MAPT-A F (qPCR_10uM)	MAPT-B F (qPCR_10uM)	MAPT-C F (qPCR_10uM)	0.75
MAPT-A R (qPCR_10uM)	MAPT-B R (qPCR_10uM)	MAPT-C R (qPCR_10uM)	0.75
RNaseP F (10 uM)			0.75
RNaseP R (10 uM)			0.75
MAPT-A_HEX	MAPT-B_HEX	MAPT-C_HEX	0.5
RNaseP_FAM			0.5
iTaq Master-Mix			12.5
DNA			1
Nuclease-free water			7.5
Total			**25**

Table 18: Assays for qPCR verification of adapter-ligated duplex library.

b) qPCR programme

The following PCR programmes were used for MAPT-amplicons (**Table 19**):

Steps	Temperature (^{0}C)	Time (min)	Repeat
Denaturation	95	3:00	x40 cycles
Annealing	95	0:10	
Extension	55	0:30	
Hold	4	∞	

Table 19: qPCR programme for verification of adapter-ligated duplex library.

3.2.4 Amplicon-based method of duplex library preparation

3.2.4.1 Pre-library amplification with polymerase chain reaction (PCR)

a) Primers

We began with the MAPT gene to optimize our duplex sequencing protocols in the FTD-Tau/Pick's case and the control as it is the main gene involved in FTD-Tau pathology.

For the amplicon-based method, the pre-amplification stage was performed with PrimeSTAR GXL DNA Polymerase (R050A, Takara) and the Platinum SuperFi II DNA Polymerase (cat. no. 12361010, Thermo Fisher Scientific) with the PCR primers described in (**Table 20**) (NCBI Reference Sequence: NC_000017.11). The primers were designed within the parameters recommended by Thermo Fisher Scientific ("PCR Primer Design Tips - Behind the Bench" 2019).

Amplicon Name	Exons covered	Primer Sequence (F) 5'… 3'	Primer Sequence (R) 5'… 3'	Nucleotide Start (5') … End (3') of amplicon	Amplicon Size
MAPT-A	Exon 8	GGCAAAGCATATCTACCACCTCC	CACTTGCTCCCAGAAGAATCAGC	106500…107708	1209 bp
MAPT-B	Exons 9-10	GACAGAGCCCTCGTACAGAAG	CATCCATATAGCTGAACCTCGAG	120153…126524	6372 bp
MAPT-C	Exons 11-12	GGTCTTGTGGGGTCATGGTTTAC	CACACTCACACAAGGTTGACATCG	128346…135213	6868 bp

Table 20: Primers designed for MAPT amplicons of interest.
Primers designed using the MAPT sequence from GenBank
*(*https://www.ncbi.nlm.nih.gov/nuccore/NC_000017.11?report=genbank&from=45894538&to=46028334*)*

b) PCR programmes

~50 ng of template DNA (FTD case or control) was put in for the following PCR reactions (**Table 21**):

Contents (For one reaction)	SuperFi DNA Polymerase Vol. (µl)	GXL DNA Polymerase Vol. (µl)
Buffer	4.0	5.0
dNTP Mix	0.5	2.0
DNA Polymerase	0.5	0.5
MAPT Forward primer (10 µM)	0.5	0.5
MAPT Reverse primer (10 µM)	0.5	0.5
DNA (~50 ng input)	Variable	Variable
Nuclease-free water	14.2 – DNA vol.	16.5 – DNA vol.
Total	**20**	**25**

Table 21: PCR tube contents for one reaction of pre-library amplification with SuperFi and GXL DNA polymerases.

The following PCR programmes were optimized to be used for the preparation of all three MAPT amplicons (**Table 22**):

Steps	Temperature (^{0}C)	Time (min)		Repeat
MAPT-A (GXL and SuperFi)				
Initial denaturation	94	1:00		x1
Denaturation	98	0:15		x30 cycles
Annealing	60	0:15		
Extension	68	1:00 (GXL)	2:00 (SuperFi)	
Hold	4	∞		
MAPT-B (GXL and SuperFi) and MAPT-C (GXL)				
Initial denaturation	94	1:00		x1
Denaturation	98	0:10		x30 cycles
Annealing	56	0:15		
Extension	68	6:30		
Final Extension	72	7:00		x1
Hold	4	∞		-
MAPT-C (SuperFi)				
Initial denaturation	94	1:00		x1
Denaturation	98	0:15		x35 cycles
Pre-annealing	68	0:10		
Annealing	60	0:15		
Extension	68	10:00		
Final Extension	72	10:00		x1
Hold	4	∞		-

Table 22: PCR programmes for pre-library amplification for the MAPT amplicons.

3.2.4.2 Agarose gel preparation and Image Acquisition

The MAPT amplicons obtained at the end of the PCR run were viewed in 0.7x agarose gel, i.e., 70 g agarose added to 100 ml of 1X TAE buffer along with 10 µl of SYBR-Safe DNA gel stain (lot no.: 2399638A, Invitrogen, Life Technologies, Thermo Fisher Scientific). This gel stain is a safer alternative to Ethidium Bromide that binds to the grooves of the DNA double-

helix and blue light or UV excitation can be utilized to view DNA bands on the gel in an appropriate gel imaging system.

Electrophoresis was done at 110V for 30 minutes before the gel was viewed on the gel imaging system (Molecular Imager ChemiDoc^{TM} XRS^{+} Imaging System with Image Lab software, Bio-Rad Laboratories).

3.2.4.3 Duplex library preparation

Equimolar proportions of the three MAPT amplicons were pooled together and the steps described in Chapter 3, section 3.2.3 was followed with 4 amol input, 20 cycles for library amplification.

Calculation for pooling amplicons in equimolar proportions: The volume of DNA (µl) required from each amplicon was calculated by first dividing the amount of DNA (ng) required to make 43.5 fmol of DNA (calculated by using the Weight to Molar Quantity Calculator for Nucleic Acids [https://www.bioline.com/media/calculator/01_07.html]) by the molecular weight (kb) of each amplicon. This was then divided by the concentration (ng/µl) of each amplicon to calculate the volume of amplicon (µl) needed to be added make the final equimolar mixture for a FTD sample, i.e., volume of each amplicon needed (µl) = $\frac{Amount\ of\ DNA\ for\ 43.5\ fmol}{Molecular\ weight\ of\ amplicon \times Amplicon\ DNA\ Concentration}$.

3.2.5 Double capture-based method of duplex library preparation – Protocol 1 (with in-house biotinylated baits)

Protocol 1 for double-capture of duplex libraries included a modified protocol for capturing DNA of interest from a homogenized mixture after duplex library preparation (Maricic, Whitten, and Pääbo 2010) and comprises of preparing duplex libraries from native genomic DNA, separately preparing a streptavidin beads-biotinylated 'baits' mixture that are complementary to our DNA libraries of interest, and then performing two rounds of capture of duplex libraries with the beads-baits mixture by performing the appropriate incubation and washing steps. The final libraries resulting from these two rounds of capture can then be sequenced on the appropriate NGS platforms with an aim to determine true mutations present in the samples with higher accuracy.

3.2.5.1 Duplex Library Preparation

Duplex libraries were prepared as described in Chapter 3, section 3.2.3 using 175 ng of native FTD case/control DNA as starting material.

After the post-ligation selection using 0.7X Ampure beads, 40 ng of adapter-ligated DNA was taken as the input for one PCR reaction, and multiple reactions made until all the DNA was used and amplified for 9 cycles using the PCR conditions as described in **Table 15**.

3.2.5.2 Biotinylated baits preparation

a) Biotinylated PCR primers

A separate set of primers with a 5' 'Biosg' modification were designed and made with the same primer sequences for the MAPT amplicons described in (**Table 20**). These biotinylated primers, along with the non-biotinylated primers, were used to make six 'baits' for the three MAPT amplicons for the double-capture protocol (**Table 23**):

Amplicon	Baits	Forward Primer (F) (10 µl)	Reverse Primer (R) (10 µl)
MAPT-A	MAPT-A (FB + R)	MAPT-A F-Biotin	MAPT-A R
	MAPT-A (F + RB)	MAPT-A F	MAPT-A R-Biotin
MAPT-B	MAPT-B (FB + R)	MAPT-B F-Biotin	MAPT-B R
	MAPT-B (F + RB)	MAPT-B F	MAPT-B R-Biotin
MAPT-C	MAPT-C (FB + R)	MAPT-C F-Biotin	MAPT-C R
	MAPT-C (F + RB)	MAPT-C F	MAPT-C R-Biotin

Table 23: Description of primers used to make baits DNA.

b) PCR programs

Eight 50 µl PCR reactions were made for each of the six baits described above using the PCR programme for PrimeSTAR GXL DNA Polymerase (R050A, Takara) (**Table 22**).

3.2.5.3 Buffers and solutions

The buffers need to be freshly prepared before application in the double-capture protocol (**Table 24**).

Buffer/Solution	Function of buffer/solution	Components of working stock	Temperature requirements before use in protocol (^{0}C)
Bind-Wash-Tween (BWT) buffer (1X)	Stringent wash buffer to remove non-specific binding	50 ml Tris + 0.1% Tween20 + 2 M NaCl	65
Wash-Tween (TET) buffer (1X)	Wash buffer to remove non-specific binding	50 ml Tris + 0.05% Tween20	65
125 mM NaOH	Mild DNA denaturant to make DNA single-stranded	Diluted from 1 M NaOH in aliquots as needed (5 µl needed for one round of capture per sample)	Room temperature

Table 24: Protocol for the preparation of buffers and solutions for double-capture protocol 1.

3.2.5.4 Baits-beads mixture preparation

A. Biotinylated baits

The three bait pairs for the three amplicons were mixed in equimolar proportions such that a total of 1 µg of baits was present in a 0.2 ml PCR tube for one round of capture per sample:

Calculation for pooling baits in equimolar proportions: The molar concentration (nM) of each bait was calculated by entering the DNA concentration (ng/µl) and the molecular weight (kb) of the bait into a Weight to Molar Quantity Calculator for Nucleic Acids (https://www.bioline.com/media/calculator/01_07.html). Next, the amount of DNA (ng) required for 50 nmol of each bait is calculated on the same calculator, and this weight (ng) is divided by the DNA concentration (ng/µl) of each bait to arrive at the DNA volume (µl) to be used as input to have all the baits in equimolar proportions.

Table 25 details the calculations done in one of the baits-mixture preparations done in equimolar proportions:

Bait	Mol. Wt. (kb)	DNA concentration (ng/µl)	DNA Molarity (nM)	Amount of DNA (ng) required for 50 nmol bait	DNA Volume (µl)
MAPT-A (FB + R)	1.209	47.00	59.90	39.23	0.83
MAPT-A (F + RB)	1.209	36.80	46.90	39.23	1.07
MAPT-B (FB + R)	6.372	79.40	19.20	206.77	2.60
MAPT-B (F + RB)	6.372	68.00	16.44	206.77	3.04
MAPT-C (FB + R)	6.868	72.00	16.15	222.91	3.10
MAPT-C (F + RB)	6.868	79.00	17.72	222.91	2.82
			Total Baits:	**937.87**	**13.46**
			Nuclease-free water (µl):		6.54
			Final Baits vol. (µl):		**20.00**

Table 25: Calculations for combining all six MAPT in-house baits in equimolar proportions.

Volume of baits as input (µl) = $\frac{Amount\ of\ DNA\ required\ for\ 50\ nmol\ of\ bait\ (ng)}{DNA\ concentration\ (\frac{ng}{\mu l})}$.

FB + R: Forward-biotin labelled primer + Reverse primer; F + RB: Forward primer + Reverse-biotin labelled primer; MAPT; Refer to *Table 23* for the detailed description of the MAPT primers.

20 µl of BWT buffer was added to the combined baits and the mixture was heated to 98^0C for 5 min in a thermocycler. Baits are denatured in this step and immediately transferred to ice and kept cold to prevent re-hybridization. In the meanwhile, beads are prepared for the next step.

B. Streptavidin Beads

Principle: Dynabeads® M-270 Streptavidin (cat. no.: 65305, ThermoFisher Scientific) are uniform and superparamagnetic beads which are 2.8 µm in diameter, with a monolayer, not a multilayer, of recombinant streptavidin covalently coupled to the surface so that the vast majority of the biotin binding sites are sterically available for binding, not only of free biotin, but also for binding of biotinylated ligands/targets. They are hydrophilic, negatively charged, and show rapid liquid-phase reaction kinetics.

Protocol: Streptavidin beads equilibrated to room temperature before use. For one sample, 30 µl of these beads was taken in a 0.2 ml PCR tube and washed with 200 µl of heated BWT buffer (washing comprises of resuspending beads in the buffer, collecting the beads with a magnet on the tube wall, and then discarding the supernatant). The beads were then washed with 200 µl of heated TET buffer. After discarding the supernatant, the baits from ice were quickly added to the dried streptavidin beads and mixed until they were resuspended in the mixture.

C. Baits-beads mixture incubation

The resultant baits-beads mixture was placed in a plate-shaker run at 2000 rpm at room temperature for about 60 min to allow the biotinylated baits to bind to the streptavidin beads. The mixture was vortexed gently every 30 min to resuspend aggregated beads at the bottom of the tube. After a minimum of one hour, the tubes were placed on a magnetic stand and the supernatant with unbound baits was discarded. The beads in the tube were then washed with 200 µl of heated BWT buffer a total of three times, and then this was repeated with 200 µl of heated TET buffer a total of three times. After the sixth wash, the beads were resuspended in 50 µl of heated TET buffer and stored in ice until needed for the next step.

3.2.5.5 Hybridization Mixture

1. Blocking oligos

The hybridization mixture has ingredients that foster the hybridization of duplex libraries with the streptavidin beads-bound biotinylated baits. However, the need for denaturing all the DNA molecules before such hybridization could lead to the cross-linking of the duplex library adapters with each other that could lead to the undesired capture of non-target library molecules. The use of blocking oligos designed to hybridize to the P5 and P7 primer sequences on the duplex adapters (except for the index sequence) prevents this from taking place.

Blocking oligos were designed to 'block' the P5 and P7 primers of duplex adapters in the system to prevent the capture of duplex libraries with non-target sequences and have been described in (**Table 26**). These blocking oligos are also phosphorylated at their 3' ends to prevent them from serving as primers in downstream PCR amplifications of the enriched duplex libraries.

Blocking oligonucleotide	Sequence (5' end to 3' end)
BO1.P5.F (200 µM)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-phosphate
BO2.P5.R (200 µM)	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-phosphate
BO3.P7.part1.F (200 µM)	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-phosphate
BO4.P7.part1.R (200 µM)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-phosphate
BO5.P7.part2.F (200 µM)	ATCTCGTATGCCGTCTTCTGCTTG-phosphate
BO6.P7.part2.R (200 µM)	CAAGCAGAAGACGGCATACGAGAT-phosphate

Table 26: List of blocking oligos used in the hybridization mixture for capture.

2. Hybridization mixture preparation

A hybridization mixture was made in a 0.2 ml PCR tube per sample with 300 ng of pooled duplex libraries, Agilent hybridization buffer (2X) and blocking agent (10X), and blocking oligos (**Table 27**).

Contents	Volume (µl)
300 ng of pooled duplex libraries	17.8
BO1.P5.F (200 µM)	0.5
BO2.P5.R (200 µM)	0.5
BO3.P7.part1.F (200 µM)	0.5
BO4.P7.part1.R (200 µM)	0.5
BO5.P7.part2.F (200 µM)	0.5
BO6.P7.part2.R (200 µM)	0.5
Agilent Blocking Agent (10X)	5.2
Agilent Hybridization Buffer (2X)	26.0
Total	**52**

Table 27: Contents of Hybridization Mixture I per sample.

This mixture was heated to 98^0C for 5 min in a thermocycler. DNA molecules are denatured in this step and immediately transferred to ice and kept cold to prevent re-hybridization until the baits-beads incubation is done.

3.2.5.6 Capture I

The beads-baits mixture was placed on a magnetic stand and the supernatant discarded after the beads were collected along the tube wall. The hybridization mixture 1 was transferred to the tube on the magnet, the tube then taken off from the stand, and the mixture pipetted until the beads were re-suspended. This baits-beads-hybridization mixture was incubated overnight at 65^0C in an appropriate thermocycler. The tubes were vortexed gently for a few seconds every 15 min for the first hour so that the beads could remain suspended in the mixture.

3.2.5.7 Post-capture I washing

The next morning, the baits-beads-hybridization mixture was placed on a magnetic stand and the supernatant discarded. The beads in the tube were then washed with 200 µl of heated BWT buffer a total of five times, and then this was repeated with 200 µl of heated TET buffer a total of five times. After the tenth wash, the beads were resuspended in 100 µl of heated TET buffer and transferred to a new PCR tube. This was placed on the magnetic stand and the supernatant was discarded. The pre-prepared master-mix of the 25 µl GXL polymerase-

mediated PCR (**Table 14**) was added immediately into the tube with the 'dry' beads and amplified for 6 cycles.

Post-amplification purification was done with 0.7X Ampure beads and 300 ng of the resultant DNA product was taken as the input for the second round of capture in hybridization mixture 2.

3.2.5.8 Capture 2

The biotinylated baits-streptavidin beads mixture was prepared for the second capture as described in Chapter 3, section 3.2.5.4.

A hybridization mixture 2 was prepared with 300 ng of pooled duplex libraries from first capture, Agilent hybridization buffer (2X) and blocking agent (10X), and blocking oligos as described in Chapter 3, section 3.2.5.5.

The hybridization mixture 2 was similarly added to the 'dry' baits-beads tube on the magnetic stand as described in Chapter 3, section 3.2.5.6, and the second capture was performed with an overnight incubation at 65^0C.

The next morning, post-capture 2 washing was done by placing the baits-beads-hybridization mixture on a magnetic stand and discarding the supernatant. The beads in the tube were then washed with 200 µl of heated BWT buffer a total of five times, and then this was repeated with 200 µl of heated TET buffer a total of five times. After the tenth wash, the beads were resuspended in 100 µl of heated TET buffer and transferred to a new PCR tube. This was placed on the magnetic stand and the supernatant was discarded. The beads were then resuspended in 5 µl of 125 mM sodium hydroxide and vortexed for 10 sec. This was incubated at room temperature for 1 min before placing the tube again on the magnetic stand. The supernatant was transferred to a new PCR tube and the beads were discarded. The DNA eluted into this 5 µl of supernatant was immediately placed in the ice.

4 amol of the eluted DNA was used for a 25 µl GXL polymerase-mediated PCR (**Table 14**) and amplification for 20 cycles (**Table 15**). The DNA obtained post-amplification was purified with 0.7X Ampure beads as described in section 3.2.3.6, performed qPCR as described in section 3.2.3.7, and then sent for sequencing on Illumina MiSeq.

3.2.6 Double capture-based method of duplex library preparation – Protocol 2 (with customized IDT probes)

Protocol 2 for double-capture of duplex libraries was taken from the Tube protocol recommended by IDT Technologies and comprises of preparing duplex libraries from native genomic DNA, adding IDT blocking oligos to duplex DNA which is followed by lyophilization, adding biotinylated 'probes' that are complementary to our DNA libraries of interest (custom-designed by IDT) to the lyophilized DNA, and then performing a sixteen-hour capture of duplex libraries with the probes followed by incubation with streptavidin beads the appropriate heated and room temperature washes. This process is done a total of two times and the final libraries resulting from these two rounds of capture can then be sequenced on the appropriate NGS platforms with an aim to determine true mutations present in the samples with higher accuracy.

3.2.6.1 Duplex library preparation

Duplex libraries were prepared as described in Chapter 3, section 3.2.3 using 500 ng of native FTD case/control DNA as starting material. The use of less than 500 ng of native DNA as starting material results in lower capture efficiency.

After the post-ligation selection using 0.7X Ampure beads, 240 ng of adapter-ligated DNA was taken the input for one PCR reaction and amplified for 9 cycles using the PCR conditions as described in **Table 15** and post-amplification selection was done with 1.2X Ampure beads.

3.2.6.2 Probes

The biotinylated probes used for the capture were custom-designed by IDT technologies after submitting a request to cover both, the '+' and '–' strands, of the entire MAPT gene sequence [GRCh38 reference genome – chromosome 17:45894554 to 17:46028334]. For our purpose here, IDT was able to design Discovery Pools (16 reactions) with 845 'good' probes for the '+' strand and 850 'good' probes for the '–' strand after discarding the 'risk' and 'remove' probes that covered the repetitive intronic regions that pose difficulties in sequencing.

For small panels (<1000 probes), the IDT protocol recommends longer hybridization times of up to 16 hours to improve the performance of the reaction.

3.2.6.3 Hybridization reaction between duplex libraries and probes

For one sample, 5 µl of Human Cot DNA (provided in the kit) and 2 µl of xGen Universal Blockers (cat. no.: 1075474, Integrated DNA Technologies) were taken in a 1.5 ml Eppendorf tube to which 500 ng of duplex libraries was added before lyophilization in a SpeedVac system (Savant SPD1Z1P SpeedVac Concentrator, Thermo Scientific). Again, the xGen Blockers have been designed to be complementary to the P5 and P7 primers of our duplex adapters to 'block' these sequences and prevent the capture of duplex libraries with non-targeted sequences.

All the contents of the xGen Hybridization and Wash kit (cat. no. 1080577, Integrated DNA Technologies) were thawed to room temperature. A hybridization master-mix was created by adding 8.5 µl of xGen 2X Hybridization Buffer (provided in the kit), 2.7 µl of xGen Hybridization Buffer Enhancer (provided in the kit), and 3 µl each of the '+' and '–' strand probes. This hybridization buffer master-mix was added directly to the tube with lyophilized DNA. The mix was pipetted well and incubated at room temperature for 10 min. Following a brief vortex and centrifuge, the components for capture were transferred to a 0.2 ml PCR tube which was then placed in thermocycler set to the HYB programme (**Table 28**).

The optimization experiments for the IDT protocol had slightly different ingredients in the hybridization master-mix in that 8.5 µl of xGen 2X Hybridization Buffer (provided in the kit), 2.7 µl of xGen Hybridization Buffer Enhancer (provided in the kit), 4 µl each of the '+' strand probes (as recommended in the protocol) and 2 µl of nuclease-free water were added.

Temperature (^{0}C)	Time
HYB programme (Lid set at 100^{0}C)	
95	30 sec
65	16 hrs
65	Hold
WASH programme (Lid set at 70^{0}C)	
65	Hold

Table 28: Thermocycler programmes for IDT capture protocol.

3.2.6.4 Preparation of buffers

1X working solutions of the xGen 2X Bead Wash Buffer, xGen 10X Wash Buffers 1, 2, and 3, and xGen 10X Stringent Wash Buffer (all stock buffers provided in the kit) was made as indicated in the kit before beginning Day 2 of the capture protocol. A Bead Resuspension Mix with 8.5 µl of xGen 2X Hybridization Buffer (provided in the kit), 2.7 µl of xGen

Hybridization Buffer Enhancer (provided in the kit), and 5.8 µl of nuclease-free water was also prepared.

3.2.6.5 Streptavidin beads wash

Streptavidin beads are to be equilibrated to room temperature before beginning the washes. After thoroughly vortexing the beads for 15 sec, 50 µl of the beads were taken in a PCR tube for a single capture of a sample. 100 µl of the 1X Bead Wash Buffer was added to the beads and pipetted many times to mix well. The tube was placed on a magnetic stand to allow the beads to separate completely from the supernatant which was then discarded when clear. Another 100 µl of the 1X Bead Wash Buffer was added to the 'dried' beads on the tube wall and this was repeated a total of three times. After the third wash, 17 µl of the Bead Resuspension Mix was added to the 'dried' beads, mixed well, and briefly centrifuged. The 17 µl of resuspended beads were aliquoted into new PCR tubes for each capture reaction.

3.2.6.6 Bead capture

After the 16 hrs incubation, the hybridized DNA sample was taken out of the thermocycler and the 17 µl of resuspended beads were transferred to this tube. This beads-hybridized DNA mixture was vortexed well and gently centrifuged. The tube was again placed in the thermocycler on the WASH programme (**Table 28**) for 45 min. The tube was gently vortexed every 10-12 min to ensure the sample remains fully resuspended. At the end of the 45 min, the samples were taken immediately for the heated washes.

3.2.6.7 Washes

One of the two aliquots of the 1X Wash Buffer 1 (110 µl) and both the aliquots of the 1X Stringent Wash Buffer (160 µl each) were pre-heated to 65^0C before starting the washes.

3.2.6.8 Heated Washes

100 µl of the heated 1X Wash Buffer 1 was added to the tube from the end of the 45 min incubation, mixed well carefully without bubble formation, and placed on the magnetic stand and the supernatant discarded. 150 µl of the heated 1X Stringent Wash Buffer was added to mix the beads carefully without bubble formation. The tube was then incubated in a thermocycler at 65^0C for 5 min, after which the tube was placed in a magnetic stand and the supernatant discarded. This wash with the 1X Stringent Wash Buffer was repeated a total of two times before proceeding to the next set of washes.

3.2.6.9 Room temperature washes

150 µl of the remaining room temperature aliquot of the 1X Wash Buffer 1 was added to the 'dried' beads from the end of the heated washes and vortexed thoroughly. The tube was incubated at room temperature while alternating between 30 sec of vortexing followed by 30 sec of rest before a brief centrifuge. The tube was then placed on a magnetic stand and the supernatant was discarded. Similarly, the washes with 150 µl of 1X Wash Buffers 2 and 3 were immediately performed. Care was taken to remove any residual supernatant after the wash with Wash Buffer 3.

20 µl of nuclease-free water was added to a capture and mixed well.

3.2.6.10 Post-capture 1 PCR amplification

A 30 µl of Amplification Reaction Mix was prepared with 25 µl of the 2X KAPA HiFi HotStart ReadyMix, 1 µl of 50 µM P5 primer, 1 µl of 50 µM of Test primer, and 3 µl of nuclease-free water. The entire 20 µl of the resuspended beads with captured DNA was used to perform post-capture 1 library amplification. The PCR programme described in (**Table 29**) (with P7 primer used here) was used to amplify the DNA for 16 cycles for Capture 1 which was then followed by 1.2X Ampure beads purification.

Contents (For one reaction)	**GXL DNA Polymerase Volume (µl)**
2X KAPA HiFi ReadyMix	25
P5 (50 µM)	1.0
P7 (50 µM)/ Index primer	1.0
DNA	Variable
Nuclease-free water	16.5 – DNA vol.
Total	**30**

Table 29: PCR tube contents for one reaction of duplex library amplification post-capture 2.
*Using any PCR master-mix other than KAPA HiFi yielded no results.

The DNA end-product of this amplification-purification was taken as the starting material for the second hybridization reaction of duplex libraries and probes.

3.2.6.11 Second round of capture

5 µl of Human Cot DNA (provided in the kit) and 2 µl of xGen Universal Blockers (cat. no.: 1075474, Integrated DNA Technologies) were taken in a 1.5 ml Eppendorf tube to which all of DNA from the end of the first capture was added before lyophilization in a SpeedVac system (Savant SPD1Z1P SpeedVac Concentrator, Thermo Scientific).

The remaining steps of the hybridization reaction and the16-hour incubation were followed as described in Chapter 3, section 3.2.6.3.

Preparation of buffers, Streptavidin beads wash, bead capture, and the heated and room temperature washes were performed as described in Chapter 3, sections 3.2.6.4 to 3.2.6.9.

3.2.6.12 Post-capture 2 PCR amplification

A 30 µl of Amplification Reaction Mix was prepared with 25 µl of the 2X KAPA HiFi HotStart ReadyMix, 1 µl of 50 µM P5 primer, 1 µl of Index primer, and 3 µl of nuclease-free water. Only 1 µl of the resuspended beads with captured DNA was used to perform post-capture 2 library amplification. The PCR programme described in section 3.2.6.10 (**Table 29**) was used to amplify the DNA for 7-8 cycles for Capture 1 which was then followed by 1.2X Ampure beads purification, before the samples for sent for sequencing on Illumina Novaseq

3.3 Analyses

The duplex libraries prepared were submitted for sequencing on the Illumina NovaSeq platform and then analysed using a well-designed computational workflow. All the bioinformatic and statistical analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

3.3.1 Duplex sequencing computational workflow

The computational workflow for DS uses a number of standard software packages to process the sequencing data (see section 3.1.4 above) as well as several custom Python scripts as described in (Kennedy et al. 2014).

The computational workflow is broken down into three major steps: (i) Tag parsing and initial alignment; (ii) single-strand consensus sequences (SSCS) assembly; and (iii) duplex consensus sequences (DCS) assembly. The latest version of the DS software package was downloaded from: https://github.com/loeblab/Duplex-Sequencing. The step-by-step instructions of the computational workflow was followed as described in (Kennedy et al. 2014) and a brief summary of the steps have been described below (**Fig. 9**).

3.3.1.1 Preparing a reference genome for use

Firstly, the reference genome of interest was downloaded and decompressed. For our analyses, the hg19 reference genome from the University of California, Santa Cruz was used.

The reference genome was indexed with BWA by following the instructions recommended on their official website (http://bio-bwa.sourceforge.net/bwa.shtml).

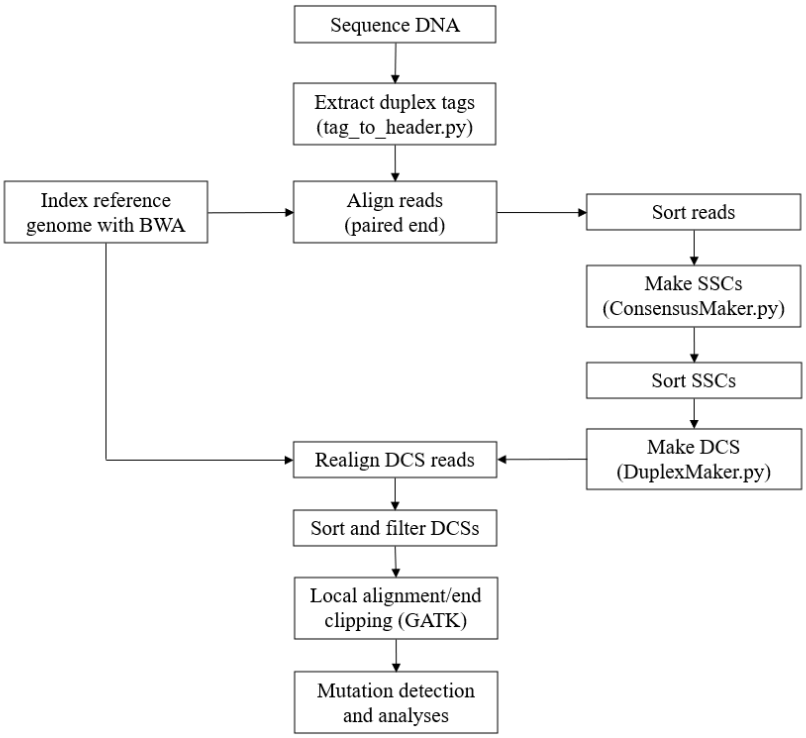


Figure 9: Schematic of the basic computational workflow for duplex sequencing.
BWA: Burrows-Wheeler Alignment; DCS: Duplex consensus sequences; DNA: Deoxyribonucleic acid; GATK: Genome Analysis Tool kit; SSCS: Single strand consensus sequences.

3.3.1.2 Parsing and filtering duplex tags

The 12-nucleotide tag sequence was concatenated from the paired reads and evaluated for the tag quality using the 'tag_to_header.py' script of the DS software package. The resultant reads were aligned to the reference genome to make a single paired-end .sam file, before converting it to .bam format and sorting by position.

3.3.1.3 Bioinformatics processing

1. Making SSCSs

The Python programme 'ConsensusMaker.py' was run to collapse the PCR duplicates into single strand consensus (SSCs) and was followed by sorting the single strand consensus sequence (SSCS) reads.

2. Making DCSs

Duplex consensus sequences (DCSs) were then constructed from SSCSs using 'DuplexMaker.py'. Each DCS .fastq was aligned to the reference genome and paired-end .sam files were made for the DCS data. This step was followed by converting the files to .bam format and sorting by position.

3. Preparing files for analyses

The final sorted DCS .bam files were indexed and all unmapped reads were filtered out.

A readgroups field was added to the header of the final DCS .bam files with Picard to allow for compatibility with the GATK using Picard tools. The final sorted DCS .bam files were indexed again (now with readgroups field). Local re-alignment of the reads was performed using GATK by first identifying the genome targets for local re-alignment followed by the actual local realignment.

A final end-trimming of five bases from both the 3' and 5' ends of each read was performed to reduce the occurrence of false-positives that occur during the enzymatic steps used during sequencing library preparation.

These processed DCS .bam files were now ready for any downstream processing desired, including, but not limited to, determining quality metrics of the reads derived, counting the number of unique mutations present in the final DCS sequences and calculating their frequencies, locating the genomic position of each mutation, etc.

3.3.2 Statistical Analyses

The statistical significance of the mutational burden seen in FTD-Tau disease cases compared to control was determined using Student's t-test within RStudio. p-value threshold for a set of data to be considered as statistically significant was set at 0.05 being: $*p < 0.05$ and $**p<0.001$.

The statistical significance of the VAF distributions seen within the data-set between FTD-Tau disease cases compared to control, Temp *vs* Occ *vs* Med, and between the three MAPT amplicons (in amplicon-based DS) were determined using Kolmogorov-Smirnov test within

RStudio. p-value threshold for a set of data to be considered as statistically significant was set at 0.05 being: *p < 0.05, **p<0.01 and ***p<0.001.

The calculations to determine the 'per-bp mutation rate per sample' as well as the 'per-bp mutation rate per genomic region per sample' were performed by the PhD candidate under the guidance of Dr Claudia Calabrese from Prof Patrick Chinnery's lab at the University of Cambridge, UK and these have been described in detail in the last Appendix of the thesis.

3.4 Summary of the different approaches to DS explored in the thesis

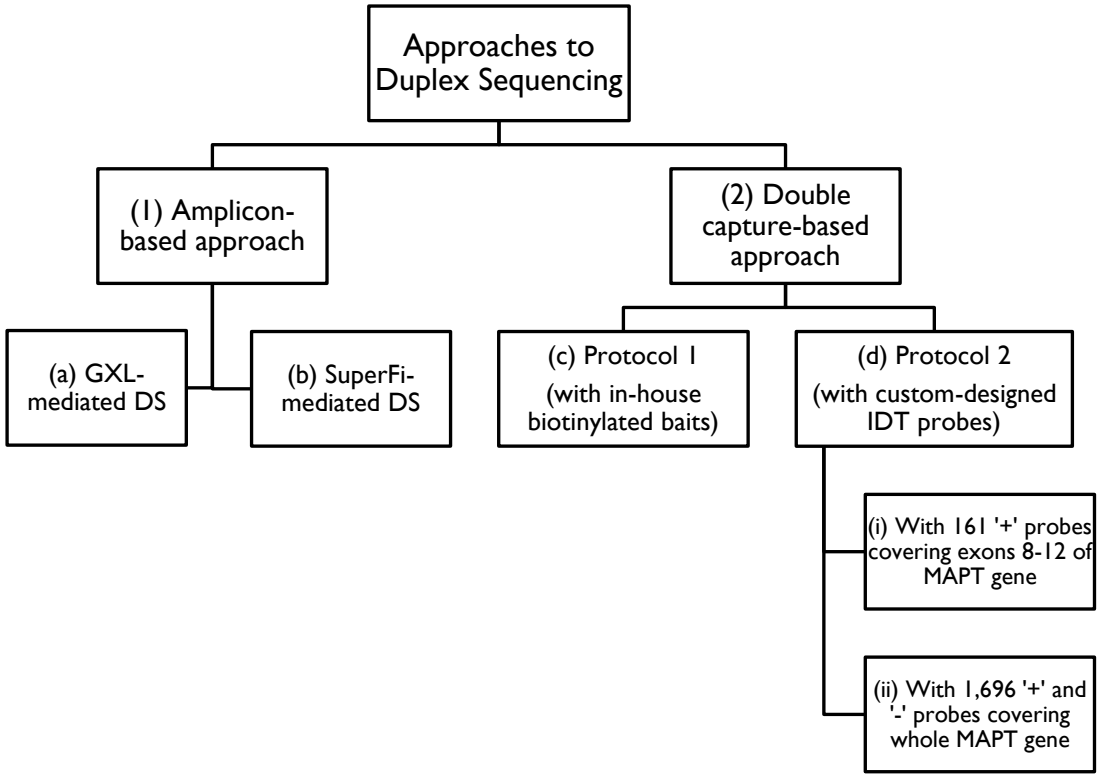


Figure 10: Flowchart to demonstrate the different approaches to duplex sequencing explored in this thesis.

The present thesis describes the wet-lab results and the computational analyses of the different approaches to duplex sequencing: (1) amplicon-based approaches: (a) GXL-mediated and (b) SuperFi-mediated; and (2) double capture-based approaches: (c) Protocol 1 (with in-house biotinylated baits) and (d) Protocol 2 (with custom-designed IDT probes) – (i) With only '+' strand probes covering exons 8-12 pf MAPT gene (as a preliminary test to optimize the protocol) and (ii) With probes for '+' and '-' strands of the whole MAPT gene (for all six FTD-Tau samples with the Temp, Occ, and Med brain regions of the disease and age-matched controls)

DS: Duplex sequencing; FTD: Frontotemporal Dementia; MAPT: Microtubule-associated protein Tau; Med: Medulla; Occ: Occipital lobe; Temp: Temporal lobe.

The flowchart above is a summary of all the approaches to DS investigated in this project to highlight its scope to detect mutations in FTD-Tau samples (**Fig. 10**) and this is followed by a tabulation of the advantages and drawbacks of two main sequencing approaches employed in the present thesis (**Table 30**).

Amplicon-based Approach	Double capture-based Approach
+ Less time (~1.5 days in total) is needed to prepare the duplex libraries that is also a low labour-intensive process. + A little amount of native DNA (as low as 50 ng) is enough as the input for library preparation. + It is possible to sequence the exact regions of interest since the amplicon can be designed as needed. + This approach is better-suited to sequence and analyse smaller regions of interest that are ≤ 10 kb in length.	+ Duplex libraries can be sequenced with high accuracy without any fears of introducing any untraceable PCR errors since this approach does not have a pre-library amplification step. + This approach is better-suited for larger areas of interest that are > 10 kb in length, with longer sequences giving better results. +Commercial kits with the ability to capture DNA from multiple samples at a time (96-well-based method) with higher reproducibility are also available.
– There exists a possibility of introducing untraceable PCR errors in the pre-library amplification step of library preparation. – This approach can handle only a few samples at a time based on the experience of the user since this is a novel approach proposed to perform DS.	– The duplex libraries require longer preparation times (~3.5-4.0 days in total) that is also a very labour-intensive process. – Higher DNA quantities of at least 500-1000 ng of native DNA will be necessary for the duplex library preparation process in this approach.

Table 30: Advantages and disadvantages of each sequencing approach to DS employed in this thesis.

Chapter IV: Optimization of Amplicon-based Duplex Sequencing

4.1 Introduction

The preparation of duplex libraries, as described in (Kennedy et al. 2014), is similar to the standard Illumina library preparation protocol and follows the steps of DNA shearing by enzymatic fragmentation, end repair and 3' dA-tailing, adapter ligation, PCR amplification, size selection with Ampure XP beads before and after the amplification step, and finally, targeted capture. Adapter ligation allows specialized duplex tagged adapters to be appended to either end of the DNA of interest that helps to later identify the "parent" DNA sequence of each read during data analyses while optimal family size, i.e., the number of reads present in the system with duplex consensus (DCS), is determined by adjusting the DNA input amount and number of cycles for the PCR amplification step, thus, making these steps the most crucial for the duplex library preparation process. The paper recommends the peak optimal family size (reads with duplex consensus) to be 6 and the entire process requires ~1-3 µg of DNA for targeted duplex sequencing that could then be captured by a long and laborious procedure. The final duplex libraries sent for sequencing must show a single sharp peak on the TapeStation Bioanalyzer with a DNA fragment size of ~300-400 bp (75 bp of adapters ligated on either end of ~200 bp DNA fragments of interest); the *TapeStation Bioanalyzer* is an automated electrophoresis machine routinely used in NGS experiments for the sample quality control of DNA samples wherein the system integrates the instrument, data processing software, reagents, and the ScreenTape devices, specifically for the DNA sample in question.

Our first approach for duplex sequencing attempted to circumvent the need for the optimization of the post-library preparation capture and the large DNA input by first amplifying our regions of interest with long-range PCR, combining them in equimolar proportions before proceeding the KAPA protocol for duplex sequencing (see Chapter 3, section 3.2.4). This resulted in a short and straight-forward protocol that needed a preparation time of no more than 1.5 days **(Fig. 11)**.

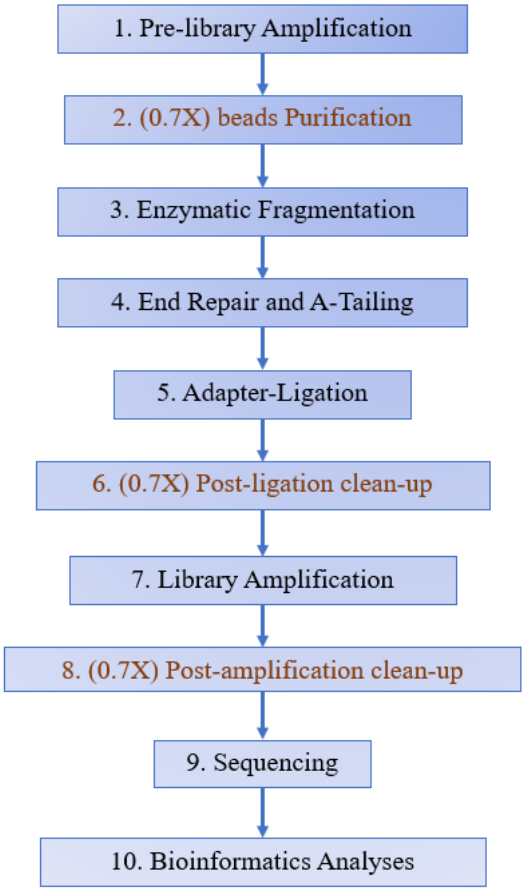


Figure 11: Flowchart of amplicon-based duplex library preparation.
Enzymatic fragmentation, End-Repair and A-tailing, Adapter-ligation, post-ligation clean-up, Library amplification, and post-amplification clean-up done as per KAPA HyperPlus protocol. Steps highlighted in brown are check-points to verify library preparation on TapeStation Bioanalyzer 2200

For our experiments, we performed the pre-library preparation amplification with two different DNA polymerases – (1) Takara PrimeSTAR GXL DNA polymerase, the conventional polymerase used routinely for long-range PCR amplifications (Control), and (2) Platinum SuperFi II DNA polymerase which is thought to have a higher fidelity rate as compared to the GXL polymerase (Test) **(Fig. 12).**

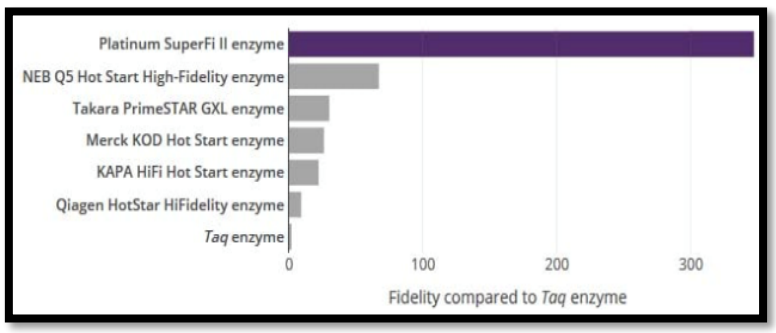


Figure 12: Fidelity rate of commonly-used DNA polymerases.
Comparison of the fidelity of the Invitrogen™ Platinum™ SuperFi™ II DNA Polymerase to other DNA polymerases. Fidelity of DNA polymerases was measured by NGS and normalized to Taq DNA polymerase. Adapted from (Thermo Fisher Scientific 2020).
DNA: Deoxyribonucleic acid; NGS: Next generation sequencing.

This chapter delves into the optimization of the PCR programmes for the preparation of MAPT amplicons A, B, and C (**Table 20**) for GXL and SuperFi-based library preparations necessary for amplicon-based duplex sequencing. Then, it gives a summary of all the optimization steps done for GXL-based as well as SuperFi-based duplex library preparations comprising of the two main parameters: (i) PCR amplification – DNA input, and (ii) PCR amplification – number of cycles, along with a brief review of the qPCR verification design. It concludes with a comparison of the family size and mutation detection analyses in duplex libraries derived from both the DNA polymerases to determine the superior polymerase for the amplicon-based DS method.

4.2 Results

4.2.1 Preparations of MAPT amplicons

4.2.1.1 GXL DNA polymerase-based amplified products

The main parameters to be optimized for PCR amplification for each of the three MAPT amplicons previously described in Chapter 3, section 3.2.4.1a, were: (i) annealing temperature, (ii) extension time, and (iii) cycle number, with an aim to obtain a single band on a 0.7x agarose gel. For GXL-based optimizations, 50 ng of cell line DNA with wildtype nuclear DNA was used as the template for the pre-library amplification.

a) MAPT-A amplicon preparation

Since the amplicon designed was only 1209 bp long, a straight-forward PCR previously established in the lab for short amplicons was used. The DNA input was 50 ng of cell line DNA and primers for MAPT-A were used (see Chapter 3, section 3.2.4.1b).

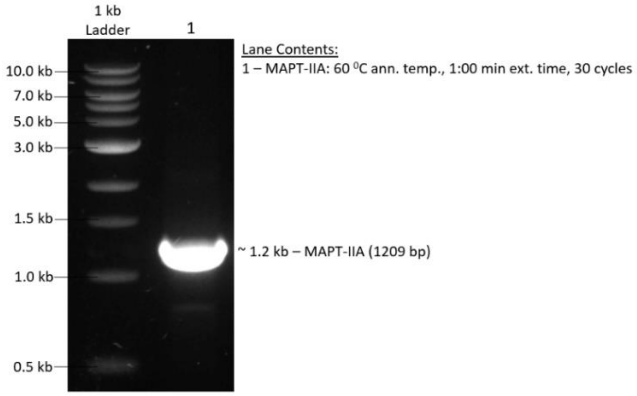


Figure 13: GXL-based MAPT-A amplicon optimization.

Gel migration for MAPT-A amplicon prepared with an annealing temperature of 60 ^{0}C and extension time of 1:00 min for 30 cycles during GXL-mediated PCR that yielded one single band within the expected range (1209 bp). Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; PCR: Polymerase chain reaction.

A single band formed at ~1.2 kb on the agarose gel **(Fig. 13)** (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol) indicated that the PCR programme used for the designed primers for MAPT-A were a success.

b) MAPT-B amplicon preparation

The MAPT-B amplicon was designed to be 6372 bp long, which necessitated the need for long-range PCR (LR-PCR). The DNA input was 50 ng of cell line DNA and primers for MAPT-B were used (see Chapter 3, section 3.2.4.1b), and several cases tested to settle upon the optimal PCR programme parameters:

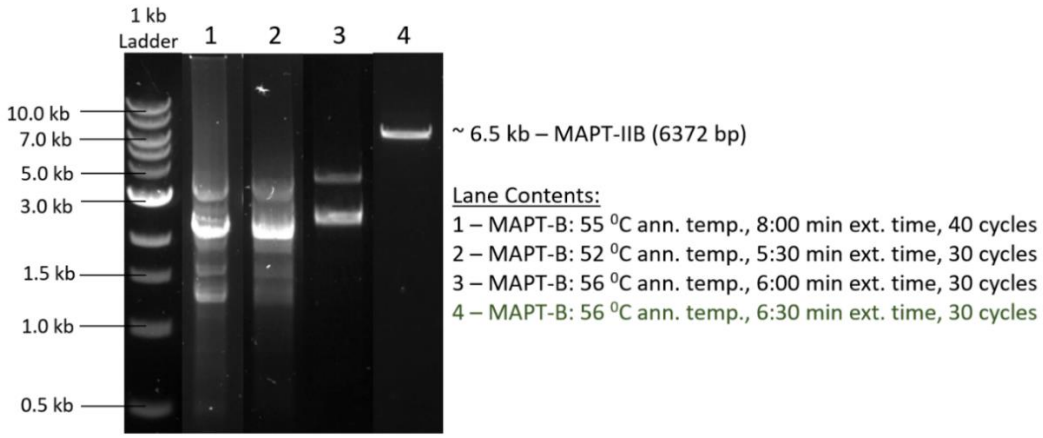


Figure 14: GXL-based MAPT-B amplicon optimizations.

Gel migration for MAPT-B amplicon prepared with: (1) An annealing temperature of 55 ^{0}C and extension time of 8:00 min for 40 cycles during GXL-mediated PCR that yielded multiple non-specific bands. (2) An annealing temperature of 52 ^{0}C and extension time of 5:30 min for 30 cycles during GXL-mediated PCR that yielded multiple non-specific bands. (3) An annealing temperature of 56 ^{0}C and extension time of 6:00 min for 30 cycles during GXL-mediated PCR that yielded multiple non-specific bands. (4) An annealing temperature of 56 ^{0}C and extension time of 6:30 min for 30 cycles during GXL-mediated PCR that yielded a single band within the expected range (6372 bp) [highlighted in **green font** to indicate successful amplification].

Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; PCR: Polymerase chain reaction.

Since lane 4 of **Fig. 14** shows a single band on 0.7x agarose gel (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol), we applied these parameters as the optimum to all GXL-based MAPT-B amplification (see Chapter 3, section 3.2.4.1b for full PCR programme).

c) MAPT-C amplicon preparation

The size of the MAPT-C amplicon, 6868 bp, is very close to that of the MAPT-B amplicon, and we tested if the PCR programme used for the latter could be applied to the last amplicon as well, with 50 ng of cell line DNA as input for amplification.

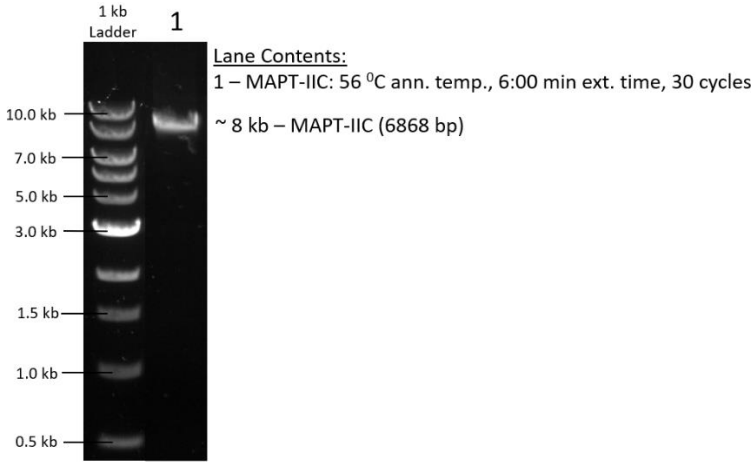


Figure 15: GXL-based MAPT-C amplicon optimization.

Gel migration for MAPT-C amplicon prepared with an annealing temperature of 56 ^{0}C and extension time of 6:00 min for 30 cycles during GXL-mediated PCR that yielded one single band within the expected range (6868 bp).

Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; PCR: Polymerase chain reaction.

A single band at ~8.0 kb on the agarose gel **(Fig. 15)** (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol) indicated that the PCR programme used was also successful for MAPT-C amplicon preparation.

All three amplicons could be successfully prepared with the primers designed using GXL DNA polymerase reagents and protocols ready for application to the brain samples.

4.2.1.2 SuperFi DNA polymerase-based amplified products

Similar strategies were used for SuperFi DNA polymerase-based amplicon preparation as the GXL-based amplicons. However, for SuperFi-based optimizations, 50 ng of FTD case/control DNA were used as the template to be amplified.

a) MAPT-A amplicon preparation

At first, the same PCR programme used for the GXL-based MAPT-A amplicon preparation was applied for this case as well; however, there was no band seen on the gel with a post-amplification Qubit reading of <1 ng/µl. The amplification improved when the extension time was extended to two minutes **(Fig. 16)**:

Figure 16: SuperFi-based MAPT-A amplicon optimization.

Gel migration for MAPT-A amplicon prepared with an annealing temperature of 60 ^{0}C and extension time of 2:00 min for 30 cycles during SuperFi-mediated PCR that yielded one single band within the expected range (1209 bp).

Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; PCR: Polymerase chain reaction.

The longer extension time resulted in a single band at ~1.2 kb on the agarose gel (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol), and the MAPT-A amplicon for C576 Med was ready to be used for duplex library preparation.

b) MAPT-B amplicon preparation

Similarly, the PCR programme used for GXL-based MAPT-B amplicon preparation was used for this case which also proved to be successful for SuperFi-based MAPT-B amplicon preparation (**Fig. 17**).

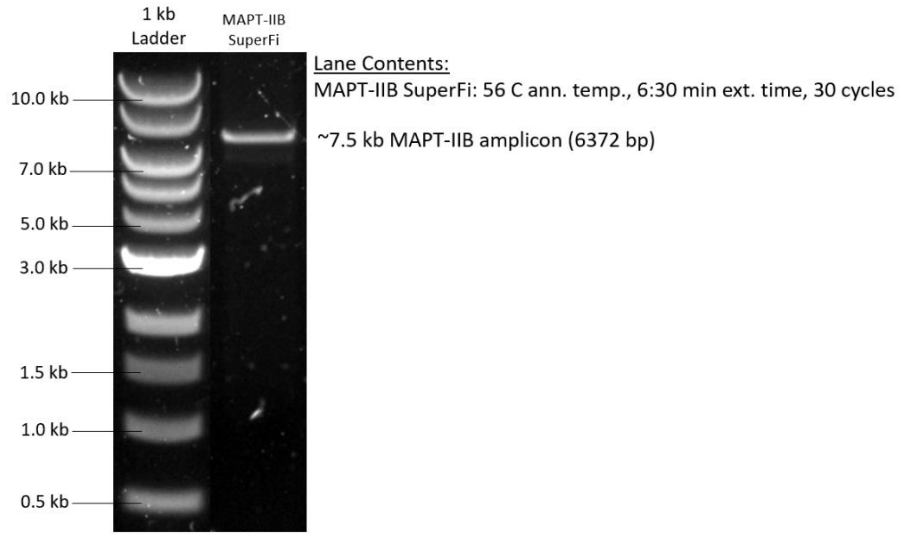


Figure 17: SuperFi-based MAPT-B amplicon optimization.

Gel migration for MAPT-B amplicon prepared with an annealing temperature of 56 ^{0}C and extension time of 6:30 min for 30 cycles during SuperFi-mediated PCR that yielded one single band within the expected range (6372 bp).

Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; PCR: Polymerase chain reaction.

The agarose gel (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol) showed a single band at ~7.5 kb which is within the expected range and the MAPT-B amplicon for C576 Med was ready to be used for duplex library preparation.

c) MAPT-C amplicon preparation

For SuperFi-based MAPT-C amplicon preparation, several optimization settings needed to be tested to obtain a single band in the expected region of interest:

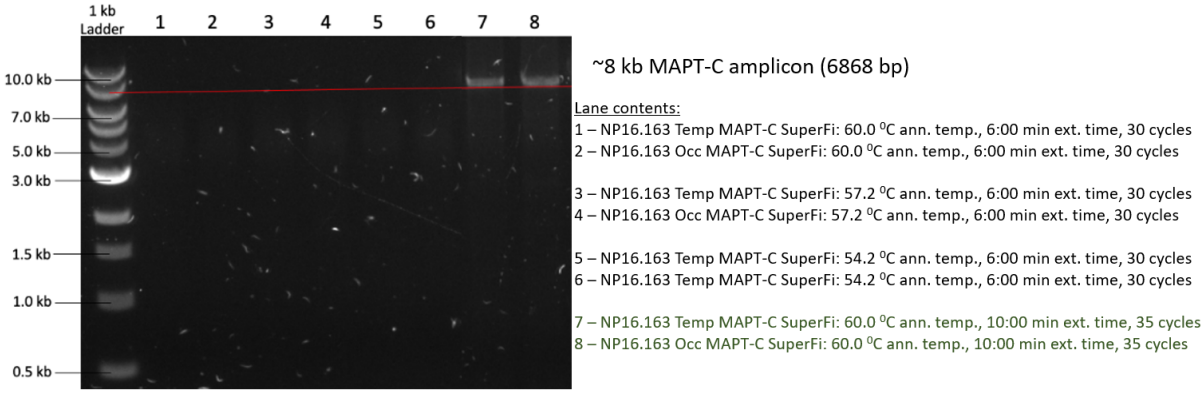


Figure 18: SuperFi-based MAPT-C amplicon optimization.

Gel migration for MAPT-C amplicon prepared with:
(1) An annealing temperature of 60.0 ^{0}C and extension time of 6:00 min for 30 cycles during SuperFi-mediated PCR that yielded no bands for NP16.163 Temp (Lane 1) and Occ (Lane 2).
(2) An annealing temperature of 57.2 ^{0}C and extension time of 6:00 min for 30 cycles during SuperFi-mediated PCR that yielded no bands for NP16.163 Temp (Lane 3) and Occ (Lane 4).
(3) An annealing temperature of 54.2 ^{0}C and extension time of 6:00 min for 30 cycles during SuperFi-mediated PCR that yielded no bands for NP16.163 Temp (Lane 5) and Occ (Lane 6).
(4) An annealing temperature of 60.0 ^{0}C and extension time of 10:00 min for 35 cycles during SuperFi-mediated PCR that yielded a single band within the expected range (6868 bp) for NP16.163 Temp (Lane 7) and Occ (Lane 8) [highlighted in **green font** to indicate successful amplification].
Ann. temp.: Annealing temperature; Ext. time: Extension time; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; min: minutes; NP1.163: FTD-Tau disease; PCR: Polymerase chain reaction; Occ: Occipital lobe; Temp: Temporal lobe.

The samples used to optimize for MAPT-C amplicon preparation were FTD-Tau patient (NP16.163) temporal lobe (Temp) and occipital lobe (Occ), and they both gave a single band at ~8 kb on a 0.7x agarose gel (see Chapter 3, section 3.2.4.2 for gel preparation and viewing protocol) only when a longer LR-PCR programme was used **(Fig. 18)**, indicating that the SuperFi-based MAPT-C amplicon preparation was now optimized.

Optimized amplification protocols for all three amplicons made using both, GXL and SuperFi DNA polymerases were ready for application to other samples for duplex library preparation.

4.2.2 Library amplification optimizations

Most of the amplicon-based duplex library preparation PCR optimization was done on an 852 bp amplicon derived from cell line DNA for another duplex sequencing project being developed at the time, and we continued with the same amplicon here since the same principles would still apply for MAPT gene amplicon-based optimized duplex library preparation. This was done to not only conserve all the reagents for MAPT library preparations, but also as a continuation of the optimizations that had already been done on the previous gene duplex

library preparation optimization at the time to ensure that we could standardize the entire protocol efficiently.

50 ng of genomic DNA was input to make the amplicon using GXL DNA polymerase reagents and the primers described in **(Table 31)**, and 50 ng of this amplicon was used for the KAPA HyperPlus protocol **(Fig. 19)**:

Forward primer (F)	5'-CTCTGACGTTCTCCTCTTCA-3'		
Reverse primer (R)	5'-TCTTTTTCCAGCTTGAGGGA-3'		
Step	**Temperature (^{0}C)**	**Time (min)**	**Cycles**
Pre-heating	94	1:00	x1
Denaturation	98	0:10	x40
Annealing	55	0:15	
Extension	68	0:30	
Final extension	72	7:00	x1
Hold	4	∞	-

Table 31: Primers used to amplify the amplicon with the PCR programme used.

A: Amplicon **B: Adapter-ligated DNA library**

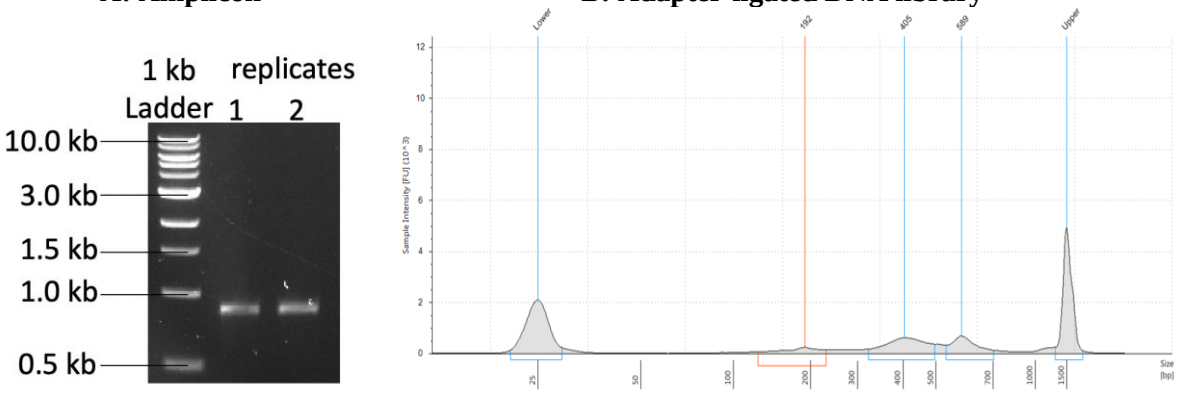


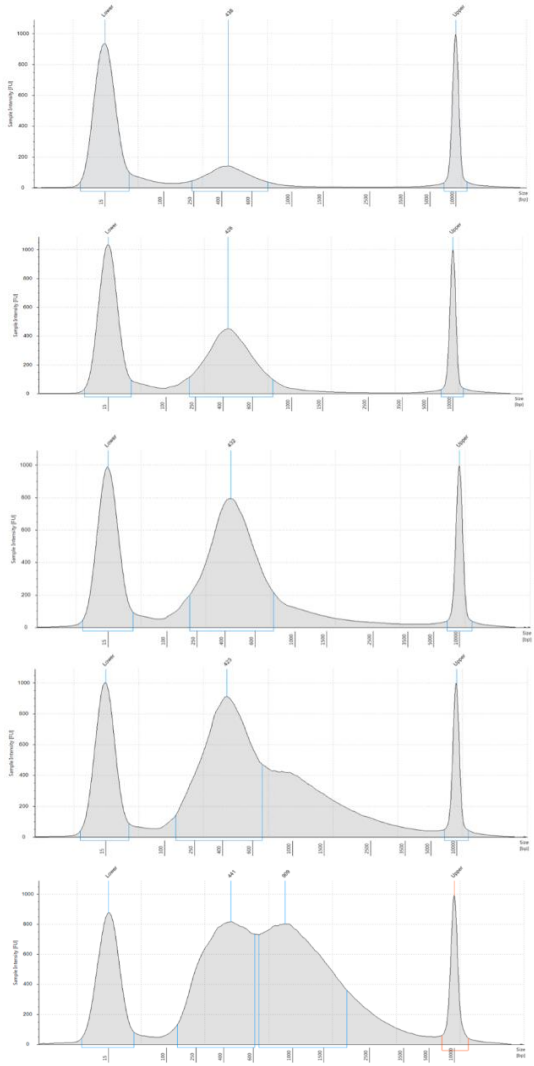
Figure 19: Amplicon-based duplex library preparation for library amplification optimizations. (A) *Pre-library amplification:* 852 bp amplicon made from cell line in two replicates (B) *Adapter-ligation and purification*: Three peaks seen as expected in the adapter-ligated DNA library (0.7X Ampure XP beads purification) on the HS D1000 ScreenTape.

4.2.2.1 Amplification cycle number – Optimizations

The number of cycles is an important variable for duplex library preparation – too many cycles can result in the creation of higher molecular weight non-specific PCR products seen as multiple peaks on the TapeStation Bioanalyzer whereas too few cycle numbers can result in not having enough number of molecular copies to have a good duplicative level that can form duplex consensus (DCS) reads.

(Kennedy et al. 2014) recommends an input DNA between 4-40 amol for PCR to obtain a certain number of reads per sample. 10 amol of input DNA was the lowest that had been tested in previous attempts of duplex library preparation and so, the following results were obtained

when 10 amol of adapter-ligated DNA (the standard at the time) **(Fig. 19b)** were amplified at varying cycle numbers:



Case 1: 16 cycles
DNA amount obtained at end = 9.6 ng

Case 2: 18 cycles
DNA amount obtained at end = 23.7 ng

Case 3: 20 cycles
DNA amount obtained at end = 57.9 ng

Case 4: 22 cycles
DNA amount obtained at end = 60.0 ng

Case 5: 24 cycles
DNA amount obtained at end = 86.1 ng

Figure 20: Library amplification cycle number optimizations
Five conditions tested to optimize the number of cycles during library amplifications: (1) 16 cycles (2) 18 cycles (3) 20 cycles (4) 22 cycles and (5) 24 cycles. Higher molecular weight non-specific fragment formation seen at cycle numbers greater than 20, suggesting 20 cycles is optimal. DNA amounts calculated from Qubit HS DNA Assay readings.

From **Fig. 20**, it is clear that as the number of amplification cycles increases, the peak size increases, along with a corresponding increase in final amount of purified, amplified adapter-ligated DNA end-product for the same amount of input DNA. At around 22 cycles, non-specific higher molecular weight products begin to form and this secondary peak size almost equals the size of our DNA of interest at 24 cycles.

Thus, the **optimum number of cycles is 20** and we did not attempt other cycle numbers for any further optimizations.

4.2.2.2 DNA input amount for amplification – Optimizations

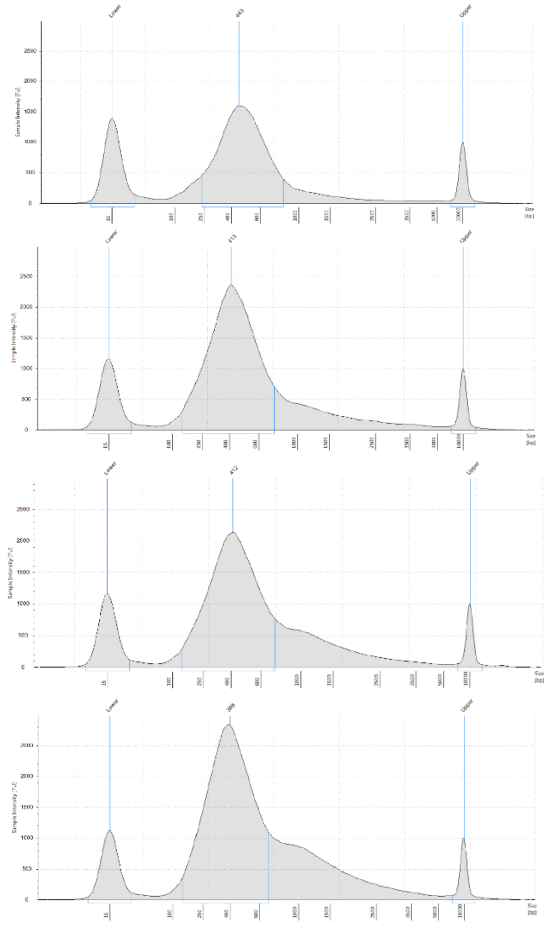
As mentioned earlier, DNA input amount is the second variable at the PCR amplification step to obtain duplex libraries with sufficiently high duplicative levels without compromising the efficiency of the sequencing run. Since previous runs with 10 amol of DNA as the PCR input proved to be non-optimal when sequenced, the next goal was to determine the lowest DNA amount needed as the input so that the highest number of copies of our molecular species of interest is present in the system and two sets of experiments were performed to this end with 20 PCR cycles.

A. Set I: 4, 8, 12, 16 amol

The first set of DNA input amounts were based on the lower end of the PCR input recommendations of (Kennedy et al. 2014) to re-test our optimal range: (i) 4 amol, (ii) 8 amol, (iii) 12 amol and, (iv) 16 amol **(Fig. 21)**.

From the figure, it is clear that increasing the DNA input amount while keeping the number of cycles constant increases the size of the peak of interest between 300-400 bp. However, input amounts greater than 8 amol also results in the formation of higher molecular weight non-specific PCR products for cycle number 20.

The conclusion from this set of experiments is that 4 amol is the best DNA input amount within the current experimental design.



Case 1: 4 amol
DNA amount obtained at end = 35.75 ng

Case 2: 8 amol
DNA amount obtained at end = 42.25 ng

Case 3: 12 amol
DNA amount obtained at end = 47.5 ng

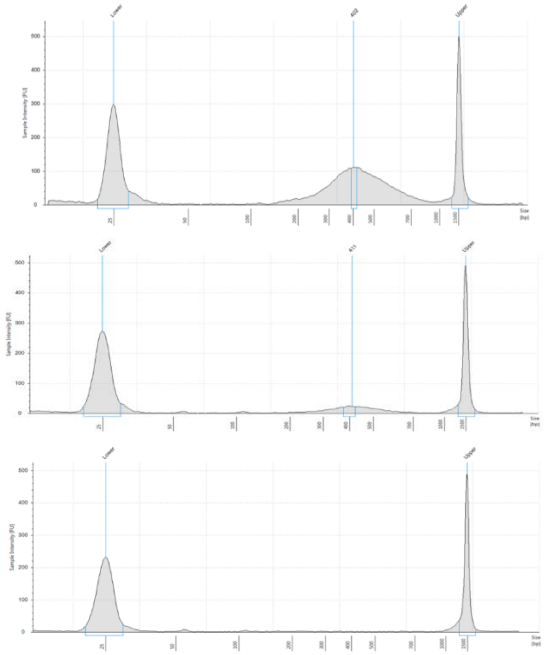
Case 4: 16 amol
DNA amount obtained at end = 55.50 ng

Figure 21: DNA input amounts for PCR – Optimizations (set 1: 4, 8, 12, 16 amol).
Four conditions tested to optimize the DNA input during library amplifications Set 1: (1) 4 amol (2) 8 amol (3) 12 amol and (4) 16 amol. Higher molecular weight non-specific fragment formation seen at DNA input greater than 8 amol, suggesting 4 amol is optimal. DNA amounts calculated from Qubit HS DNA Assay readings.
amol: attomoles; DNA: Deoxyribonucleic acid.

B. Set 2: 10, 1.67, 0.28 amol

We also designed an experiment to see if it were possible to further reduce the DNA input in the hopes that more DNA copies would be made of the target library upon amplification. Since the lowest amount of input DNA determined in previous optimizations in the lab was 10 amol, lower amounts were tested by reducing the input by a factor of 6 to study peak sizes at input amounts below 2 amol: (i) 10 amol, (ii) 1.67 amol, and (iii) 0.28 amol **(Fig. 22)**.

From the figure, it is clear that decreasing the DNA input amount to below 2 amol while keeping the number of cycles constant at 20 does not improve the size of the peak of interest between 300-400 bp, and is even too low to be detected by the TapeStation Bioanalyzer and Qubit reader machines.



Case 1: 10.00 amol
DNA amount obtained at end = 39.9 ng

Case 2: 1.67 amol
DNA amount obtained at end = 15.3 ng

Case 3: 0.28 amol
Amount too low to be detected by the machine

Figure 22: DNA input amounts for PCR - Optimizations (set 2: 10, 1.67, 0.28 amol).
Three conditions tested to optimize the DNA input (below 2 amol) during library amplifications Set 2: (1) 10.00 amol (2) 1.67 amol and (3) 0.28 amol. No peaks seen below 2 amol DNA input, further suggesting 4 amol is optimal. DNA amounts calculated from Qubit HS DNA Assay readings.
amol: attomoles; DNA: Deoxyribonucleic acid.

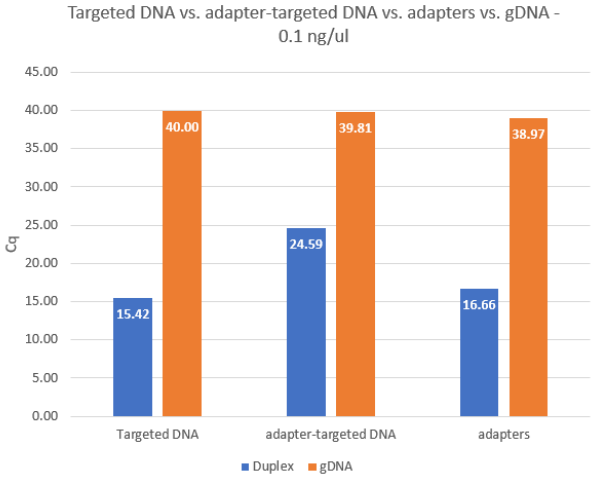
Thus, 4 amol is the optimum DNA input amount capable of generating 8 million reads per sample (Table 2, Kennedy et al. 2014), and has been used for all future amplicon-based duplex library preparations.

4.2.3 qPCR verification test

With all the size selection and library amplification optimizations done, an aliquot of the purified amplified products of the 4 amol, 20 cycles sample (**Fig. 20, Case 1**) was taken, diluted to 0.1 ng/µl and prepared for qPCR verification as described in Chapter 3, section 3.2.3.7, but for the gene used for amplicon-based duplex sequencing optimization.

qPCR readings for each molecular species present in the system was obtained in three replicates and the average of the C_q values from the three replicates was used to make the ddCq ($\log_2$ expression fold change) calculations.

A



B

Assay	Molecular probe	Duplex	gDNA
1	Targeted DNA	15.42	40.00
2	adapter-targeted DNA	24.59	39.81
3	adapters	16.66	38.97

(Duplex - gDNA)	dCq	ddCq
Assay 1	-24.58	25079408.73
Assay 2	-15.22	38212.97
Assay 3	-22.31	5211883.57

***$ddC_q = 2 \wedge (-dC_q) = \log_2$ expression fold change**

<u>Principle</u>: Lower the C_q value, more are the number of initial molecules

No. of 'targeted DNA' molecules **>** *No. of 'adapter' molecules* **>>** *No. of 'adapter-ligated targeted DNA' molecules >>> Native genomic DNA molecules*

Figure 23: qPCR verification for the optimization of amplicon-based duplex libraries.

Assay 1 corresponds to 'targeted DNA' molecules, Assay 2 helps identify 'adapter-ligated targeted DNA' molecules, and Assay 3 corresponds to 'adapter' molecules as seen in the duplex library *vs* native genomic DNA.

(a) *<u>Graphical representation of the varying C_q values between each assay for 4 amol, 20 cycles</u>*: 'Targeted DNA' has the lowest C_q (15.42 ± 0.42) of the duplex library sample which means this is the predominant molecular species in the system while the 'adapter-ligated targeted DNA' has the highest C_q (24.59 ± 0.22) of the duplex library sample which means this has the lowest presence in the system, with the C_q value of 'adapters' detected (16.66 ± 0.35) being closer to the 'targeted DNA' C_q. In contrast, the qPCR detected the lowest levels of native genomic DNA sample (experimental control) with the highest C_q values (39.60 ± 0.63).

(B) *<u>dC_q and ddC_q values</u>*: give a relative estimate of the number of each molecular species relative to one another. The ratio of the number of 'adapter-ligated targeted DNA' molecules detected to experimental control is ~ (1:38,000) which is suggestive of successful duplex library preparation.

amol: attomoles; C_q: Quantitative threshold cycle number, dC_q: Difference between two threshold cycle numbers, ddC_q: base-2 logarithmic expression fold change; DNA: Deoxyribonucleic acid; gDNA: Genomic DNA.

In **fig. 23**, the total numbers of gene DNA or adapter molecules represent probes binding to molecular species present in free as well as ligated forms, and it is clear that a large percentage of these are present in "adapter-ligated targeted DNA" PCR products and the number of this molecular species is also very high when compared to the experimental control (adapter-ligated targeted DNA control: experiment = 1: 38213), both of which are suggestive of successful duplex library construction.

With the amplicons successfully prepared using GXL and SuperFi DNA polymerases, library construction process optimized, and a verified qPCR test design, we then prepared amplicon-based duplex libraries for the MAPT gene with both the polymerases for the FTD-Tau samples and analysed the sequencing data for a comparison of duplex libraries constructed with both the DNA polymerases.

4.2.4 Computational analyses of sequencing data

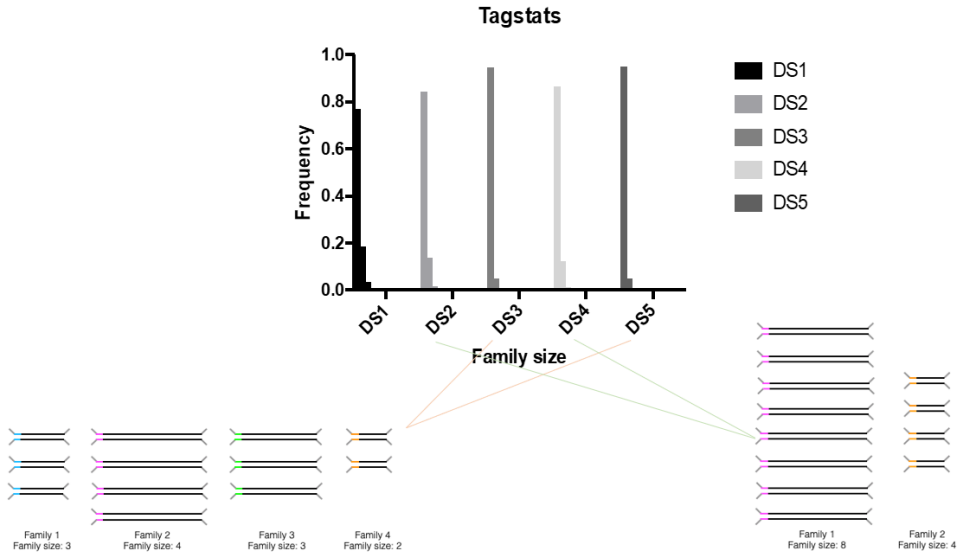
The computational analyses to compare duplex libraries generated using both GXL-mediated amplicon preparations as well as SuperFi-mediated amplicon preparations was limited to the temporal lobe *vs* medulla regions of the disease *vs* age-matched control FTD samples.

All the bioinformatic and statistical analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

1) Family Size

The first analysis done for duplex sequencing is to check if there are enough number of reads to have an optimum family size to be able to generate double consensus. Again, **family size** refers to the number of 'sequencing reads' that share the same tag sequence that form one family that can eliminate false-positive calls using appropriate bioinformatic pipelines (reads with duplex consensus).

Previous attempts in the lab to modulate library amplification parameters resulted in poor family sizes for amplicon-based DS (**Fig. 24**) (Nie Y, unpublished data).



Higher DNA input, lower cycle number for lib. amp

Lower DNA input, higher cycle number for lib. amp

Figure 24: Amplicon-based DS - Family size influenced by library amplification parameters.
The family sizes (reads with duplex consensus) of a total five samples after amplicon-based DS were calculated. The samples showed mostly poor peak family sizes (~ 1) with two distinct trends: (1) A higher DNA input with lower cycle numbers than optimal resulted in fewer family members to call single and duplex consensus confidently; and, (2) A lower DNA input with higher cycle numbers than optimal resulted in some families having too many or too few members to call single and duplex consensus confidently, indicating inefficient DS. (*Nie Y, unpublished data*).
DNA: Deoxyribonucleic acid; DS: Duplex sequencing; Lib. amp.: library amplification.

Our present attempt at amplicon-based DS mediated by two different polymerases yielded good average peak family sizes for all the samples, as seen in **Fig. 25** [x-axis: family size, y-axis: proportion of total reads] and highlights the importance of optimal parameters for library amplification for good family size and successful library construction.

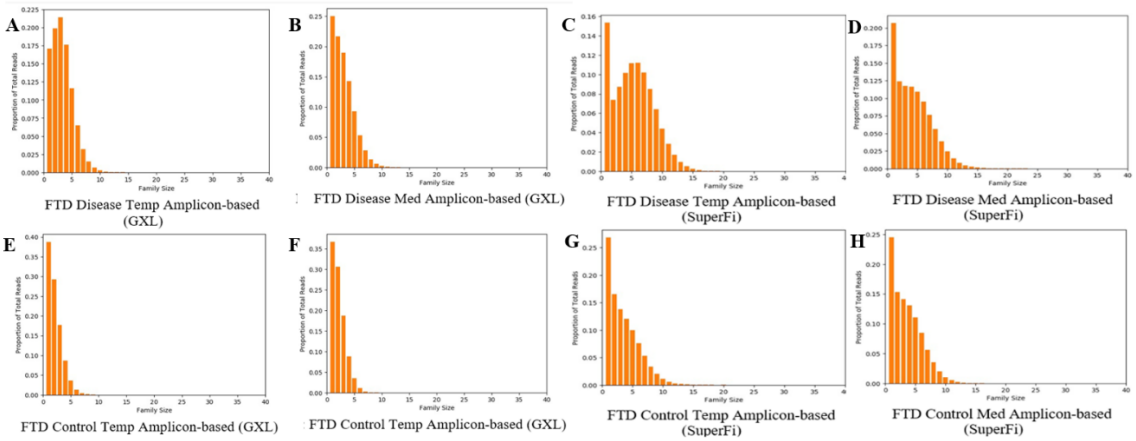


Figure 25: Amplicon-based duplex libraries – Family size calculations.
An average peak family size (reads with duplex consensus) between 3-4 was obtained for the GXL-mediated samples, and an average peak size of ~6-8 was obtained for the SuperFi-mediated samples, indicating that the Super-Fi based duplex libraries are of superior quality (Optimal peak family size ~ 6).
C576: Age-matched control; Med: Medulla; NP16.163: FTD disease samples; Temp: Temporal lobe.

2) Mutation Detection

True mutations detected after duplex consensus were broadly divided into two classes: those with a variable allele frequency (VAF) > 0.25, termed as germline mutations, and those with a VAF < 0.05, termed as somatic mutations.

Fig. 26 [x-axis: brain region sample, y-axis: number of paired reads/mutations] describes the mutational burden seen in GXL-based duplex libraries *versus* SuperFi-based duplex libraries.

Quantitative analysis of mutation detection showed that the number of somatic mutations (VAF < 0.25) were higher than germline mutations (VAF > 0.25) in all the cases in both brain regions of FTD disease and control in duplex libraries derived from both DNA polymerases. However, the number of germline mutations in the temporal lobe and medulla of NP16.163 disease states in both polymerases sequencing data (VAF > 0.25) is also higher than in their age-matched controls, indicative of the association of the accumulation of both, acquired and inherited, variations in FTD patients.

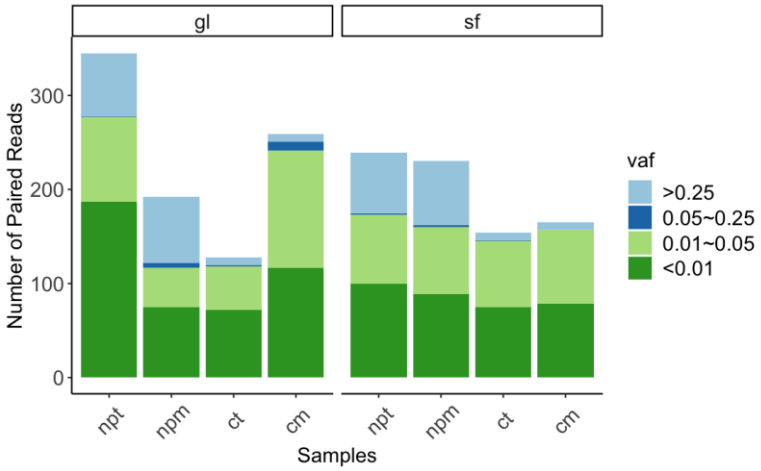


Figure 26: Amplicon-based DS: Mutation frequencies detected.

FTD control *vs* disease and Med *vs* Temp in GXL-based *vs* SuperFi-based duplex libraries. SuperFi-based data is indicative of higher accuracy as more false-positives are eliminated. Overall mutational burden is greater in disease than in control. The presence of somatic SNVs with very high VAF ~ 0.05-0.25 (dark blue), VAF ~ 0.01-0.05 (pale green), and ultra-low VAF < 0.01 (dark green) indicate that these are likely to contribute to ageing and disease state in FTD patients. However, germline mutational burden with VAF > 0.25 (light blue) is also greater in disease than controls as seen in libraries made from both polymerases, indicating that these are also likely to be associated with FTD pathology.

FTD: Frontotemporal Dementia, Med: Medulla, Occ: Occipital lobe, Temp: Temporal lobe, SNVs: Single nucleotide variations; VAF: Variable allele frequency.

It is also quite clear from **Fig. 26** that the Platinum SuperFi II DNA polymerase was the superior polymerase to use for the initial pre-library amplification for duplex library preparation due to the reduced background noise indicating a more successful elimination of false-positives. However, it is possible that some of the somatic mutations detected in SuperFi-mediated libraries may still be attributed to the incorporation of untraceable PCR errors in the pre-library amplification step in addition to true mutations, making it imperative that we also explored alternative approaches to DS.

4.3 Discussion

This chapter explores the variations in PCR protocols for pre-library amplification of the three MAPT-amplicons via GXL and SuperFi DNA polymerases as well as the various optimizations done for amplicon-based duplex library amplification and concludes with computational data analyses to compare the quality of the duplex libraries generated using both DNA polymerases and the mutation detection power of each.

We initially focused on the three amplicons designed to capture the regions of highest genetic variation within the MAPT gene and the regions with the greatest number of established pathogenic mutations, labelled here as MAPT-A, B, and C that covered the regions of exons 8, 9-10, and 11-12. We optimized for amplifying our template DNA using two polymerases – the

standard Takara GXL PrimeSTAR DNA polymerase and the new Platinum SuperFi II DNA polymerase with higher fidelity. Since the two polymerases have different properties, we needed different PCR programmes to successfully amplify our amplicons, all of which have been described in detail in Chapter 3, section 3.2.4.

We then optimized the crucial parameters for successful duplex library preparation: (i) DNA input for library amplification and (ii) the amplification cycle number. The two parameters of library amplification needed to be modulated so that we could have the optimal numbers of duplex libraries with good duplicative levels that would not compromise the efficiency of the sequencing run which meant the lowest DNA input and the highest cycle number possible. A good result of such a scenario on the TapeStation Bioanalyzer would show the final duplex library with a single, sharp, and large peak within 300-400 bp range along with the absence of any other non-specific higher molecular size fragments. These results were also supported by our qPCR verification design. With the optimized protocols, we prepared duplex libraries generated using both the DNA polymerases for sequencing and analysed the sequencing data.

For the amplicon-based method, we found an input of **4 amol amplified with 20 cycles** to be optimal, as seen with the average peak family size being close to the recommended range in all the samples analysed. all the samples also showed a higher somatic mutational burden (~ 150-200 somatic SNVs detected in all the brain regions considered for both, the disease case and the control) than germline mutational burden (~50 germline SNVs detected in both brain regions of disease *versus* ~10 germline SNVs detected in both brain regions of control). However, the quantitative mutation frequency analyses also showed that the germline mutational burden was higher in both brain regions tested of disease than those of control, indicative of the contribution of both types of mutations to ageing and disease in the patient.

Thus, the amplicon-based approach was useful in quickly and more easily detecting mutations in FTD-Tau samples. The higher fidelity of Platinum SuperFi II proved to play an important role in detecting the mutation rate in brain regions more accurately with a lower number of mutations detected indicating that this method was likely to be more successful in eliminating false-positives successfully.

Although seemingly straight-forward, it is evident that the amplicon-based method of duplex library preparation required lots of optimizations at various steps of the protocol. It

emphasises the need to be vigilant and thorough while performing the protocol, particularly at the sensitive stages of adapter-ligation and library amplification to obtain the best results. However, the pre-library amplification before performing the duplex library preparation could introduce untraceable PCR errors into the system which could be difficult to eliminate with the bioinformatic pipeline designed for DS, which is why alternative approaches to DS were also considered.

Chapter V: Optimization of Double capture-based Duplex Sequencing – Protocol 1

(With in-house biotinylated baits)

5.1 Introduction

The principle of the double capture-based approach to duplex sequencing is to first prepare duplex libraries and then perform "targeted-capture" of the libraries of interest with custom-designed "biotinylated baits/probes," twice. The lack of amplifications before using the specific duplex adapters would reduce the incorporation of untraceable PCR errors that would make 'true mutations' easier to identify by making it easier to eliminate false-positives during computational analyses, thus, overcoming the limitations of the amplicon-based approach.

To this end, the protocol to be adapted for our first attempt at a double capture-based duplex library preparation included a double-capture protocol to capture the DNA of interest from a homogenized DNA mixture wherein duplex sequencing is performed first, followed by the enrichment and targeted capture of the DNA of interest from complex DNA mixtures. Previous work in the lab had shown that this method was successful in capturing mitochondrial DNA as described in the paper (Maricic, Whitten, and Pääbo 2010) and so we attempted to adapt this method to capture genomic DNA containing MAPT sequence (exons 8-12) as well by using home-prepared biotinylated baits designed to capture the three previously-described amplicons of the MAPT gene.

This chapter presents the results from this first double-capture protocol (with in-house biotinylated baits) modified for targeted-capture of MAPT DNA after preparing duplex libraries on an FTD sample as a test. An overview of this protocol 1 has been summarized in a schematic in **Fig. 27**.

The chapter then ends with a discussion on the quality of sequencing data procured after using this method, followed by a summary of the limitations of this protocol for the targeted-capture of nuclear DNA.

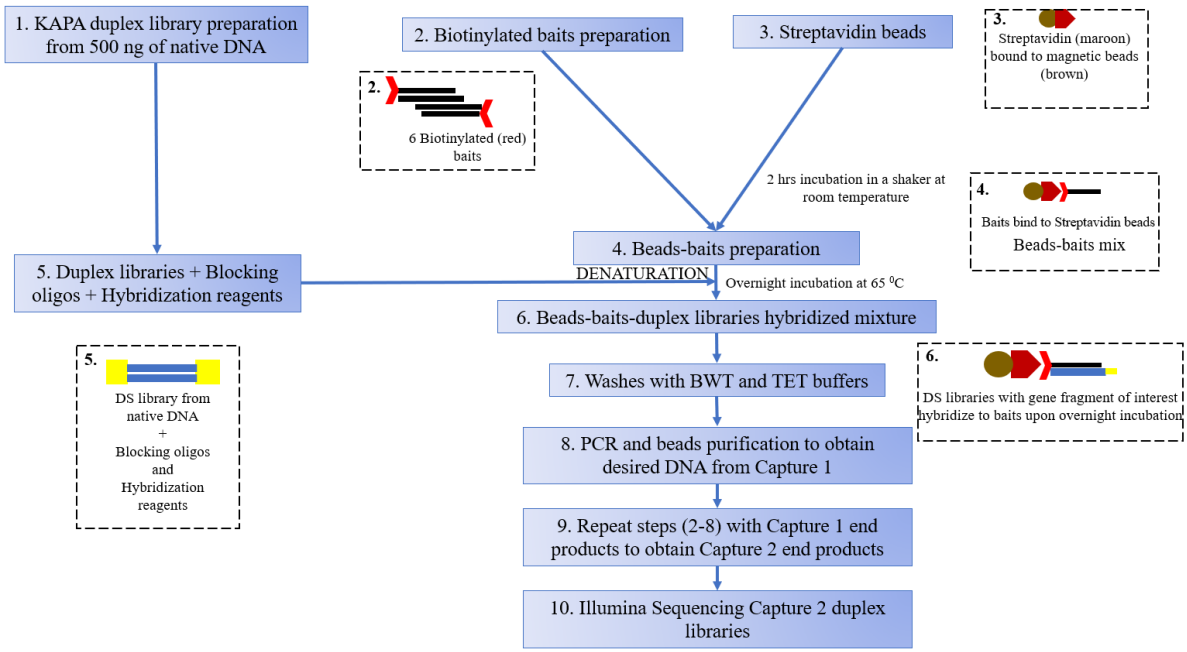


Figure 27: Double capture-based duplex library preparation - Protocol I (with in-house biotinylated baits).

DAY 1: Duplex libraries prepared with native DNA with KAPA HyperPlus kit.
DAY 2: Preparation of biotinylated baits for the three MAPT amplicons.
DAY 3: Preparation of streptavidin beads for capture, denaturation of biotinylated baits to single-strands, and Streptavidin beads-biotinylated baits mixing on a shaker for 1-2 hours. Then, duplex libraries added to the beads-baits mix followed by an overnight incubation at 65 ^{0}C.
DAY 4: Post-capture I washes and library amplification. Steps to set up a new beads-baits mix repeated and this time, duplex libraries from post-capture I amplification added for the overnight incubation at 65 ^{0}C.
DAY 5: Post-capture 2 washes and library amplification. Double capture-based duplex libraries now ready for sequencing on a suitable Illumina platform.
DNA: Deoxyribonucleic acid; MAPT: Microtubule-associated protein tau.

5.2 Results

5.2.1 Duplex Library preparation

Duplex libraries were prepared for all six samples using the method as described in Chapter 3, section 3.2.4. The HS D1000 ScreenTape results to verify adapter-ligation and HS D5000 ScreenTape results to check library amplification were successful (see *Appendix I, column 1* for results).

5.2.2 Biotinylated baits preparation

Biotinylated baits were prepared for all three MAPT amplicons such that one of the PCR primers (forward or reverse) was always biotinylated, i.e., six biotinylated baits were prepared with each amplicon having a "Forward-biotin + Reverse" and a "Forward + Reverse-biotin" primer sets. These were amplified and run on a 0.7x agarose gel to verify the bait sizes corresponded to the MAPT amplicon intended to be captured (**Fig. 28**).

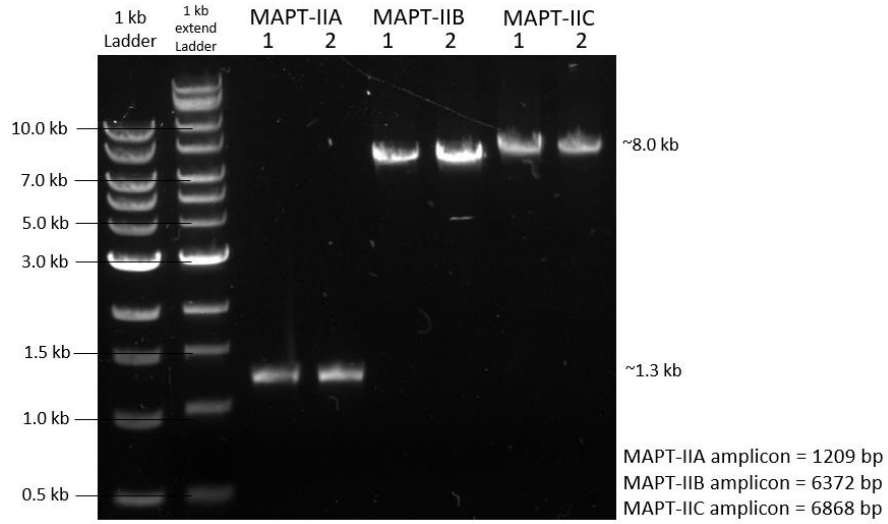


Figure 28: In-house biotinylated baits preparation for double capture (Protocol 1).
Gel migration for amplicons MAPT-A (first set of two lanes), MAPT-B (second set of two lanes), and MAPT-C (third set of two lanes). A single band was seen within the expected range of each amplicon (MAPT-A = 1209 bp, MAPT-B = 6372 bp, MAPT-C = 6868 bp). Lane 1 for each amplicon is the (Forward-biotin + Reverse) amplicon and Lane 2 for each amplicon is the (Forward + Reverse-biotin) amplicon.
MAPT: Microtubule-associated protein tau.

The gel results show that the biotinylated baits preparation for both PCR primer sets for all three amplicons were successful with each having the band size as expected. These baits were then mixed in equimolar proportions (see Chapter 3, section 3.2.5.4a) and made ready for capturing duplex libraries prepared from the end of section 5.2.1 above.

5.2.3 Capture protocol optimizations

The first and second rounds of capture were mainly optimized at the library amplification stage after the overnight incubation step and several attempts were made to this end:

1. Attempt 1

For our first attempt at the double capture protocol, we adapted the protocol as described in the paper (Maricic, Whitten, and Pääbo 2010) for our MAPT baits on the sample NP16.163 disease Med as a test (see chapter 3, section 3.2.5). Library amplification was done for both the rounds of capture as: Capture 1 – All the DNA from 'dried' beads, 16 cycles; Capture 2 - All the DNA from 'dried' beads, 20 cycles. The final library after two rounds of capture was verified on HS D5000 ScreenTape. Unfortunately, the ScreenTape results were not as desired with only a small percentage of the final libraries within our expected size range (**Fig. 29a**).

We then repeated the first round of capture alone with a new aliquot of duplex libraries and baits to check if the libraries were as expected at least at this stage of post-capture 1

amplification. However, there was, again, a greater presence of non-specific higher molecular weight fragments (**Fig. 29b**), suggesting that the DNA input and cycle numbers needed to be optimized for both rounds of capture.

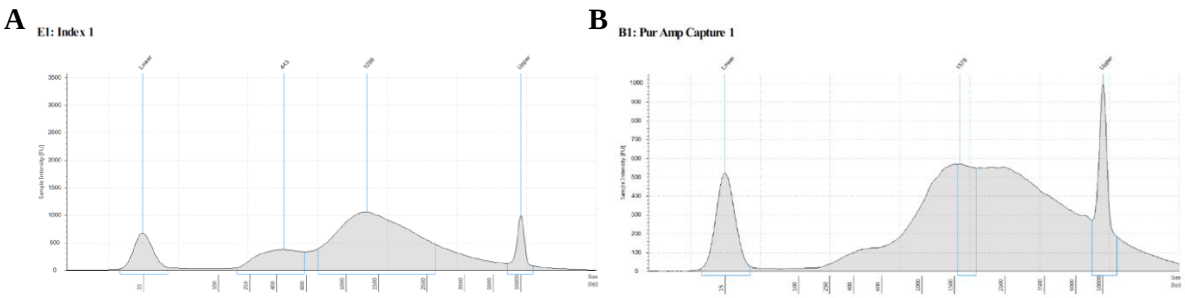


Figure 29: Attempt 1: Library amplification optimizations post-both captures.
The figure is the HS D5000 ScreenTape results for NP16.163 Med after:
(a) *Post-capture 2* – Non-specific high molecular weight fragment formation seen, making this unsuitable for sequencing.
To understand if the non-specific high molecular weight fragments were formed only after the second capture, or if they were formed after the first capture itself, a single capture was done on the native FTD-tau DNA sample and checked on the ScreenTape:
(b) *Post-capture 1* – Non-specific high molecular weight fragments seen here as well, indicating the need to optimize the library amplifications after both captures.
DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; Med: Medulla; NP16.163: FTD-Tau disease case.

2. Attempt 2

a. Capture 1 optimizations

To optimize the library amplification after first round of capture, we performed the capture protocol as is until the BWT and TET washes (Chapter 3, sections 3.2.5.1 to 3.2.5.7) on control medulla (C576 Med). Then, instead of taking the 'dried' beads for PCR, we instead eluted the beads into 5 µl of NaOH and incubated for 5 min. Then, we collected beads on a magnetic stand and discarded the beads after transferring the eluate to a fresh tube. Of the 5 µl of eluate, 1 µl was added each into three different PCR mixtures (**Table 14**) and we performed three amplifications at different cycle numbers: (a) 8 cycles (b) 16 cycles and (c) 30 cycles which were then verified on HS D5000 ScreenTape.

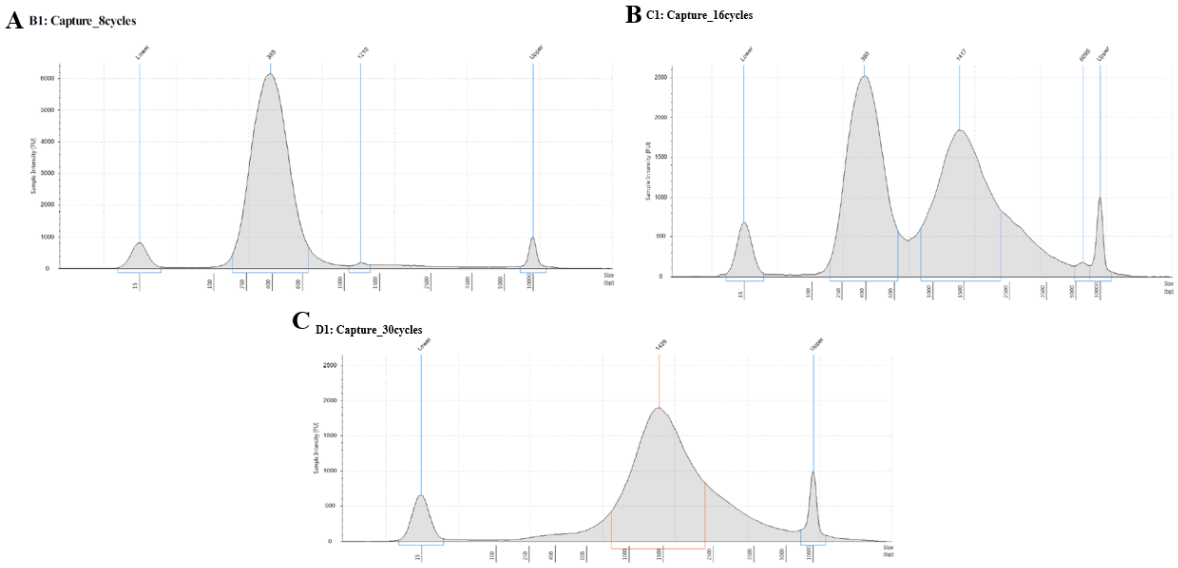


Figure 30: Attempt 2: Library amplification optimizations post-capture 1.
(a) *Case 1*: 1 µl, 8 cycles – only one single, sharp peak within 300-400 bp range as desired and is suitable for the second capture. (b) *Case 2*: 1 µl, 16 cycles – Two peaks seen wherein one of the peaks is within desired range, but the other peak shows the presence of non-specific high molecular weight fragments which is unsuitable for the second capture. (c) *Case 3*: 1 µl, 30 cycles – One large peak seen indicating the presence of non-specific high molecular weight fragments which is highly unsuitable for the second capture.

Out of the three cycle numbers, it is evident that '(a) 8 cycles' is the optimum as it alone gives a single sharp peak within 300-400 bp range (**Fig. 30a**). In contrast, '(b) 16 cycles' gives the multiple peaks at the expected as well as undesired ranges indicating that this is a higher cycle number than necessary (**Fig. 30b**) while '(c) 30 cycles' gives a large peak with fragment sizes larger than 1000 bp (**Fig. 30c**), thus showing that cycle numbers greater than 15 overwhelms the system.

We used the '1 µl, 8 cycles' sample for library amplification optimizations post-capture 2 done on NP16.163 Occ.

b. Capture 2 optimizations

Libraries obtained from Capture 1 '1 µl, 8 cycles' were taken for a second round of capture as described in Chapter 3, section 3.2.5.8. Again, once the captured DNA is eluted into 125 mM NaOH, 1 µl was taken as the input each into three different PCR mixtures (**Table 14**) and three amplifications were done at different cycle numbers: (a) 8 cycles (b) 10 cycles and (c) 12 cycles which were then purified with (-0.64x; 1.0x) double-tail beads selection (Chapter 3, section 3.2.2.3) and verified on HS D5000 ScreenTape. Double-tail beads selection was done to prevent any non-specific higher molecular weight fragments from being selected.

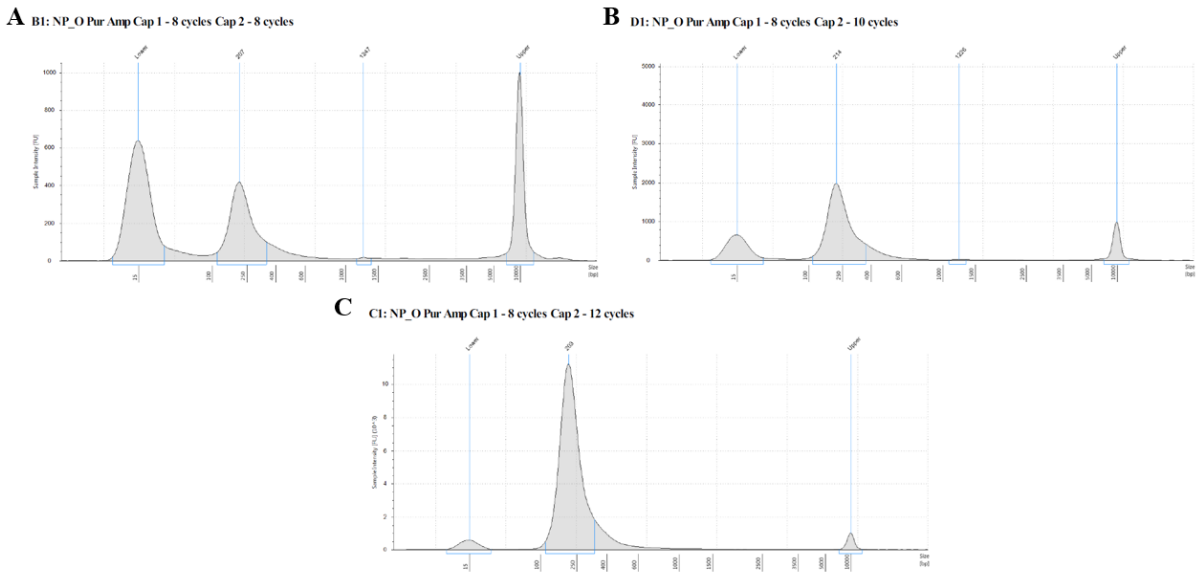


Figure 31: Attempt 2: Library amplification optimizations post-capture 2.
All three samples derived from Cap 1: 1 µl, 8 cycles duplex libraries. Double-tail beads selection (-0.64x; 1.0x) done for all three samples: (a) *Case 1*: 1 µl, 8 cycles, (b) *Case 2*: 1 µl, 10 cycles, (c) *Case 3*: 1 µl, 12 cycles – only one single, sharp peak within 300-400 bp range as desired for all three cases; however, all three cases show the presence of non-specific high molecular weight fragments despite double-tail beads selection, indicating that these DNA input and cycle number still overwhelm the system.
Cap 1: Capture 1; DNA: Deoxyribonucleic acid.

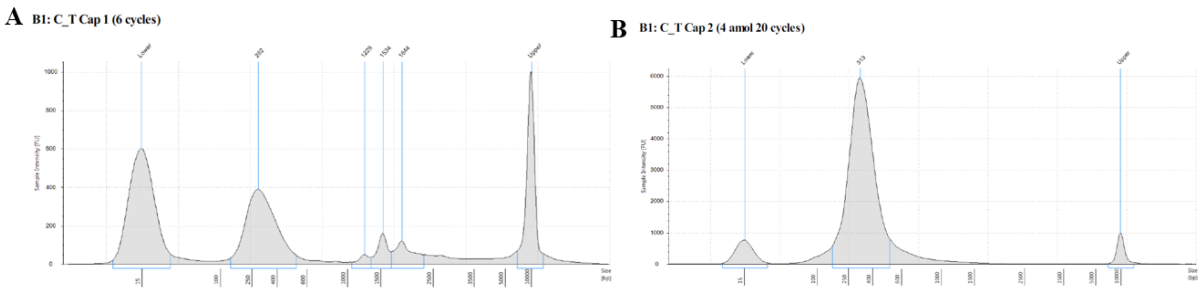
From **Fig. 31**, while it is clear that increasing the cycle number while maintaining a DNA input of 1 µl for all three cases also increases the amount of DNA present in the final amplified libraries (greater area covered in grey from case A to C), the presence of higher non-specific fragments remains a concern despite using a double-tail bead selection for all three cases (area under each peak represents number of molecules of a species detected by the TapeStation programme; non-specific higher molecular weight fragments undetected by machine due to relatively higher proportion of molecules under peak of interest compared to lower and upper markers in **Fig. 31c**). A double-tail selection such as this also reduces the number of library molecules present in the amplified-only DNA, which suggests that these cycle numbers may be too large for the present system; however, reducing the cycle numbers any further could limit the family size in the final data analyses.

To overcome these short-comings in optimizations, a final attempt was made to modify the Maricic, Whitten, and Pääbo method for MAPT targeted-capture.

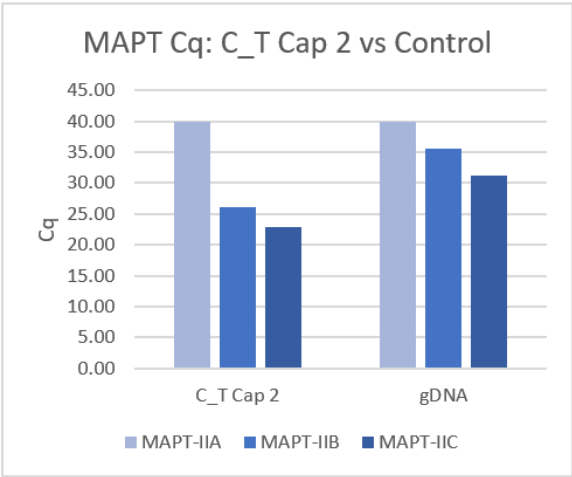
3. Attempt 3

Since, '8 cycles' was the optimum for 1 µl of input DNA for library amplification post-capture 1, we then attempted a final double capture optimization on C576 Temp by amplifying the libraries post-capture 1 with all the DNA on the 'dried' beads for 6 cycles, which gave a

significant large peak as desired at ~300 bp, along with negligible higher molecular weight fragments (**Fig. 32a**), and for post-capture 2 library amplification, 4 amol of DNA was taken as the input from the eluted DNA for 20 cycles, which gave a single sharp peak as desired at ~300 bp without the presence of any higher fragment peak sizes (**Fig. 32b**) (See *Appendix I* for HS D5000 (column 2) ScreenTape results of pre-capture amplified and purified duplex library preparation from native DNA).



C



(C_T) - gDNA	dCq	ddCq
MAPT-IIA	0.00	1.00
MAPT-IIB	-9.44	694.58
MAPT-IIC	-8.19	292.04

Figure 32: Attempt 3: Library amplification optimizations post-both captures.
(a) *Cap 1*: All DNA from 'dried' beads, 8 cycles (HS D5000 ScreenTape) – One main peak within 300-400 bp range as desired along with negligible peaks indicating favourable post-capture 1 results.
(b) *Cap 2*: 4 amol, 8 cycles (single-tail beads selection) (HS D5000 ScreenTape) – A single, sharp peak within 300-400 bp range as desired and is suggestive of the suitability of these duplex libraries for sequencing.
(c) *qPCR verification*: The duplex libraries made for MAPT amplicons in an FTD sample, C576 Temp, was compared with native DNA that underwent no enrichment as the control for the qPCR test. The C_q and ddC_q values of the presence of duplex libraries corresponding to the three MAPT amplicons in C576 Temp sample is indicative of the enrichment of DNA molecules of these three regions compared to the non-enriched control DNA, further increasing the confidence for sequencing.
amol: attomoles; C576: Age-matched control; Cap 1: Capture 1; Cap 2: Capture 2; C_q: Quantitative threshold cycle number, dC_q: Difference between two threshold cycle numbers, ddC_q: base-2 logarithmic expression fold change, DNA: Deoxyribonucleic acid; gDNA: non-captured genomic DNA; MAPT: Microtubule-associated protein tau; qPCR: Quantitative polymerase chain reaction; Temp: Temporal lobe.

A qPCR check of the final amplified libraries using the qPCR primers as described in Chapter 3, section 3.2.3.7 also showed good amplification of desired regions in the duplex libraries as compared to native genomic DNA (**Fig. 32c**).

This final capture ScreenTape is exactly as desired and with acceptable results on qPCR, and these duplex libraries were submitted for Illumina NovaSeq to check the family size in order to apply the protocol to the remaining FTD disease and control samples, if successful.

5.2.4 Computational sequencing data analyses

All the bioinformatic analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

5.2.4.1 Family size

The first analysis done for duplex sequencing is to check if there are enough reads to have an optimum family size to be able to generate double consensus. For good analyses, an optimum peak family size (reads with duplex consensus) of 6 is required for the sample as calculated previously (Kennedy et al. 2014).

The sample, control temporal lobe (C576 Temp), was double-captured with Capture 1 – all DNA from 'dried' beads for 6 cycles, Capture 2 – 4 amol for 20 cycles. The family size obtained in this case was only 1, which indicates poor sequencing duplex library preparation (**Fig. 33**).

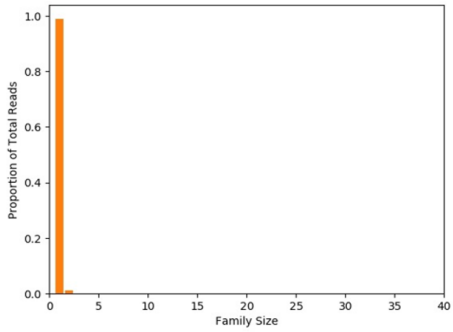


Figure 33: Double-capture Protocol 1 – Family size calculations.
The sequencing data gave a peak family size (reads with duplex consensus) of 1, which shows that library construction with the optimized double capture-based duplex library preparation protocol 1 for C576 Temp was unsuccessful (Optimal peak family size ~ 6).
C576: Age-matched control; Temp: Temporal lobe.

5.2.4.2 On-target ratio

To analyse the efficiency of capture, an analysis of the on-target ratio, i.e., the proportion of reads that cover chromosome 17 (our region of interest) with respect to the whole genome.

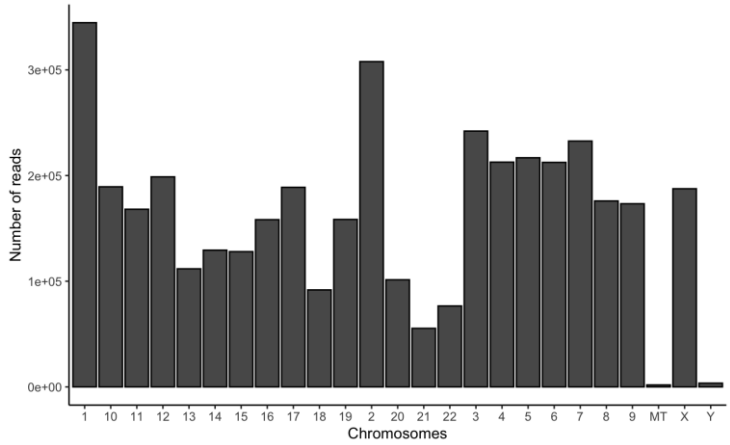


Figure 34: Double-capture Protocol 1 – On-target ratio calculations.
The proportion of reads covering chromosome 17 (containing MAPT gene) is comparable to the whole genome showing that a significant enrichment and targeted-capture of MAPT DNA did not occur with the optimized double capture-based duplex library preparation protocol 1, suggesting that the in-house biotinylated baits were not powerful enough for our purposes.
DNA: Deoxyribonucleic acid; MAPT: Microtubule-associated protein tau.

For the present sample obtained after double-capture of C576 Temp, substantial enrichment of the MAPT amplicons present on chromosome 17 is not seen (**Fig. 34**), suggesting that this protocol is not suitable for our purposes. This may be due to the significantly low number of biotinylated baits used (six in total covering a total of ~14 kb) to be identified within a genome that is ~6 x 10^9 bp (diploid), making it imperative to explore other protocols for double capture-based duplex library preparation.

5.3 Discussion

This chapter delves into the modifications done to the protocol used by (Maricic, Whitten, and Pääbo 2010) for double capture-based duplex library preparation.

Our optimization experiments showed that while the protocol may have been successful for previous experiments to capture mitochondrial DNA whose entire genomic sequence is ~16 kb with hundreds-thousands of copies per cell; in contrast, the biotinylated baits that need to be designed for the protocol do not have enough sensitivity to capture such a small target (~13-14 kb) of the nuclear genome (~3.2 × 10^9 bp) which would contain only two copies of the gene per cell which complicates capture efficiency substantially. Indeed, a lot of time and resources were spent in optimizing the parameters for library amplifications after both the captures and the final duplex libraries that did pass the wet-lab quality checks had an insufficient family size and poor on-target ratios upon analyses.

Thus, our results suggested that while the biotinylated baits prepared and used were theoretically-sound, they were not strong enough to capture duplex libraries containing our gene of interest from a complex pool of duplex libraries made from native genomic DNA. Thus, performing a duplex library preparation based on double-capture targeting a larger MAPT region with a much higher number of probes may yield better results for our purposes.

An interesting point of note is that our qPCR verification system is vulnerable in that while it can quantify the enrichment of adapter ligated-DNA libraries after the two captures; however, it does not give any three-dimensional structural information as to whether these duplex libraries have been constructed correctly such that they are capable of successfully producing good data upon sequencing, i.e., the correct duplex library will be a Y-shaped right-angled structure in 3-D with the sequence described as in **Fig. 8b** (Post-second round of amplification structure). Hence, we did not perform the qPCR check in our future experiments for DS.

Chapter VI: Optimization of Double capture-based Duplex Sequencing – Protocol 2

(With custom-designed IDT probes)

6.1 Introduction

Our previous attempt at the double capture-based duplex library preparation (protocol 1) yielded a poor quality of sequencing data due to the nature of the in-house biotinylated baits. We then decided to adapt and optimize another double capture-based duplex sequencing protocol as described by (Schmitt et al. 2015) wherein duplex sequencing was performed followed by the targeted capture of a nuclear gene. We decided to modify this protocol for targeted-capture of MAPT via double-capture in our duplex samples.

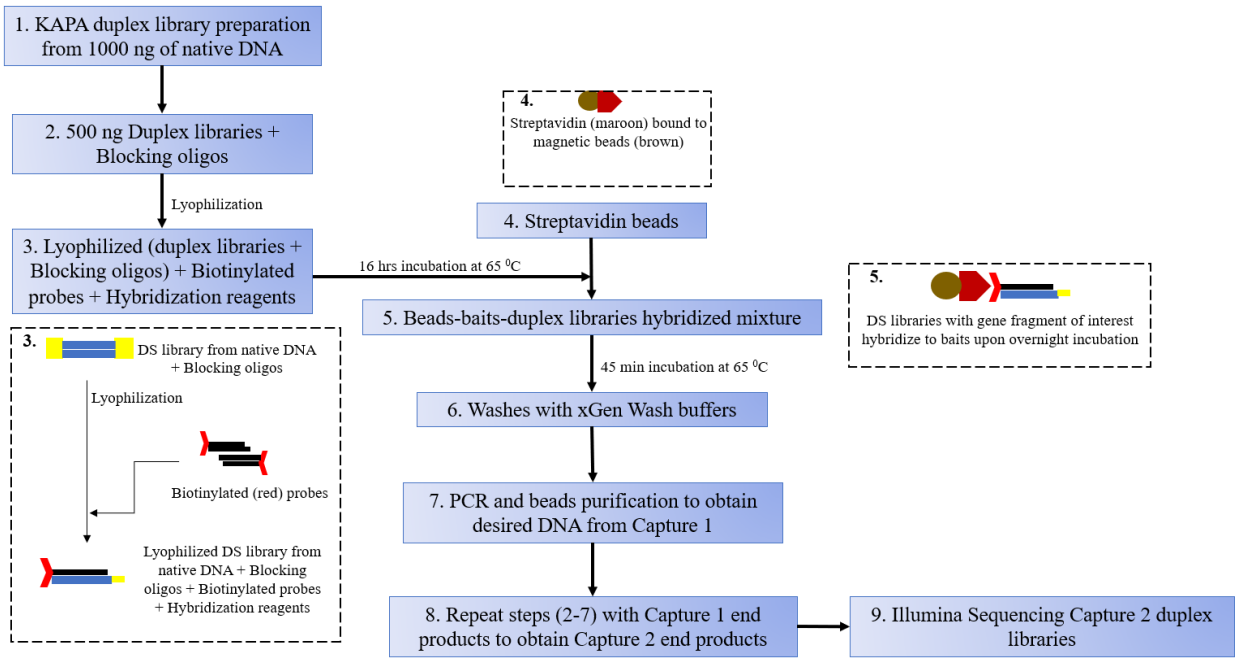


Figure 35: Double capture-based duplex library preparation - Protocol 2 (with custom-designed IDT probes).

DAY 1: Duplex libraries prepared with native DNA with KAPA HyperPlus kit.
DAY 2: Blocking oligos added to the duplex libraries before lyophilization Hybridization reagents and biotinylated probes added to lyophilized DNA followed by 16-hour incubation at 65 ^{0}C.
DAY 3: Streptavidin beads prepared and added to the incubated mix followed by 45-min incubation. Post-capture 1 washes and library amplification. Steps to set up a new biotinylated probes-hybridization reagent mix repeated and this time, duplex libraries from post-capture 1 amplification lyophilized and added for the 16-hour incubation at 65 ^{0}C.
DAY 4: Post-capture 2 washes and library amplification. Double capture-based duplex libraries now ready for sequencing on a suitable Illumina platform.
DNA: Deoxyribonucleic acid; oligos: oligonucleotides.

After consulting with Integrated DNA Technologies (IDT) about the custom-designed probes described in the aforementioned previous study (Schmitt et al. 2015), we then decided to use their updated 'xGen Hybridization Capture of DNA libraries v4' protocol that required little optimization as it was a commercial protocol. We used ~160+ custom-designed probes for the entire MAPT region containing exons 8-12. An overview of this protocol 2 has been outlined in a schematic in **Fig. 35**.

This chapter describes the optimizations done on genomic test DNA as well as two FTD samples, the quality of sequencing data procured after using this method and concludes with discussions on the number and nature of mutations detected within the two FTD samples under investigation and possible further improvements to this approach of duplex sequencing.

6.2 Results

The IDT Hybridization and capture protocol recommended duplex sequencing followed by a single round of targeted capture while the (Schmitt et al. 2015) protocol advocated duplex sequencing followed by two rounds of targeted capture. With this in mind, and our previous optimization with double capture-based DS protocol 1, three conditions were explored on 'test DNA' with varying rounds of capture and DNA input levels, with an additional condition for the FTD samples NP16.163 and C576 Temp samples (replicates). Here, 'Test DNA' refers to genomic DNA isolated from cell lines routinely cultured in the lab.

Thus, the IDT protocol was applied for five samples in total to test four conditions:

a) Test DNA single capture (as per IDT protocol),
b) Test DNA double capture (as per Schmitt *et al* 2015),
c) Test DNA double capture (Capture 1 – all DNA 16 cycles as in IDT protocol, Capture 2 – 4 amol for 18 cycles as per double-capture Protocol 1 optimization), and,
d) Double capture (as per Schmitt *et al* 2015):
    a. FTD disease temporal lobe (NP16.163 Temp) (protocol replicate 1),
    b. Control temporal lobe (C576 Temp) (protocol replicate 2).

A GXL-mediated amplicon-based duplex library was also prepared from Test DNA as an experimental control to further compare the efficiencies of both the approaches to duplex

sequencing. The HS D1000 and HS D5000 ScreenTape results for the verification of adapter-ligation and library amplification (4 amol for 20 cycles) was successful (see *Appendix II*).

6.2.1 Library Preparation optimizations for Protocol 2

6.2.1.1 Duplex library preparation

As mentioned earlier, 'Test DNA' refers to genomic DNA isolated from cybrid cell lines with wild-type nuclear genome routinely cultured in the lab for other research projects. About 2 µg of test DNA was used as the starting material for performing duplex library preparation as described in Chapter 3, section 3.2.6.1.

For the FTD disease and control samples, 500 ng of native DNA was used as described in the same section (See *Appendix II* for HS D1000 and HS D5000 ScreenTape results).

6.2.1.2 Probes

For the initial test run and optimization experiments of the IDT protocol, probes were custom-designed for only the positive '+' (template) strand of genomic DNA as recommended by the protocol. For the chromosomal coordinates for MAPT gene as described in Chapter 3, section 3.2.6.2, ~160 probes of the total 240 designed were classified as "good" probes and were synthesized in the IDT Discovery Pools (see *Appendix III* for full list; 'good' probes will capture DNA sequences that will yield good reads that can be properly mapped to the reference genome; 'risk' and 'remove' probes will capture repetitive sequences whose resultant reads will be hard to map correctly to their respective genomic location on the reference genome).

6.2.1.3 Capture 1

The first round of capture including lyophilization, 16-hour incubation, mixing with streptavidin beads, 'heated' and 'room temperature' washes, and post-capture 1 amplification was done as described in Chapter 3, sections 3.2.6.3 to 3.2.6.10 (only '+' strand probes used here).

A total of five samples underwent the first capture. An index primer was used for the Test DNA single capture amplification (**Fig. 36b**) while the P5 and test primers were used for the Test DNA double-capture amplifications (**Figs. 36c, d, e, f**) (**Table 29**). An amplicon-based duplex library with index primer was also prepared with Test DNA to compare with the double-capture protocol 2 (**Fig. 36a**).

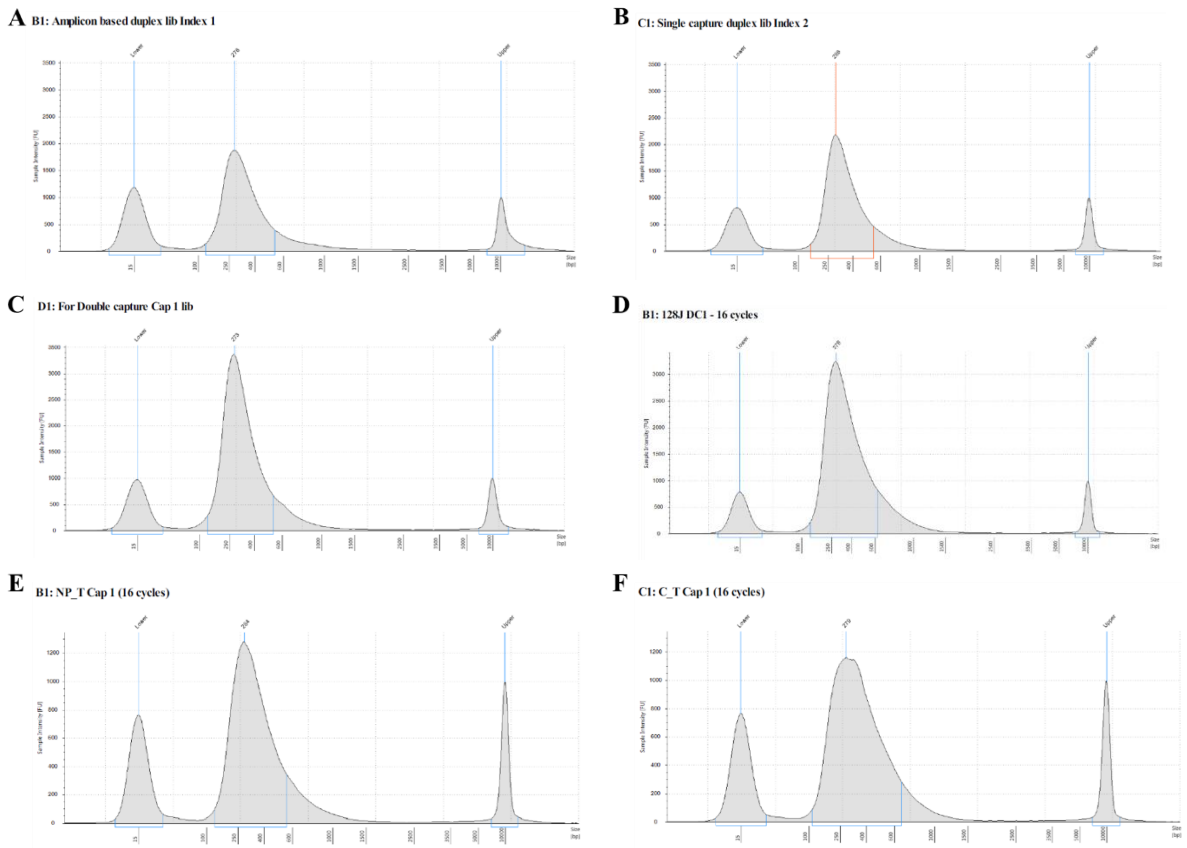


Figure 36: Amplicon-based *vs* Double capture-based duplex library preparation Protocol 2 - Post-capture 1 library amplifications

A single, sharp peak within 300-400 bp as desired on HS D5000 ScreenTape for:
(a) Amplicon-based (GXL) Test DNA (4 amol 20 cycles)
(b) Single capture Test DNA: all DNA 16 cycles (as per IDT protocol)
(c) Test DNA Cap 1: all DNA 16 cycles (as per Schmitt *et al* 2015)
(d) Test DNA Cap 1: all DNA 16 cycles (for testing optimized double-capture protocol 1 condition)
(e) NP16.163 Temp Cap 1: all DNA 16 cycles (as per Schmitt *et al* 2015)
(f) C576 Temp Cap 1: all DNA 16 cycles (as per Schmitt *et al* 2015)
Cap 1: Capture 1; C576: Age-matched control; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe; Test DNA: DNA extracted from cybrid cell lines with wild-type nuclear genome routinely cultured in lab.

The HS D5000 ScreenTape results showed a single peak within the desired range of 300-400 bp for all the six samples (**Fig. 36**).

6.2.1.4 Capture 2

The protocol for the second capture was followed as described in Chapter 3, section 3.2.6.11. The final amplification post-capture 2 was done by adding 30 µl of the PCR master-mix + primers (**Table 29**) to:

1. Test DNA double capture (as per Schmitt *et al* 2015): all DNA (20 µl), 9 cycles
2. Test DNA double capture (Capture 1 – 16 cycles) (as per double-capture Protocol 1 optimization):

    a. Capture 2 – 4 amol 18 cycles
    b. Capture 2 – 4 amol 20 cycles
3. Double capture (as per Schmitt *et al* 2015) Capture 2: half the DNA amount (10 µl), 18 cycles:
    a. NP16.163 Temp (protocol replicate 1)
    b. C576 Temp (protocol replicate 2)

The library amplifications post-capture 2 for the FTD samples used only half the DNA amount and doubled the cycle number compared to (Schmitt et al. 2015) to further modulate the duplex libraries in an attempt to obtain a good peak family size when analysed computationally.

The final HS D5000 ScreenTape results for the post-second capture library amplifications have been illustrated in **Fig. 37**.

The second captures also show a single sharp peak within the desired range of fragment size at higher quantities as seen by the proportion of the size of the peak of interest with respect to the lower and upper markers for the corresponding single captures, indicating that the second capture is highly successful in amplifying desirable molecules in the system. While there also appear to be non-specific higher molecular weight molecules in these second captures, they can be ignored as their peak sizes are almost the same size or smaller than the lower and upper markers of their ScreenTape run (**Figs. 37a, b, d, e**). The only exception to this would be the second capture with 4 amol for 20 cycles (**Fig. 37c**), which is why this was discarded for sequencing.

Thus, a total of six duplex library samples were submitted for Illumina MiSeq in accordance with optimizations for double capture-based duplex library preparations protocol 2 (with custom-designed IDT probes) – amplicon-based (**Fig. 36a**), single-capture (**Fig. 36b),** Test DNA double capture (as per Kennedy *et al* 2014) (**Fig. 37a**), Test DNA double capture (Cap 1: 16 cycles, Cap 2: 4 amol 18 cycles) (**Fig. 37b**), NP16.163 Temp double capture: half the DNA amount (10 µl), 18 cycles (**Fig. 37d**), and C576 Temp double capture: half the DNA amount (10 µl), 18 cycles (**Fig. 37e**).

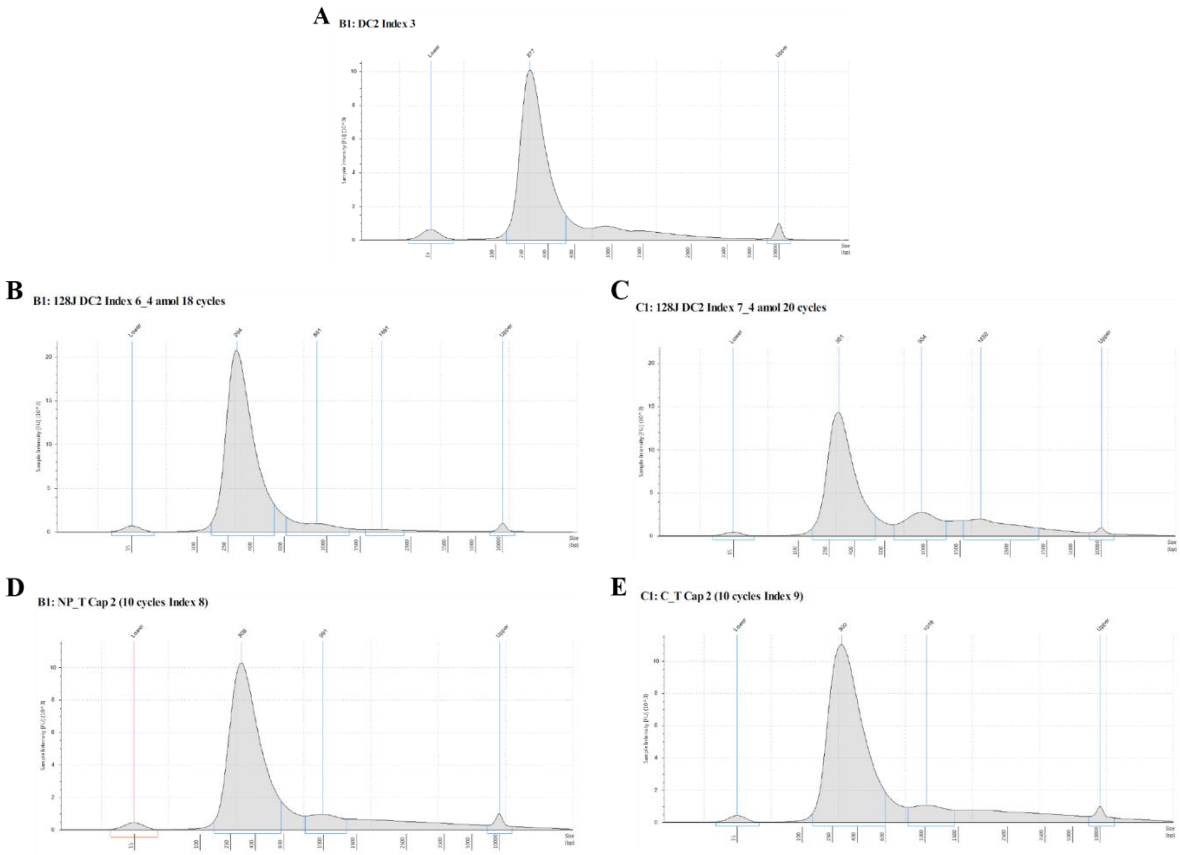


Figure 37: Double capture-based duplex library preparation Protocol 2 – Post-capture 1 library amplifications.

HS D5000 ScreenTape results post-capture 2 library amplifications for:
(a) Test DNA Cap 2: all DNA 9 cycles (as per Schmitt *et al* 2015) – A single, sharp peak within 300-400 bp as desired which is suitable for sequencing
(b) Test DNA Cap 2: 4 amol 18 cycles (as per optimized double-capture protocol 1) – A single, sharp peak within 300-400 bp as desired with negligible larger fragments which is suitable for sequencing
(c) Test DNA Cap 2: 4 amol 20 cycles (as per optimized double-capture protocol 1) – Multiple peaks seen with sharp peak within 300-400 bp but also a significant presence of larger fragments which is unsuitable for sequencing
(d) NP16.163 Temp Cap 2: half DNA 18 cycles (as per Schmitt *et al* 2015) – A single, sharp peak within 300-400 bp as desired with negligible larger fragments which is suitable for sequencing
(e) C576 Temp Cap 2: half DNA 18 cycles (as per Schmitt *et al* 2015) – A single, sharp peak within 300-400 bp as desired with negligible larger fragments which is suitable for sequencing
Cap 2: Capture 2; C576: Age-matched control; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe; Test DNA: DNA extracted from cybrid cell lines with wild-type nuclear genome routinely cultured in lab.

6.2.2 Computational sequencing data analyses

All the bioinformatic analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

6.2.2.1 Family Size

The first analysis done for duplex sequencing is to check if there are enough number of reads to have an optimum family size to be able to generate double consensus. An average peak family size of 6 (reads with duplex consensus, i.e., peak family size with 6 DCS) with around

40-50 raw reads required to make one DCS is optimal for duplex sequencing efficiency (Kennedy et al. 2014).

As seen in **Fig. 38** [x-axis: family size, y-axis: proportion of total reads], the family size varies from sample-to-sample:

A. The GXL-mediated amplicon-based libraries (**Fig. 38a**) have an average family size of 2-3 indicating low DS efficiency; however, the number of raw reads required to make a DCS is close to 44 which suggests that this approach is moderately acceptable for our purposes.
B. A single capture (**Fig. 38b**) is not enough for DS as not only is the average family size close to 1, but the number of raw reads required to make a DCS is close to 2000, which shows that the IDT capture protocol cannot be performed as suggested for DS.

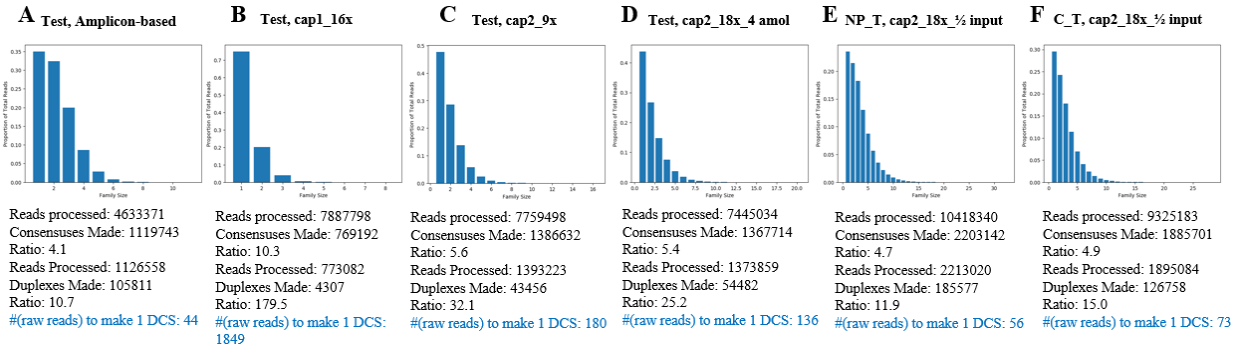


Figure 38: Double capture-based Protocol 2 – Family size calculations.
(a) *Amplicon-based Test DNA* – shows a reasonably good peak family size as desired. (b) *Single capture Test DNA* – shows poor peak family size and poor capture quality. (c) *Test DNA Cap 2 (all DNA 9 cycles)* (as per Schmitt *et al* 2015) – shows an adequately good peak family size. (d) *Test DNA Cap 2 (4 amol 18 cycles)* – Further optimization as per double-capture protocol 1 gives a comparable peak family size to (Schmitt *et al* 2015), suggesting this is not a particularly advantageous optimization. (e) *NP16.163 Temp Cap 2 (half DNA 18 cycles) and* (f) *C576 Temp Cap 2 (half DNA 18 cycles)* – both the FTD samples give the best family size with a peak family size of 6 (optimal).
Cap 2: Capture 2; C576: Age-matched control; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe; Test DNA: DNA extracted from cybrid cell lines with wild-type nuclear genome routinely cultured in lab.

C. A double capture as adapted from (as per Schmitt *et al* 2015) (**Fig. 38c**) gives a lower average family size than the amplicon-based approach with ~150+ raw reads required to make a DCS, suggesting that this protocol is not as good for our purposes.
D. A double capture with the first capture amplification performed as in (Schmitt *et al* 2015) but with the second capture amplification performed with 4 amol for 18 cycles (**Fig. 38d**) offers only a slight improvement in average size and the number

of raw reads required, which proves that not only do DNA input and cycle number influence the family size and number of raw reads needed to make a DCS, these are also extremely crucial and difficult parameters to modulate without actual sequencing.

E. The final condition of performing first capture amplification as in (Schmitt *et al* 2015) and performing the second capture amplification with half the DNA input for 18 cycles for NP16.163 (**Fig. 38e**) & C576 (**Fig. 38f**) Temp samples show the best average family size of all the present samples at ~5-6 and required ~50-70 raw reads to make a DCS which makes this the condition with the most efficient parameters.

6.2.2.2 On-target ratio

The next analysis for the double captures is to look at the on-target ratio, i.e., the number of raw reads found within the region of interest i.e., the MAPT gene on chromosome 17.

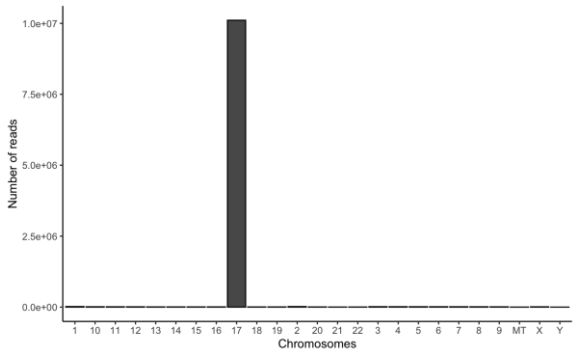


Figure 39: Double capture-based Protocol 2 – On-target ratio calculations.
The proportion of reads covering chromosome 17 (containing MAPT gene) is significantly higher than the rest of the genome indicating a notable enrichment and targeted-capture of MAPT DNA, suggesting that the custom-designed IDT probes were more than adequate for our purposes.
DNA: Deoxyribonucleic acid; MAPT: Microtubule-associated protein tau.

Fig. 39 is the calculation of the on-target ratio of a double-capture sample [x-axis: chromosomes, y-axis: number of reads] and the high number of reads corresponding to the MAPT gene on the reference genome shows that there is a highly significant enrichment of our regions of interest, indicating that the double capture protocol was successful.

6.2.2.3 Number of reads: Q30 vs DCS

The number of raw reads that passed the Q30 filters *versus* the duplex consensus reads that show an alignment to the region of interest further illuminates the efficiencies of DS in each sample by giving a graphical representation of the number of raw reads corresponding to chromosome 17 *vs* the rest of genome (top row) and the number of DCS reads corresponding

to chromosome 17 *vs* the rest of genome (bottom row) (**Fig. 40**) [x-axis: method, y-axis: number of reads, light blue: chromosome 17, dark blue: rest of the genome]:

a) When comparing the number of raw reads (top row) *versus* DCS (bottom row) across samples, it is clear that duplex consensus results in significant enrichment of targeted genome by a power of at least 10^5 with respect to the whole genome, an expected outcome considering the design of DS.
b) Within each type of read:
    i. The amplicon-based approach, as the experimental control, has a decent enrichment ratio (**Fig. 40a**) with comparable enrichment ratios in raw *vs* DCS reads as only the MAPT regions were amplified first and then duplex libraries prepared for sequencing.
    ii. The enrichment shown in the single capture is the lowest of all samples (**Fig. 40b**) in raw reads. While the DCS reads help eliminate a lot of false-positives from non-targeted genomic regions, the enrichment ratio at this level is still the lowest of all DCS enrichment ratios.
    iii. The targeted enrichment ratios of the double capture (Schmitt *et al* 2015) protocols adapted as is (Cap 2: all DNA, 9 cycles) (**Fig. 40c**) and modified (Cap 2: 4 amol, 18 cycles) (**Fig. 40d**) are higher than the experimental control which indicates that the double capture method is the superior method. However, they have comparable enrichment ratios at the levels of both raw reads as well as DCS reads, suggesting that this modification of input and cycle numbers for library amplification did not improve the DCS reads.
    iv. Finally, the double capture of the NP16.163 & C576 Temp (**Fig. 40e & f**) samples (Cap 2: half input of DNA, 18 cycles) have the most significant enrichment of chromosome 17 as compared to the rest of the genome, especially at the DCS level and reflects the results obtained from family size qualitative and quantitative calculations.

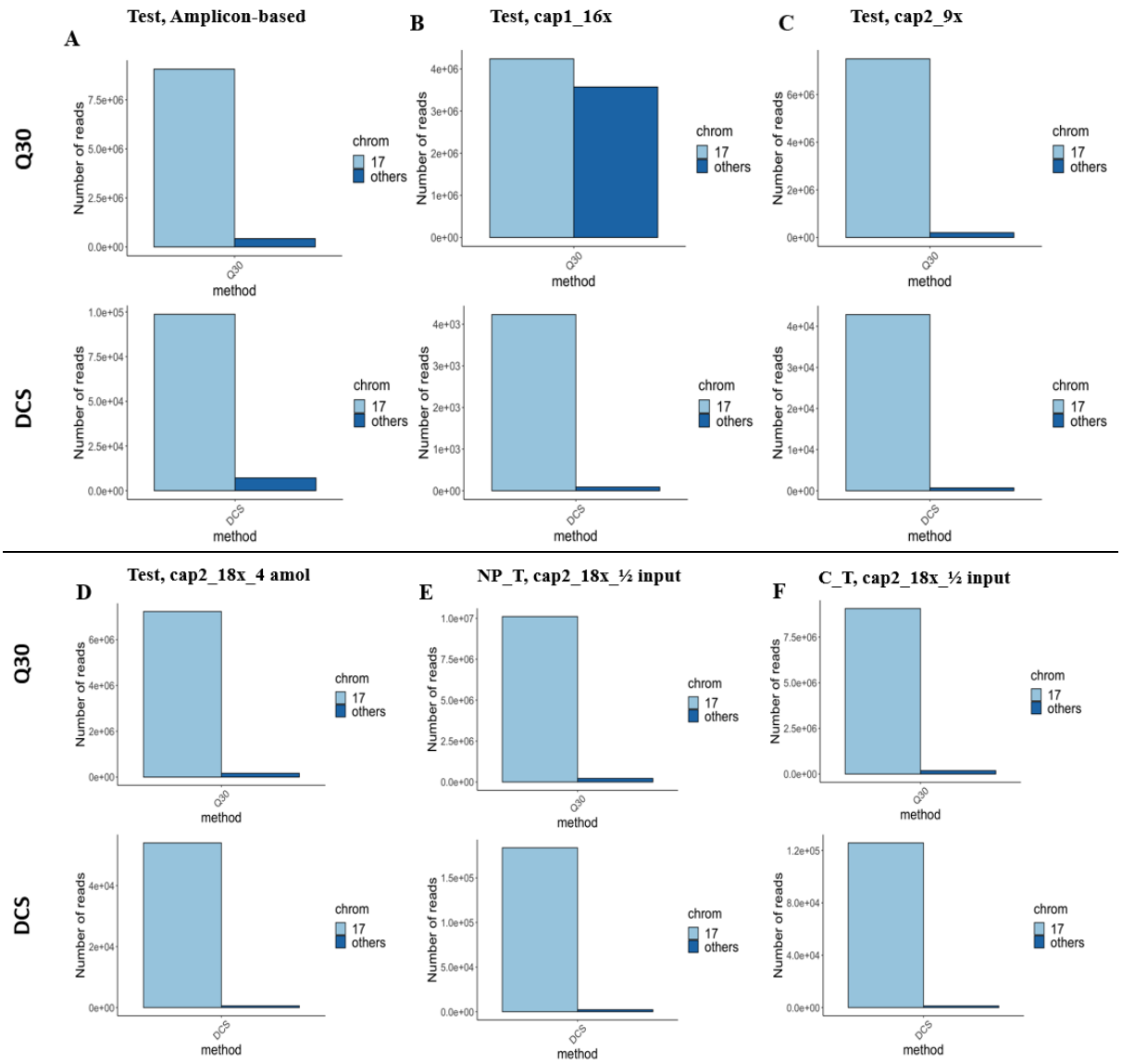


Figure 40: Double capture-based Protocol 2 – Number of reads.

The number of raw reads (Q30) (Top row) *versus* DCS reads (bottom row) for each sample:

(a) *Amplicon-based Test DNA* – shows a reasonably good enrichment of chromosome 17 compared to rest of the genome. Number of DCS reads is comparable to Q30 within chromosome 17 as expected.

(b) *Single capture Test DNA* – shows poor peak family size and poor capture quality. of chromosome 17 compared to rest of the genome. Number of DCS reads is comparable to Q30 within chromosome 17, further indicating the weakness of a single capture for duplex sequencing.

(c) *Test DNA Cap 2 (all DNA 9 cycles)* (as per Schmitt *et al* 2015) – shows an adequately good enrichment of chromosome 17 compared to rest of the genome. Number of DCS reads is lower than Q30 within chromosome 17, indicating a good targeted capture of MAPT.

(d) *Test DNA Cap 2 (4 amol 18 cycles)* – Further optimization as per double-capture protocol 1 gives a comparable enrichment of chromosome 17 as (Schmitt *et al* 2015) sample above, both, with respect to chromosome 17 enrichment compared to rest of the genome and DCS *vs* raw reads, suggesting this is not a particularly advantageous optimization.

(e) *NP16.163 Temp Cap 2 (half DNA 18 cycles) and* (f) *C576 Temp Cap 2 (half DNA 18 cycles)* – both the FTD samples give the best enrichment of chromosome 17 compared to rest of the genome. Number of DCS reads is much lower than Q30 within chromosome 17, indicating a substantial targeted capture of MAPT.

Cap 2: Capture 2; C576: Age-matched control; DCS: Duplex consensus sequences; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Q30: raw reads; Temp: Temporal lobe; Test DNA: DNA extracted from cybrid cell lines with wild-type nuclear genome routinely cultured in lab.

6.2.2.4 Depth of Sequencing

Sequencing depth refers to the number of times that a given nucleotide in the genome has been read while sequencing so that base calls may be made with a higher degree of confidence with higher depth.

The sequencing depth for the double capture experiments was measured by comparing base calls using raw reads (passed Q30 filters) *versus* single consensus (SSCS) *versus* duplex consensus (DCS) (**Fig. 41**) [x-axis: chromosomal coordinates, y-axis: depth].

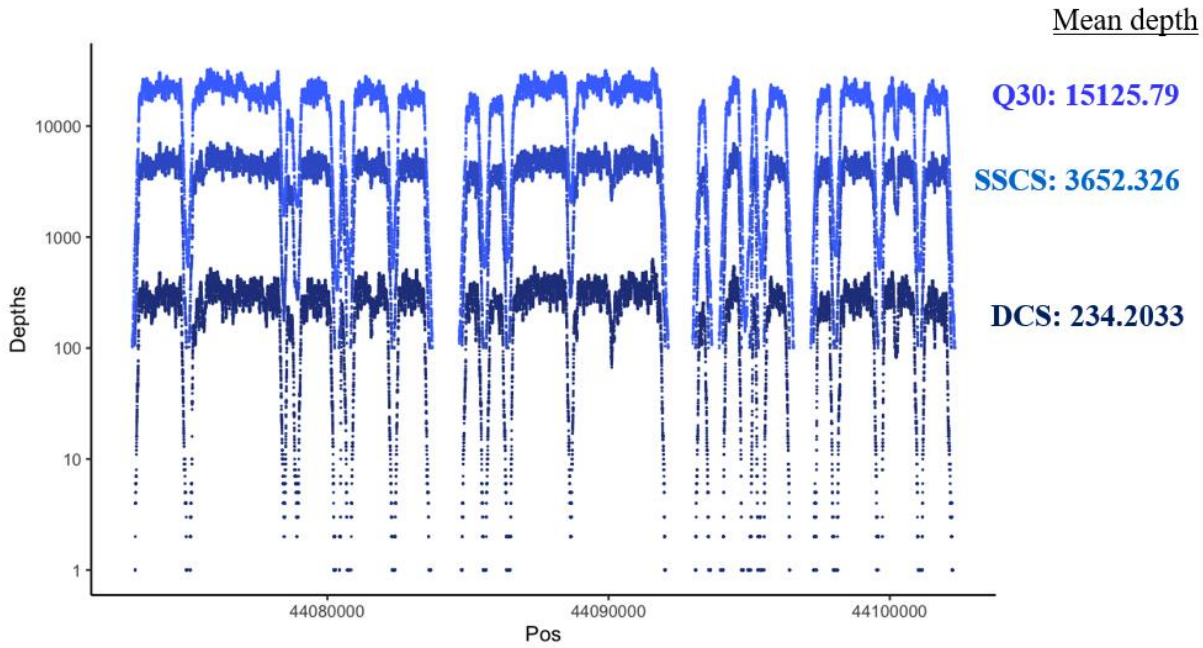


Figure 41: Double capture-based Protocol 2 – Depth of sequencing.
The depth of raw reads is the highest (with an average coverage of ~15,000X), followed by single consensus read depth (with an average coverage of ~3700X), both of which are much higher than the average coverage of ~230X after duplex consensus, indicative of an adequate quality of sequencing data.
DCS: Duplex consensus sequences; Q30: raw reads; SSCS: Single strand consensus sequences.

There is a good sequencing coverage across all the three amplicons designed with an expected decrease in the depth of sequencing comparing raw reads (light blue) to single-strand consensus (blue) to duplex consensus (dark blue) with an average depth of 15,125.79, 2652.33, and 234.20 respectively. The depth is consistent across the three amplicons in each case, and the staggered spots below are most likely to be representative of repetitive and other polymeric regions of the introns which are quite difficult to capture and get an accurate depth for and so can be ignored at this point.

6.2.2.5 Mutation detection

Mutation detection was done for the two FTD samples made from the IDT protocol (Cap 1: all DNA, 16 cycles and Cap 2: half the DNA, 18 cycles) – FTD-Tau disease temporal lobe (NP16.163 Temp) and control temporal lobe (C576 Temp).

The number of mutations with both, VAF > 0.25 and VAF < 0.25, were calculated and compared (**Fig. 42**).

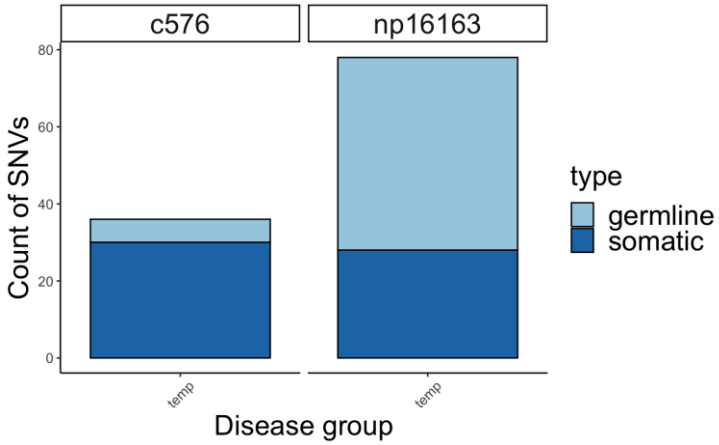


Figure 42: Double capture-based Protocol 2 – Detection of mutation frequencies.
The number of mutations detected in Control Temp (left) *versus* FTD Disease Temp (right) samples. Number of germline mutations (light blue) in disease is much higher than in control. Somatic SNVs (dark blue) is comparable in both cases.
C576: Age-matched control; FTD: Frontotemporal dementia; SNVs: Single nucleotide variations; Temp: Temporal lobe.

The quantitative analysis of mutation detection done from IDT capture-mediated duplex libraries showed that while the somatic mutational burden in disease (28 SNVs) was comparable to those in control (30 SNVs), the germline mutational burden (VAF > 0.25) in FTD disease (50 SNVs) is much higher than in FTD control (6 SNVs), suggesting that germline mutations could also be associated with FTD pathology in this patient who is a sporadic case of FTD-Tau/Pick's.

6.3 Discussion

This chapter delves into the optimization done for the xGen IDT protocol for hybridization and capture, referred to as 'double capture-based duplex library preparation protocol 2' in this thesis. In this protocol, custom-designed probes for the chromosomal coordinates on chromosome 17 that coded for MAPT gene were used. With ~160 the probes designed for the '+' strand, we needed to only optimize for library amplification for capture 2 of the protocol. Several conditions were tested on 'test' DNA extracted from cell lines to determine the best

parameters for our purposes and two FTD temporal samples of disease and control, along with one amplicon-based (GXL) sample on 'test' DNA as an experimental control.

While it was immediately clear that the single capture as recommended by IDT was not enough for duplex sequencing, it was interesting to note that the two double captures with 'test' DNA showed comparable family sizes for 'capture 2: all DNA, 9 cycles' and 'capture 2: 4 amol 18 cycles', suggesting that the relationship between DNA input and cycle number (for library amplification) to the family size is more complex than expected. Duplex libraries made with 'half the input, 18 cycles' had the best peak family size and the 'number of raw reads required to make a DCS', indicating that this condition was optimal for double capture. All these IDT double captures had better family sizes than the amplicon-based (GXL) experimental control, which, along with the excellent on-target ratios, show that the IDT double capture method may be the superior approach of all those attempted for optimization of the duplex sequencing protocol.

The number of reads as well as the depth of sequencing decreased from raw reads to single consensus (SSCS) to duplex consensus (DCS) amongst the double capture samples. The germline mutational burden in the FTD-Tau disease Temp was approximately that eight times of Control Temp while the number of germline SNVs was about 1.8 times higher compared to somatic SNVs in FTD-Tau disease Temp, indicating that the disease state could also be associated with inherited SNVs in this patient who was a sporadic case of FTD-Tau/Pick's.

With the two optimized approaches to duplex sequencing, we prepared duplex libraries on three FTD-Tau disease and their corresponding age-matched controls via the SuperFi DNA polymerase-mediated amplicon-based method as well as double capture-based duplex library preparation protocol 2 (with custom-designed IDT probes), and these results are described in the next chapter.

Chapter VII: Optimized Duplex Sequencing on all FTD-Tau samples

7.1 Introduction

This chapter comprises of the results generated from the pre-library amplifications done for the three FTD-Tau brain regions their age-matched control with the high-fidelity Platinum SuperFi II DNA polymerase, generating a total of eighteen MAPT amplicons for the six FTD-Tau samples. Then, the three MAPT amplicons made from each sample were combined in equimolar proportions that generated a final of six mixes that underwent then duplex library preparation. Library amplification of the adapter-ligated DNA was done with 4 amol DNA input for 20 cycles and the ScreenTape results for the final twelve duplex libraries that were then taken for Illumina MiSeq have been illustrated. This is followed by an outline of the primary findings after the appropriate computational analyses of the mutation frequency of germline and somatic mutations found in each sample and explores the comparisons between each sample corresponding to the different brain regions (Temp *vs* Occ *vs* Med) in FTD-Tau disease case *versus* control.

The chapter then delves into the optimized xGen IDT protocol for double capture-based approach (with custom-designed IDT probes) that calls for duplex library preparation for all the DNA samples by first constructing duplex libraries with the KAPA HyperPlus Kit, adding blocking oligos to 500 ng of these duplex libraries, lyophilization followed by adding the custom-designed probes for both the '+' and '-' strands of regions of interest, 16 hours of heated incubation, mixing streptavidin beads to the baits-duplex libraries mixture, performing the heated and room temperature washes, library amplification and finally, repeating the capture process on the amplified libraries obtained from capture 1, before sequencing on the Illumina NovaSeq platform. Descriptions of the HS D5000 ScreenTape results generated after the library amplifications of both the captures as well as the computational analyses done to determine the mutation frequencies of germline and somatic mutations detected and the possible points of origin of these mutations in the entire MAPT gene (~140 kb) in all the six DNA samples have been reported. Since the probes captured libraries from both, '+' (template) and '-' (non-template), strands of the entire MAPT gene as opposed to only '+' strand of exons

8-12 of the MAPT gene in the previous chapter, further optimizing that needed to be done on the post-capture 2 library amplification step have also been documented in this chapter. A discussion on the primary findings of the computational analyses follows. Finally, the chapter concludes with a comparison of the various approaches to duplex sequencing and the implications of the information gleaned on our understanding of the true mutation frequencies on FTD-Tau pathogenesis.

7.2 Results

7.2.1 Amplicon-based approach to duplex sequencing

7.2.1.1 Pre-library amplifications

50 ng of the DNA from the Temp, Occ, and Med samples of NP16.163 (disease) as well as C576 (control) were taken as the initial DNA input for the pre-library amplification of each of the three MAPT amplicons with Platinum SuperFi II polymerase (see Chapter 3, section 3.2.4.1 for amplicons primer description and PCR programmes used).

A single band was seen for all 18 lanes at desirable band sizes on a 0.7x agarose gel (see Chapter 3, section 3.2.4.2 for gel preparation and image acquisition): ~1.3 kb for MAPT-A amplicons and ~7.0 kb for MAPT-B and C amplicons (**Fig. 43**). These amplicons were also now ready for duplex library preparation.

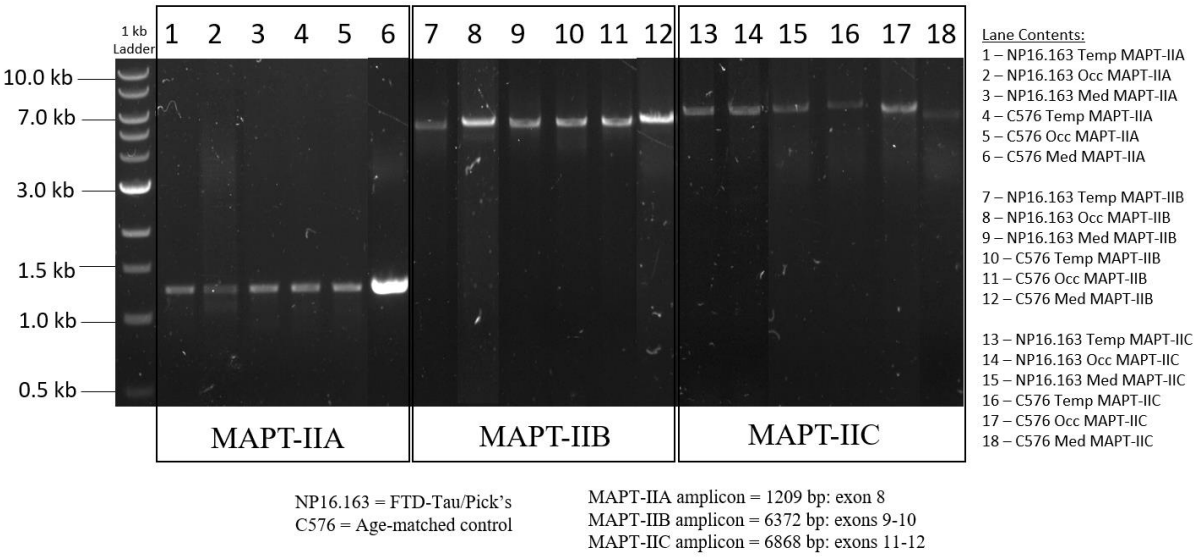


Figure 43: Pre-library amplifications of the three MAPT amplicons with the SuperFi polymerase.
SuperFi-mediated pre-library amplification done with all six DNA samples for MAPT-A (lanes 1-6), MAPT-B (lanes 7-12), and MAPT-C (lanes 13-18) amplicons against 1 kb and 1 kb-extend DNA ladders.
C576: Age-matched control; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; Med: Medulla; NP16.163: FTD-Tau disease case; Occ: Occipital lobe; Temp: Temporal lobe.

7.2.1.2 Combining the three MAPT amplicons in equimolar proportions for each FTD sample

As per our calculations using the Weight to Molar Quantity for Nucleic Acids Calculator (https://www.bioline.com/media/calculator/01_07.html), the addition of 43.5 fmol of the three MAPT amplicons makes the final amount of DNA to be ~100 ng which is then used for duplex library preparation. The volume of DNA required from each amplicon was calculated by first dividing the amount of DNA required to make 43.5 fmol of DNA by the molecular weight of each amplicon. This was then divided by the concentration of each amplicon to calculate the volume of amplicon needed to be added make the final equimolar mixture for a FTD sample.

For instance, the amount of each amplicon added to the final mixture of NP16.163 Temp was calculated as follows (**Table 32**):

Amplicon	Mol. Wt. (kb)	Amt. of DNA (ng) for 43.5 fmol	Amplicon DNA concentration (ng/µl)	Amplicon DNA vol. needed (µl)
MAPT-A	1.209	35.57	5.58	5.27
MAPT-B	6.372	179.89	5.75	4.91
MAPT-C	6.868	193.89	4.32	6.54

Table 32: Calculations for combining the three MAPT amplicons in equimolar proportions.
Calculations done for all six DNA samples (NP16.163 Temp (SF) here as an example):

$$Volume\ of\ amplicon\ DNA\ needed\ = \frac{Amount\ of\ DNA\ for\ 43.5\ fmol}{Molecular\ weight\ of\ amplicon\ \times Amplicon\ DNA\ Concentration}.$$

DNA: Deoxyribonucleic acid; fmol: femtomoles; NP16.163: FTD-Tau disease case; SF: SuperFi DNA polymerase; Temp: Temporal lobe.

Similar calculations were done for the remaining five SuperFi-mediated amplicon preparations.

7.2.1.3 Duplex Library Preparations

The duplex libraries were prepared as described in Chapter 3, section 3.2.4.3, using the SuperFi-mediated MAPT amplicons (in equimolar) preparations. The HS D1000 and HS D5000 ScreenTape results after adapter-ligation – three main peaks at ~200, 400 and 600 bp – and library amplification (4 amol DNA input, 20 cycles) – one sharp peak at ~300 bp – were as desired for all six FTD samples after SuperFi-mediated amplicon-based duplex library preparation.

An example of these ScreenTape results for NP16.163 Temp (SuperFi) have been outlined as follows (**Fig. 44**) (See *Appendix IV* for all SuperFi-mediated amplicon-based duplex library preparation ScreenTape results):

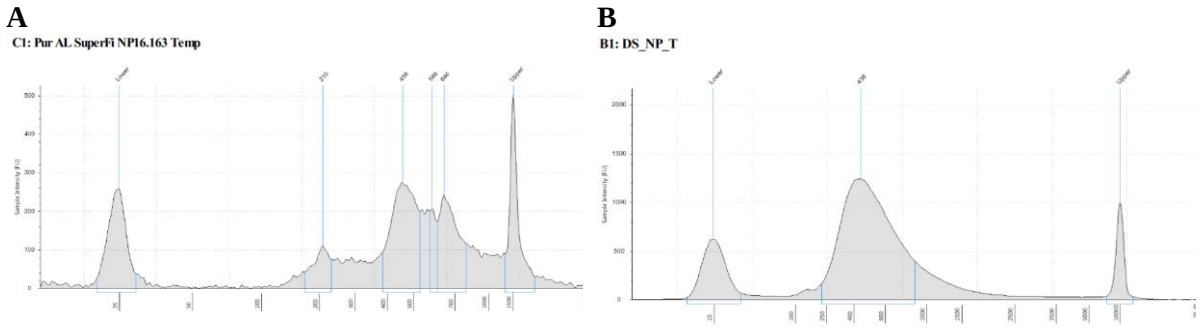


Figure 44: Amplicon-based duplex library preparation (SuperFi-mediated).
(A) *Adapter-ligation and purification*: Three peaks seen as expected in the adapter-ligated DNA library (0.7X Ampure XP beads purification) on the HS D1000 ScreenTape. (B) *Library amplification and purification*: A single peak within 300-400 bp range as expected in the duplex library (0.7X Ampure XP beads purification) on the HS D5000 ScreenTape. Here, for SuperFi-mediated NP16.163 Temp as an example.
DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe.

These amplified libraries were then prepared for sequencing on the Illumina NovaSeq platform.

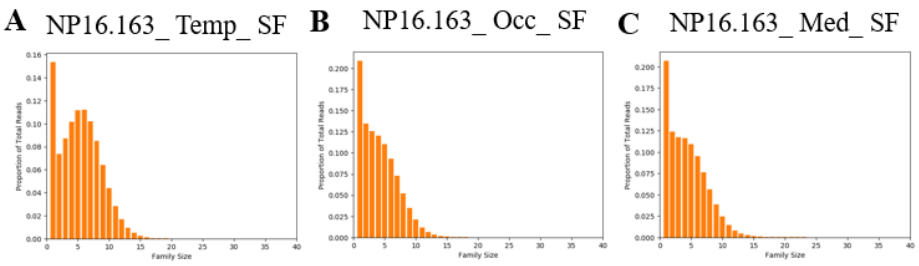
7.2.1.4 Computational Sequencing Data Analyses

All the bioinformatic analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

1) Family Size

The first analysis done for duplex sequencing is to check if there are enough number of reads to have an optimum family size to be able to generate double consensus. Again, **family size** refers to the number of 'sequencing reads' that share the same tag sequence that form one family that can eliminate false-positive calls using appropriate bioinformatic pipelines. An average peak family size of 6 (reads with duplex consensus, i.e., peak family size with 6 DCS) with around 40-50 raw reads required to make one DCS is optimal for duplex sequencing efficiency (Kennedy et al. 2014).

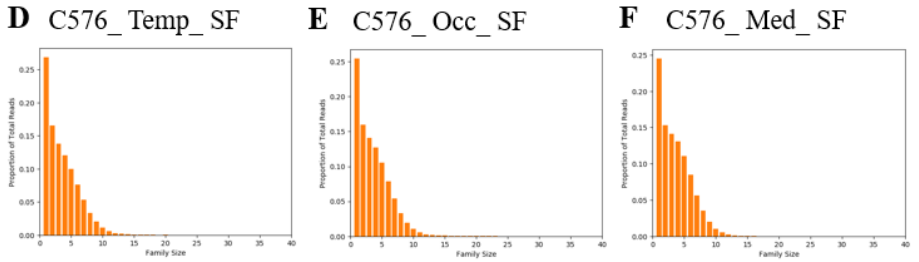
Fig. 45 [x-axis: family size, y-axis: proportion of total reads] shows that the average peak family size as well as the number of raw reads to make a DCS were at the optimal levels, indicating a successful duplex library construction for all the six FTD-Tau samples generated using SuperFi-mediated amplicon-based DS.



Summary Statistics:
Reads processed:5058956
Bad reads: 4603144
Reads with Bad Flags: 778003
Overlapping Reads: 3626863
Soft-clipped Reads: 197844
Repetitive Duplex Tag: 434
Reads with Less Common Cigar Strings: 11
Consensuses Made: 68388
Consensuses with Too Many Ns: 25
Q30/SSCS: 73.97
Reads Processed: 68344
Duplexes Made: 12272
Unpaired Duplexes: 5178
N-clipped Duplexes: 0
SSCS/DCS: 5.57
#raw reads to make 1 DCS: 412.01

Summary Statistics:
Reads processed:5802074
Bad reads: 5329571
Reads with Bad Flags: 1008031
Overlapping Reads: 4097055
Soft-clipped Reads: 223978
Repetitive Duplex Tag: 507
Reads with Less Common Cigar Strings: 9
Consensuses Made: 69706
Consensuses with Too Many Ns: 34
Q30/SSCS: 83.24
Reads Processed: 69662
Duplexes Made: 12215
Unpaired Duplexes: 5833
N-clipped Duplexes: 0
SSCS/DCS: 5.70
#raw reads to make 1 DCS: 474.468

Summary Statistics:
Reads processed:4309686
Bad reads: 3931633
Reads with Bad Flags: 521289
Overlapping Reads: 3223504
Soft-clipped Reads: 186506
Repetitive Duplex Tag: 334
Reads with Less Common Cigar Strings: 29
Consensuses Made: 53660
Consensuses with Too Many Ns: 43
Q30/SSCS: 80.31
Reads Processed: 53583
Duplexes Made: 9449
Unpaired Duplexes: 4695
N-clipped Duplexes: 0
SSCS/DCS: 5.57
#raw reads to make 1 DCS: 447.33



Summary Statistics:
Reads processed:4349815
Bad reads: 3269834
Reads with Bad Flags: 375102
Overlapping Reads: 2652557
Soft-clipped Reads: 241462
Repetitive Duplex Tag: 713
Reads with Less Common Cigar Strings: 63
Consensuses Made: 154956
Consensuses with Too Many Ns: 56
Q30/SSCS: 28.07
Reads Processed: 154868
Duplexes Made: 34226
Unpaired Duplexes: 13752
N-clipped Duplexes: 0
SSCS/DCS: 4.52
#raw reads to make 1 DCS: 126.88

Summary Statistics:
Reads processed:6395874
Bad reads: 5853928
Reads with Bad Flags: 790230
Overlapping Reads: 4782685
Soft-clipped Reads: 280540
Repetitive Duplex Tag: 473
Reads with Less Common Cigar Strings: 23
Consensuses Made: 77847
Consensuses with Too Many Ns: 51
Q30/SSCS: 82.16
Reads Processed: 77773
Duplexes Made: 13987
Unpaired Duplexes: 6871
N-clipped Duplexes: 0
SSCS/DCS: 5.56
#raw reads to make 1 DCS: 456.81

Summary Statistics:
Reads processed:5354982
Bad reads: 4881562
Reads with Bad Flags: 775160
Overlapping Reads: 3902345
Soft-clipped Reads: 203641
Repetitive Duplex Tag: 416
Reads with Less Common Cigar Strings: 8
Consensuses Made: 70212
Consensuses with Too Many Ns: 26
Q30/SSCS: 76.27
Reads Processed: 70173
Duplexes Made: 12198
Unpaired Duplexes: 5548
N-clipped Duplexes: 0
SSCS/DCS: 5.75
#raw reads to make 1 DCS: 438.56

Figure 45: SuperFi-mediated amplicon-based DS – Family size calculations.
(Peak family size, number of raw reads to make a DCS) calculated for: (a) NP16.163 Temp (~6, 412), (b) NP16.163 Occ (~ 3, 475), (c) NP16.163 Med (~ 4, 447), (d) C576 Temp (~ 3, 127), (e) C576 Occ (~ 3, 457), (f) C576 Med (~ 3, 439). ~400+ raw reads needed to make a DCS indicating lowered DS efficiency; however, adequate peak family size seen for each sample (optimal ~ 6) indicating successful duplex library construction. DCS: Duplex consensus sequences; DS: Duplex sequencing; C576: FTD-Tau control; FTD: Frontotemporal dementia; Med: Medulla; NP16.163: FTD-Tau disease case; Occ: Occipital lobe; SF: SuperFi polymerase; Temp: Temporal lobe.

2) Depth of Sequencing

Depth of Sequencing refers to the number of times that a given nucleotide in the genome has been read while sequencing so that base calls may be made with a higher degree of confidence as the depth increases.

Fig. 46 [x-axis: chromosomal coordinates, y-axis: depth] shows the depth of sequencing across the three MAPT amplicons which were mixed in equimolar proportions before duplex library preparation for all the six DNA samples. The samples showed high depth and were mostly homogeneous and consistent across the MAPT amplicons for each sample which indicated that the sequencing run was of good quality for further analyses.

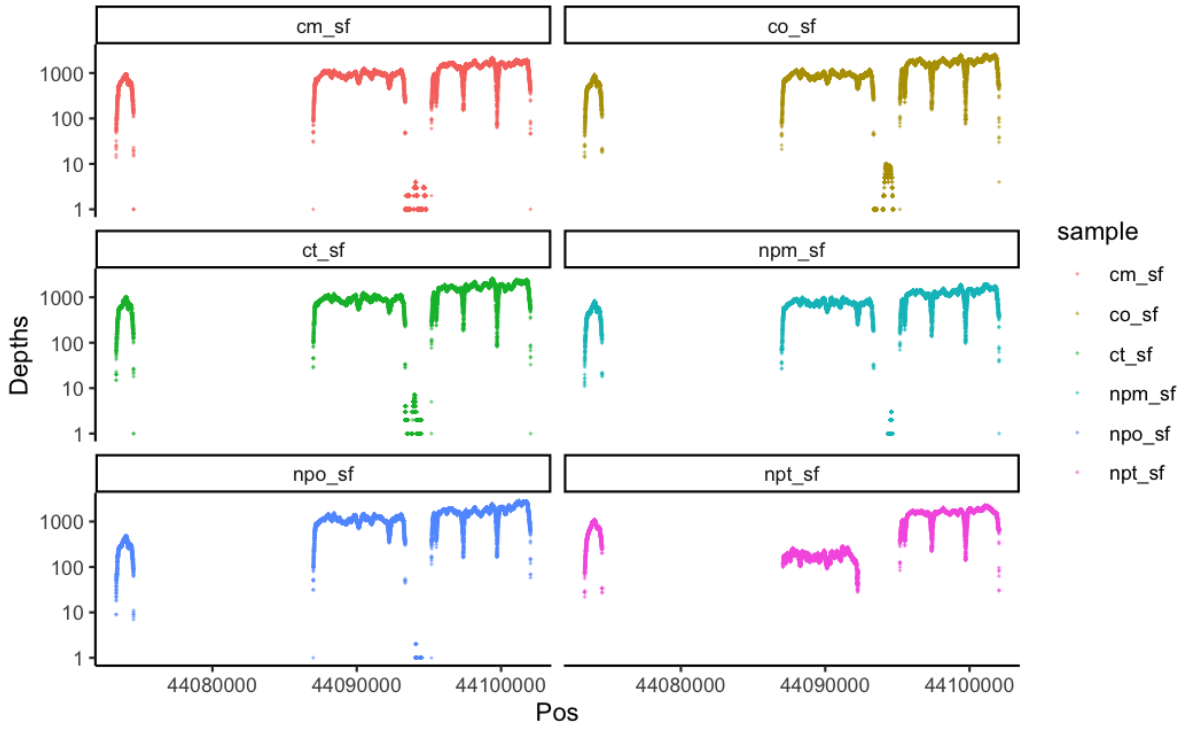


Figure 46: SuperFi-mediated amplicon-based DS – Depth of Sequencing.
Depth of sequencing for all six DNA samples: (a) NP16.163 Temp: 913.32X, (b) NP16.163 Occ: 925.59X, (c) NP16.163 Med: 1,296.06X, (d) C576 Temp: 1,147.76X, (e) C576 Occ: 1,037.98X, (f) C576 Med: 1,146.39X. High and homogeneous coverage across all three amplicons indicating the mixing the MAPT amplicons in equimolar proportions was successful.
C576: Age-matched control; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; MAPT: Microtubule-assisted protein tau; Med: Medulla; NP16.163: FTD-Tau disease case; Occ: Occipital lobe; Temp: Temporal lobe.

3) Mutation Detection

- ***Number of mutations detected:***

The overall mutation burden in FTD-disease case was higher than in control.

These 'true' mutations detected after duplex consensus were then broadly divided into two classes: those with a variable allele frequency (VAF) > 0.25, termed as germline mutations, and those with a VAF < 0.25, termed as somatic mutations.

First, the mutational burdens of somatic and germline SNVs were calculated (**Fig. 47**) [x-axis: FTD-Tau samples, y-axis: number of paired reads]. In control, the number of somatic SNVs was much higher than the number of germline SNVs per brain region while in FTD-disease case, the number of somatic SNVs was only slightly higher than the number of germline SNVs per brain region, indicating that despite being a sporadic case, the FTD-disease state was likely to be associated with both, somatic and inherited, mutations in the patient.

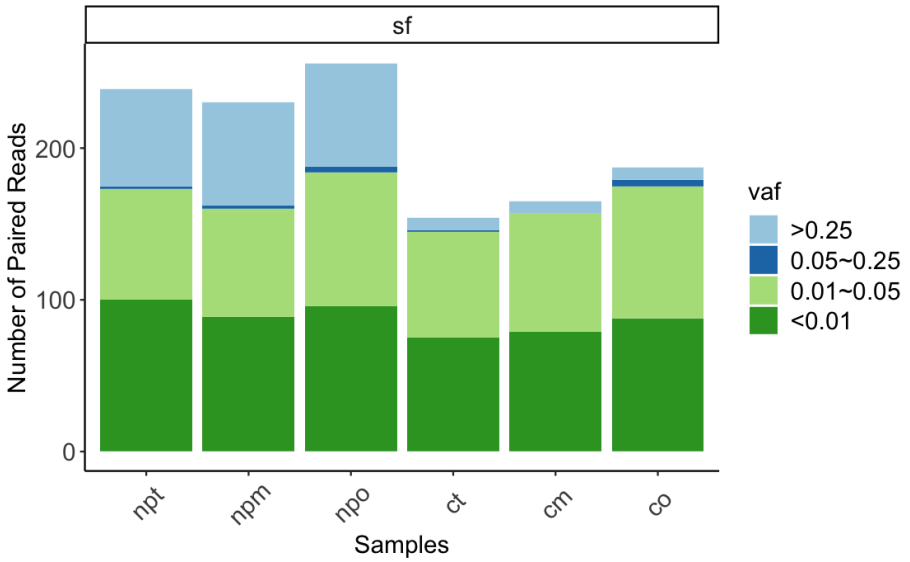


Figure 47: SuperFi-mediated amplicon-based DS – Mutational burden detection.

Mutational burden detection in the MAPT gene for the three FTD-Tau samples and age-matched controls: Germline SNVs with VAF > 0.25 (light blue) were higher in the three brain regions of FTD-Tau disease than control. Somatic SNVs with VAF ~ 0.05-0.25 (dark blue) represented a few early-arising somatic SNVs that were propagated in more brain cells. While the number of somatic SNVs with low (VAF ~ 0.01-0.05 (pale green)) and ultra-low frequencies (VAF < 0.05 (dark green)) were comparable with each other in all three brain regions of both disease case and control, the total somatic mutational burden was higher than the germline mutational burden in all the six DNA samples being considered. Thus, this figure indicates that somatic SNVs (VAF < 0.05) as well as germline SNVs (VAF > 0.25) may be contributing to FTD pathogenesis in the patient.
C_T, O, M: Age-matched control Temporal lobe, Occipital lobe, Medulla, respectively; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; NP_T, O, M: FTD-Tau disease case Temporal lobe, Occipital lobe, Medulla, respectively; SNVs: Single nucleotide variations; VAF: Variable allele frequency.

This statement is also supported by the statistical analysis, Student's t-test, which revealed that not only was the number of ultra-low frequency somatic mutations (VAF < 0.01) in FTD-disease *vs* control statistically significant (Student's t-test, *p = 0.0476), but also that the germline mutational burden (VAF > 0.25) in FTD-Tau disease case was statistically very significant (Student's t-test, **p = 0.0052) compared to control (**Table 33**).

Variable Allele Frequency (VAF)	Type of mutation	p-value	Significance
VAF ≥ 0.25	Germline	0.00052	**
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	0.93170	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.89730	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.04762	*

Table 33: SuperFi-mediated amplicon-based DS – Student's t-test to calculate significance of mutational burden in FTD-disease *vs* control.

Note: ***Student's t-test*** *is a statistical method of testing hypotheses about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown. The* ***p-value****, or probability value, gives the likelihood of the given data occurring under the null hypothesis; here, the null hypothesis is that there is no relationship between the variables of interest or that there is no difference among groups being considered.* ***Statistical significance*** *is a measure of whether the p-value of the test is small enough to reject the null hypothesis of the test. The most common threshold is p < 0.05.*

- ***Distribution of mutations detected:***

The VAF distribution of these mutations within each sample was calculated in **Fig. 48.** The top row of the figure represents data analyses of germline SNVs (VAF > 0.25) while the bottom row represents data analyses of somatic SNVs (VAF < 0.25).

Figs. 48 a & c are density plots of the mutations detected in FTD-disease and control [x-axes for each brain region: VAF (0.00-1.00); y-axis: Density (0-60)]. **Fig. 48a** shows that the relative peak density, distribution, and density curves of somatic SNVs (VAF < 0.25) is only slightly higher than and is quite comparable with those of germline SNVs (VAF > 0.25) in all the three brain regions of FTD-disease while in control, the relative peak densities and distributions of somatic SNVs (VAF < 0.25) are much higher and sharper density curves than those of germline SNVs (VAF > 0.25) are seen in all the three brain regions. When further focusing only on the SNVs with VAF < 0.15, the relative peak densities, distributions, and density curves of these somatic SNVs were comparable across all the six brain regions FTD-disease and control (**Fig. 48c**). Thus, the comparable densities of somatic and germline SNVs in FTD-disease and the contrasting density distributions of both mutation types in control indicates that

higher numbers of somatic and germline SNVs are likely to be associated with the disease state and the lower VAF somatic SNVs are likely associated with ageing.

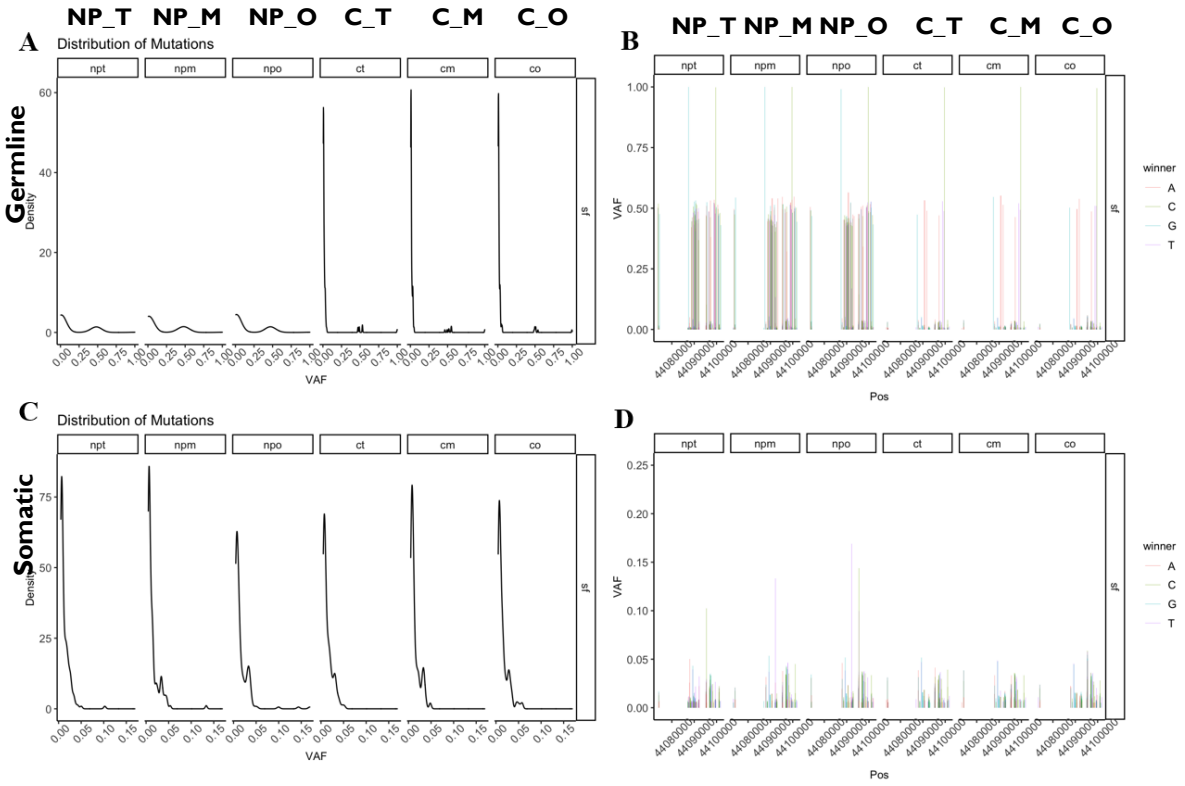


Figure 48: SuperFi-mediated amplicon-based DS – Distributions of detected mutations.

Distributions of mutations seen in –

Germline mutations (top row):

(a) *Density plot* – (x-axes for each brain region: VAF ranging from 0.00 to 1.00; y-axis: Density ranging from 0-60) The relative peak density of somatic SNVs (VAF < 0.25) was comparable (at an average density of 6) with that of germline SNVs (VAF ~ 0.50) (at an average density of 4) for each brain region in FTD-disease case which indicated that both mutation types had comparable distributions and density curves in the patient while the relative peak density of somatic SNVs (VAF < 0.05) is much higher (at an average density of 60) than that of germline SNVs (VAF ~ 0.50) (at an average density of 2) for each brain region in control which indicated that somatic SNVs had a much greater difference in the distribution and a sharper density curve than that of germline SNVs in the age-matched control.

(b) *VAF of each nucleotide across chromosomal positions* – (x-axes for each brain region: chromosomal position of the MAPT gene, y-axis: VAF ranging from 0.00 to 1.00)

A higher proportion of variations in nucleotides of germline SNVS (VAF > 0.25) is seen in each chromosomal position of the MAPT gene of FTD-Tau disease compared to control whereas variations in nucleotides of somatic SNVs (VAF < 0.25) in each chromosomal position of MAPT gene is comparable across disease and control cases.

Somatic mutations (bottom row):

(c) *Density plot* – (x-axes for each brain region: VAF ranging from 0.00 to 0.15; y-axis: Density ranging from 0-75) The relative peak density of somatic SNVs (VAF < 0.15) is comparable (at an average density of 75) between all brain regions in FTD-disease case and control which indicated that somatic SNVs had comparable distributions and sharp density curves in all the six DNA samples for VAF < 0.15.

(d) *VAF of each nucleotide across chromosomal positions* – (x-axes for each brain region: chromosomal position of the MAPT gene, y-axis: VAF ranging from 0.00 to 0.25)

The proportion of variations of nucleotides in each chromosomal position of the MAPT gene of FTD-Tau disease and control is comparable across all six the DNA samples for VAF < 0.25.

*A Kernel density plot is used to plot the graph on a continuous interval/time period whereby the values from a selected column are charted into equally binned distributions and the background noise is smoothened out using kernel smoothing. The peaks of a density plot display where the values are concentrated over an interval.

A: Adenine (red), C: Cytosine (green), G: Guanine (blue), T: Thymine (violet); C_T, O, M: FTD-Tau control Temporal lobe, Occipital lobe, Medulla, respectively; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; NP_T, O, M: FTD-Tau disease case Temporal lobe, Occipital lobe, Medulla, respectively; SNVs: Single nucleotide variations; VAF: Variable allele frequency.

These results are also reflected in **Figs. 48 b & d** which show the distributions of the VAF of each nucleotide across each chromosomal position of the MAPT gene [x-axes for each brain region: MAPT chromosomal position (44080000-44100000), y-axis: VAF (0.00-1.00)]. The proportions of variation of mutations detected across the MAPT gene with VAF > 0.25 is much greater in the brain regions of FTD-disease than control (**Fig. 48b**) while the proportions of variation of mutations detected across the MAPT gene with VAF < 0.25 is comparable across all the six DNA samples of FTD-disease and control (**Fig. 48d**).

Again, statistical analysis with the Kolmogorov-Smirnov test revealed that not only was the distribution of somatic SNVs with VAF < 0.01 in FTD-Tau disease compared to control significant (Kolmogorov-Smirnov test, **p = 0.0012), but also that the VAF distribution of germline SNVs in FTD-Tau disease compared to control was very highly significant (Kolmogorov-Smirnov test, ***p = 0.0002) (**Table 34**).

VAF	Type of mutation	D-value	p-value	Significance
FTD disease Temp vs FTD control Temp				
VAF ≥ 0.25	Germline	0.42188	0.1262	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	0.50000	1.0000	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.13503	0.4717	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.11667	0.5609	No significance
FTD disease Temp vs FTD disease Occ				
VAF ≥ 0.25	Germline	0.20037	0.1178	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	0.50000	0.9333	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.24502	0.01300	*
VAF < 0.01	Somatic (ultra-low VAF)	0.09542	0.70570	No significance
FTD disease temp vs FTD disease Med				
VAF ≥ 0.25	Germline	0.13327	0.5415	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	0.50000	1.0000	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.18580	0.1404	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.11337	0.5206	No significance
FTD disease Occ vs FTD disease Med				
VAF ≥ 0.01 and VAF < 0.01	Somatic	0.10067	0.7707	No significance
FTD disease (total) vs FTD control (total)				
VAF ≥ 0.25	Germline	0.44500	0.0002	***
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	0.62500	0.1259	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.07073	0.6033	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.16791	0.0012	**

FTD disease Temp vs FTD control Temp				
MAPT-A				
VAF ≥ 0.25	Germline	N/A	N/A	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	N/A	N/A	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	1.00000	1.0000	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.37500	0.9778	No significance
MAPT-B				
VAF ≥ 0.25	Germline	0.51515	0.2085	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	1.00000	1.0000	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.47368	0.0327	*
VAF < 0.01	Somatic (ultra-low VAF)	0.44839	0.0411	*
MAPT-C				
VAF ≥ 0.25	Germline	0.38889	0.5455	No significance
VAF ≥ 0.05 and VAF < 0.25	Somatic (very high VAF)	N/A	N/A	No significance
VAF ≥ 0.01 and VAF < 0.05	Somatic	0.20078	0.2252	No significance
VAF < 0.01	Somatic (ultra-low VAF)	0.21003	0.3098	No significance

Table 34: SuperFi-mediated amplicon-based DS – Kolmogorov-Smirnov test to calculate significance of mutational burden in FTD-disease *vs* control.

Note: ***Kolmogorov-Smirnov test*** *is a non-parametric goodness-of-fit test that measures the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (the one-sample case), or to compare two samples (the two-sample case). The Kolmogorov–Smirnov statistic quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples, given as the* ***D-value****. The null distribution is calculated under the null hypothesis that the sample is drawn from the reference distribution (in the one-sample case) or that the samples are drawn from the same distribution (in the two-sample case).* ***Statistical significance*** *is a determined by calculating the* ***p-value*** *to reject the null hypothesis, which is computed from the maximum distance between the cumulative frequency distributions (D-value), accounting for sample size in the two groups. The most common threshold to determine statistical significance is* $p < 0.05$.

Comparing between the two brain regions of FTD-Tau disease, temporal lobe *versus* occipital lobe, the distribution of somatic mutations (VAF ≥ 0.01 and VAF < 0.05) had slight significance (Kolmogorov-Smirnov test, *p = 0.013), indicating that the temporal lobe of the FTD-Tau patient may be susceptible to a higher mutation accumulation which could be associated with FTD pathogenesis.

Comparing the temporal lobe samples of FTD-Tau disease *versus* control, the VAF distribution within exons 9-10 of the MAPT gene (MAPT-B amplicon) also showed statistical significance in somatic SNVs (VAF ≥ 0.01 and VAF < 0.05) (Kolmogorov-Smirnov test, *p = 0.0327) and ultra-low frequency somatic SNVs (VAF < 0.01) (Kolmogorov-Smirnov test, *p = 0.0411), indicating that this region had a significantly higher mutation rate which corresponds to the genomic region within the MAPT gene

with a higher incidence of established pathogenic mutations as per literature (Ensembl database).

7.2.1.5 Summary of results for SuperFi-mediated amplicon-based DS

Thus far, this chapter describes the preparation of all the three MAPT amplicons for each of the brain regions Temp, Occ, and Med for FTD disease and control cases using SuperFi DNA polymerase in the pre-library amplification step resulting in eighteen amplicons. The three MAPT amplicons for each brain region were then mixed in equimolar proportions as the template for the KAPA-mediated duplex library preparations, resulting in a final of six amplified duplex libraries made from 4 amol, 20 cycles of adapter-ligated DNA. The libraries were prepared as desired when checked on the HS D1000 and HS D5000 ScreenTape results for adapter ligation and library amplification stages as verifications for duplex library constructions. These libraries were then submitted for sequencing on Illumina NovaSeq platform.

The sequencing run was successful as verified by the family size calculations – all the duplex libraries had an average family size of ~3-6 as recommended by (Kennedy et al. 2014); however, the number of raw reads for 1 DCS for each sample was much higher than the recommended number of 50 reads, indicating inefficient DS. The excellent DCS depth of sequencing obtained at ~1000X for all six samples also indicated successful duplex library preparation.

A greater number of somatic mutations (VAF < 0.25) were detected in all the six DNA samples; however, the number of germline SNVs (VAF > 0.25) was higher in all the three brain regions of FTD-disease with a comparable distribution of mutations to the somatic SNVs than the control, whose somatic SNVs (VAF <0.25) had higher numbers and higher density curves than the germline SNVs (VAF > 0.25) in all its three brain regions. The numbers and VAF distributions of germline SNVs (VAF > 0.25) and ultralow frequency somatic SNVs (VAF < 0.01) was found to be statistically significant when comparing the total mutational burdens of FTD-disease versus control. When comparing two lobes of the FTD-disease patient, the mutational burden detected in temporal lobe was significantly higher than the occipital lobe, suggesting that an accumulation of mutational burden in the temporal lobe may be associated with FTD pathogenesis. Lastly, a higher number of mutations within exons 9-10 of MAPT gene was also determined to a significant degree, a finding that is consistent with literature that

annotates this region to have the highest number of established pathogenic mutations (Ensembl database).

7.2.2 Double capture-based approach to duplex sequencing (with custom designed IDT probes for whole MAPT gene)

7.2.2.1 Duplex library preparation

500 ng of native DNA was used as the starting material for performing duplex library preparation as described in Chapter 3, section 3.2.6.1. (See *Appendix V* for HS D1000 and HS D5000 ScreenTape results of the duplex library preparation from native DNA before the application of the capture protocol).

7.2.2.2 Probes

A total of 1,696 probes were custom-designed for the '+' and '-' strands for the chromosomal coordinates for MAPT gene (see in Chapter 3, section 3.2.6.2): 845 probes of the total 1,115 designed for the '+' strand and 851 probes of the total 1,116 designed for the '-' strand were classified as "good" probes and were synthesized in the IDT Discovery Pools (see *Appendix VI* for list of "good" probes for both strands, i.e., probes that do not code for any repetitive sequences).

7.2.2.3 Capture 1

The first round of capture included lyophilization, 16-hour incubation, mixing with streptavidin beads, heated and room temperature washes, and post-capture 1 amplification which were done with all the DNA for 16 cycles as described in Chapter 3, sections 3.2.6.3 to 3.2.6.10.

The HS D5000 ScreenTape results showed a single peak within the desired range of 300-400 bp for all the six samples (**Fig. 49**).

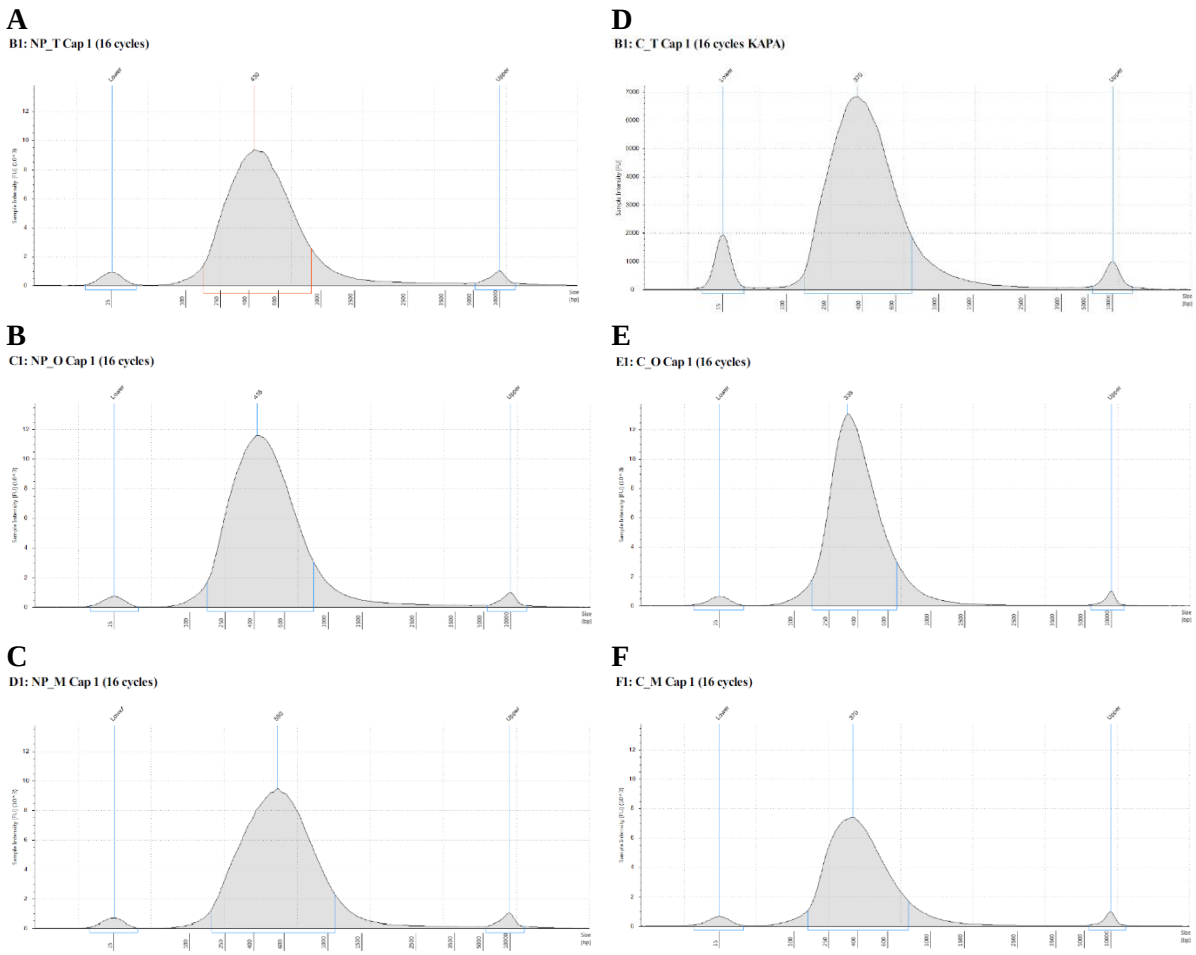


Figure 49: Double capture-based DS: Post-capture I library amplification.
A single, sharp peak within 300-400 bp as desired on HS D5000 ScreenTape after Cap I (all DNA, 16 cycles) for: (a) NP16.163 Temp (b) NP16.163 Occ (c) NP16.163 Med (d) C576 Temp (e) C576 Occ (f) C576 Med Cap I: Capture I; C576: Age-matched control; FTD: Frontotemporal dementia; Med: Medulla NP16.163: FTD-Tau disease case; Occ: Occipital lobe; Temp: Temporal lobe.

7.2.2.4 Capture 2

The protocol for the second capture was followed as described in Chapter 3, section 3.2.6.11.

The final amplification post-capture 2 for the samples needed optimization to obtain a single sharp peak at ~300 bp since we had double the number of duplex libraries captured in this experiment. The following conditions were tested on post-capture 2 duplex libraries until only a single peak within desired range were obtained (**Figs. 50-54**).

Different samples needed to be used to test each case as the total volume of DNA-beads was only 20 µl for each FTD sample. We attempted only one test sample per condition, until a positive result. We accepted a condition for replication in all samples only if it gave positive results for at least two samples.

Optimization 1: Our first attempt of optimization used '10 µl for 10 cycles' for library amplification. We used half the input based on family size results from Chapter 5, section 5.2.2.1, and we used the higher end of the recommended cycle number by (Schmitt et al. 2015) – 10 cycles – on NP16.163 Temp (**Fig. 50**). While there was a significant peak within our desired size range, there also was a larger peak containing non-specific higher molecular weight fragments, and this test condition was ruled unusable.

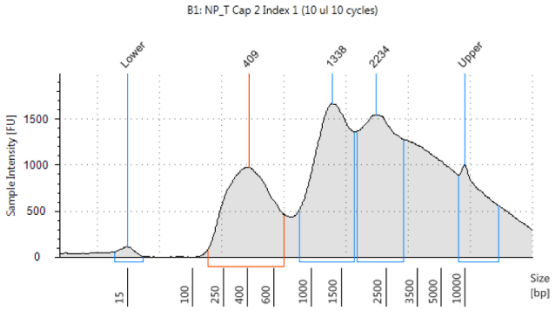


Figure 50: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 1.

Optimization 1: Cap 2 – 10 µl, 10 cycles. Verified on HS D5000 ScreenTape. Sample tested on: NP16.163 Temp – Multiple peaks seen indicating that this test condition is unsuitable for sequencing.
Cap 2: Capture 2; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe.

Optimization 2: Reducing the cycle number by one while maintaining the DNA input quantity in C576 Temp gave a single sharp peak at ~300 bp (**Fig. 51a**); however, higher molecular weight fragments were seen when the same condition was tested on C576 Med (**Fig. 51b**), so this condition was discarded.

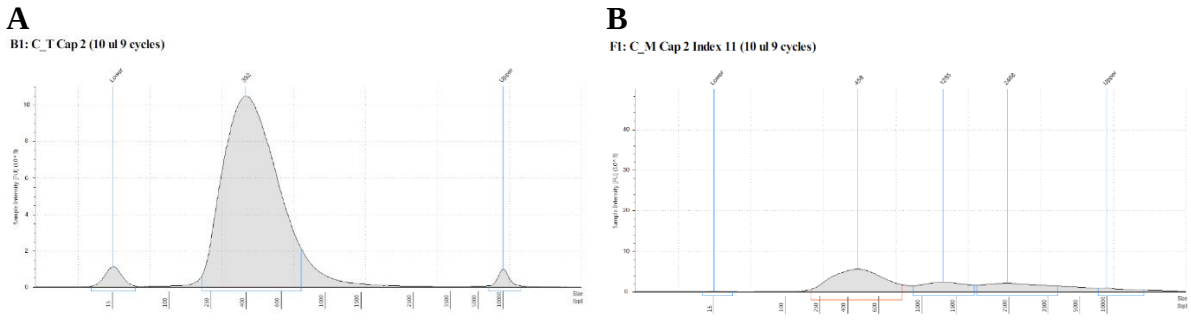


Figure 51: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 2.

Optimization 2: Cap 2 – 10 µl, 9 cycles. Verified on HS D5000 ScreenTape. Sample tested on: (a) C576 Temp – Gave a single sharp peak within desirable range – taken for sequencing; however, (b) C576 Med – Multiple peaks seen indicating that this test condition is ultimately unsuitable for sequencing.
Cap 2: Capture 2; C576: Age-matched control; Med: Medulla; Temp: Temporal lobe.

Optimization 3: Similar peaks comprising of high molecular weight fragments was also seen when the DNA input was reduced by 2 µl and cycle number was further decreased by one to of '8 µl for 8 cycles' on NP16.163 Occ (**Fig. 52**).

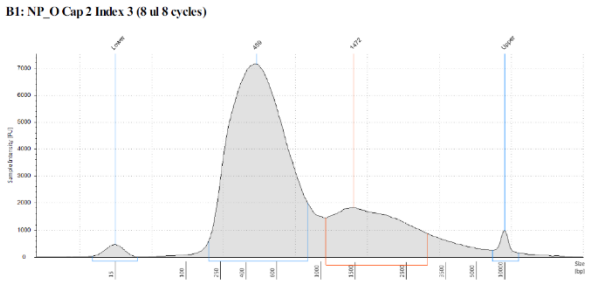


Figure 52: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 3.

Optimization 3: Cap 2 – 8 µl, 8 cycles. Verified on HS D5000 ScreenTape. Sample tested on: NP16.163 Temp – Multiple peaks seen indicating that this test condition is unsuitable for sequencing.
Cap 2: Capture 2; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe.

Optimizations 4 & 5: The conditions '2.5 µl for 10 cycles' and '1 µl for 10 cycles' were tested simultaneously. While the former gave us a single peak as desired for C576 Temp (**Fig. 53a**), the latter condition showed the presence of larger fragments for NP16.163 Temp (**Fig. 53b**), indicating that 10 cycles may be overwhelming the system and we were not confident to adopt the '2.5 µl for 10 cycles' for the remaining samples as well.

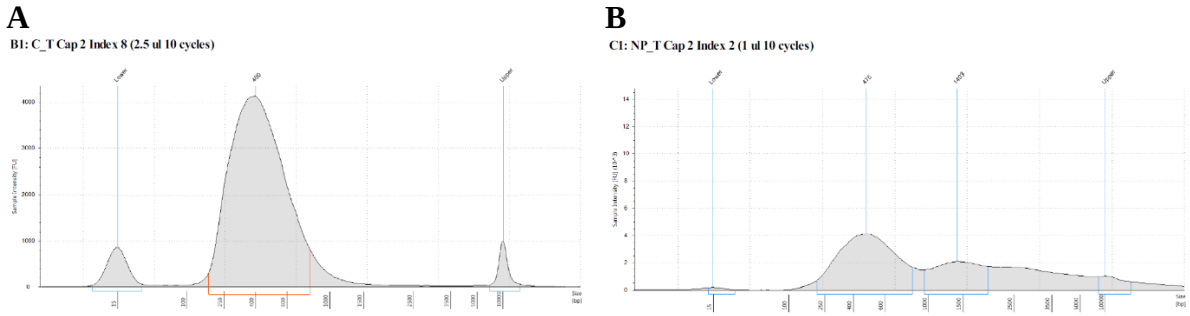


Figure 53: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimizations 4 and 5.

Verified on HS D5000 ScreenTape.
Optimization 4: Cap 2 – 2.5 µl, 10 cycles. Sample tested on: (a) C576 Temp – Gave a single sharp peak within desirable range – taken for sequencing; however,
Optimization 5: Cap 2 – 1 µl, 10 cycles. Sample tested on (b) NP16.163 Temp – Multiple peaks seen indicating that test conditions with cycle number ≥ 10 may be unsuitable for sequencing for remaining samples.
Cap 2: Capture 2; C576: Age-matched control; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Temp: Temporal lobe.

Optimization 6: We tested a final condition of '1 µl for 9 cycles' on two samples of FTD – NP16.163 (**Fig. 54a**) and C576 Occ (**Fig. 54b**) – and the ScreenTape results showed a single peak at ~300-400 bp for both.

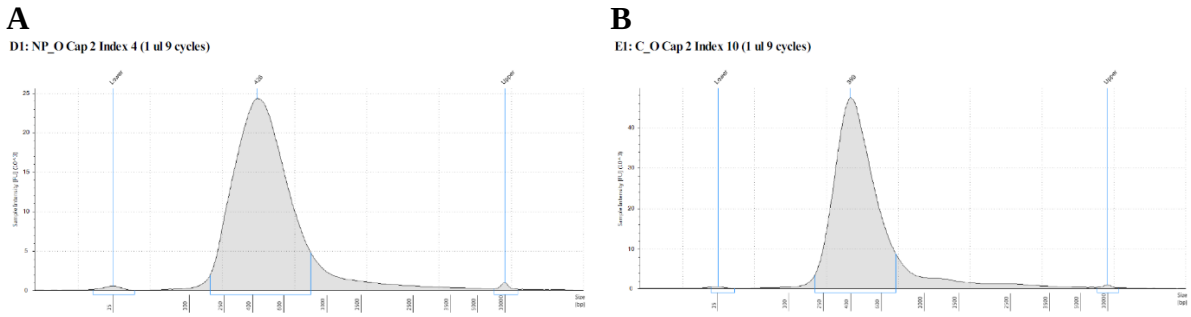
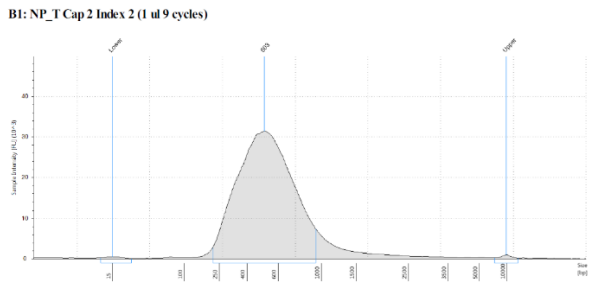


Figure 54: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – post-capture 2 library optimization 6.

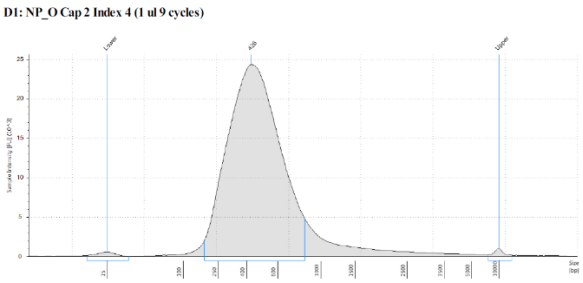
Optimization 6: Cap 2 – 1 µl, 9 cycles. Verified on HS D5000 ScreenTape. Sample tested on: (a) NP16.163 Occ and (b) C576 Occ – Both replicates gave a single sharp peak within desirable range indicating that this test condition is suitable for sequencing.

Cap 2: Capture 2; C576: Age-matched control; FTD: Frontotemporal dementia; NP16.163: FTD-Tau disease case; Occ: Occipital lobe.

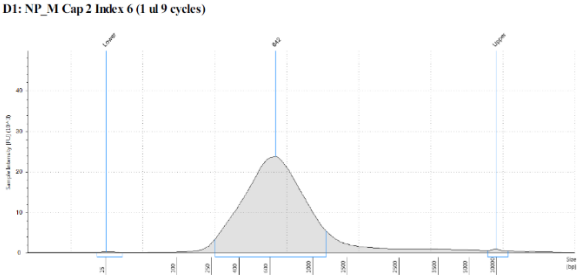
Thus, we replicated this condition for all the remaining FTD samples sent for sequencing on Illumina NovaSeq, except for C576 Temp for which we submitted the '10 µl for 9 cycles' as well as '2.5 µl for 10 cycles' to further understand the relationship between the library amplification parameters and family size (**Fig. 55**).



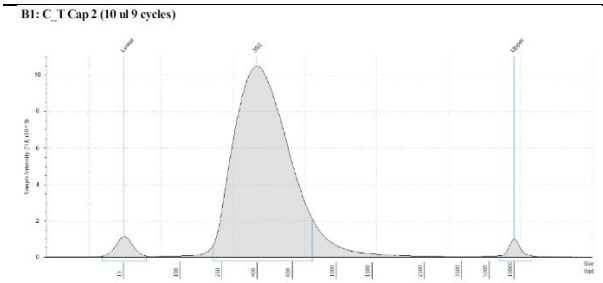
A: NP16.163 Temp Cap 2 (1 µl, 9 cycles)



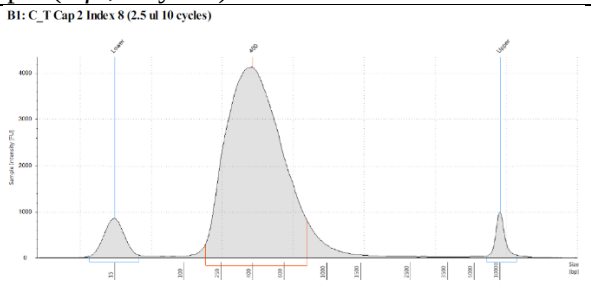
B: NP16.163 Occ Cap 2 (1 µl, 9 cycles)



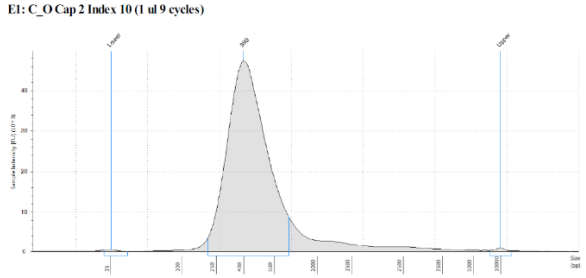
C: NP16.163 Med Cap 2 (1 µl, 9 cycles)



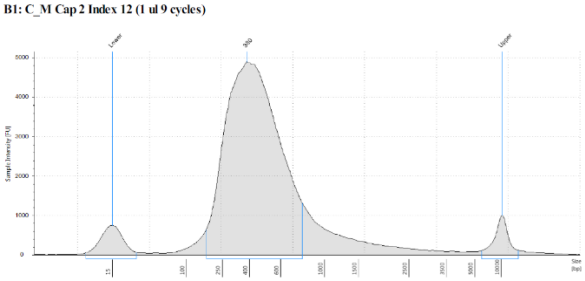
D: C576 Temp Cap 2 (10 µl, 9 cycles)



E: C576 Temp Cap 2 (2.5 µl, 10 cycles)



F: C576 Occ Cap 2 (1 µl, 9 cycles)



G: C576 Med Cap 2 (1 µl, 9 cycles)

Figure 55: HS D5000 results for final amplified duplex libraries for NovaSeq.

All samples gave one single peak within desirable range of ~300-400 bp. Verified on HS D5000 ScreenTape.
FTD Disease samples for NovaSeq: (a) NP16.163 Temp Cap 2 (1 µl, 9 cycles) (b) NP16.163 Occ Cap 2 (1 µl, 9 cycles) (c) NP16.163 Med Cap 2 (1 µl, 9 cycles); and,
Control samples for NovaSeq: (d) C576 Temp Cap 2 (10 µl, 9 cycles) (e) C576 Temp Cap 2 (2.5 µl, 10 cycles) (f) C576 Occ Cap 2 (1 µl, 9 cycles) (g) C576 Med Cap 2 (1 µl, 9 cycles).
Cap 2: Capture 2; C576: Age-matched control; FTD: Frontotemporal dementia; Med: Medulla; NP16.163: FTD-Tau disease case; Occ: Occipital lobe; Temp: Temporal lobe.

7.2.3 Computational sequencing data analyses

Most of the bioinformatic analyses described hereafter were performed by Dr Yu (Nell) Nie from Prof Patrick Chinnery's lab at the University of Cambridge, UK; the mutational burden and frequency calculations done in '**5) Mutational burden**' were performed by the

PhD candidate under the guidance of Dr Claudia Calabrese from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

1) Family Size

Again, the first analysis done for duplex sequencing was to determine the family size for each sample. An average peak family size of 6 (reads with duplex consensus, i.e., peak family size with 6 DCS) with around 40-50 raw reads required to make one DCS is optimal for duplex sequencing efficiency (Kennedy et al. 2014).

Fig. 56 [x-axis: family size, y-axis: proportion of total reads] shows that the peak family size as well as the number of raw reads to make a DCS were well above the optimal levels.

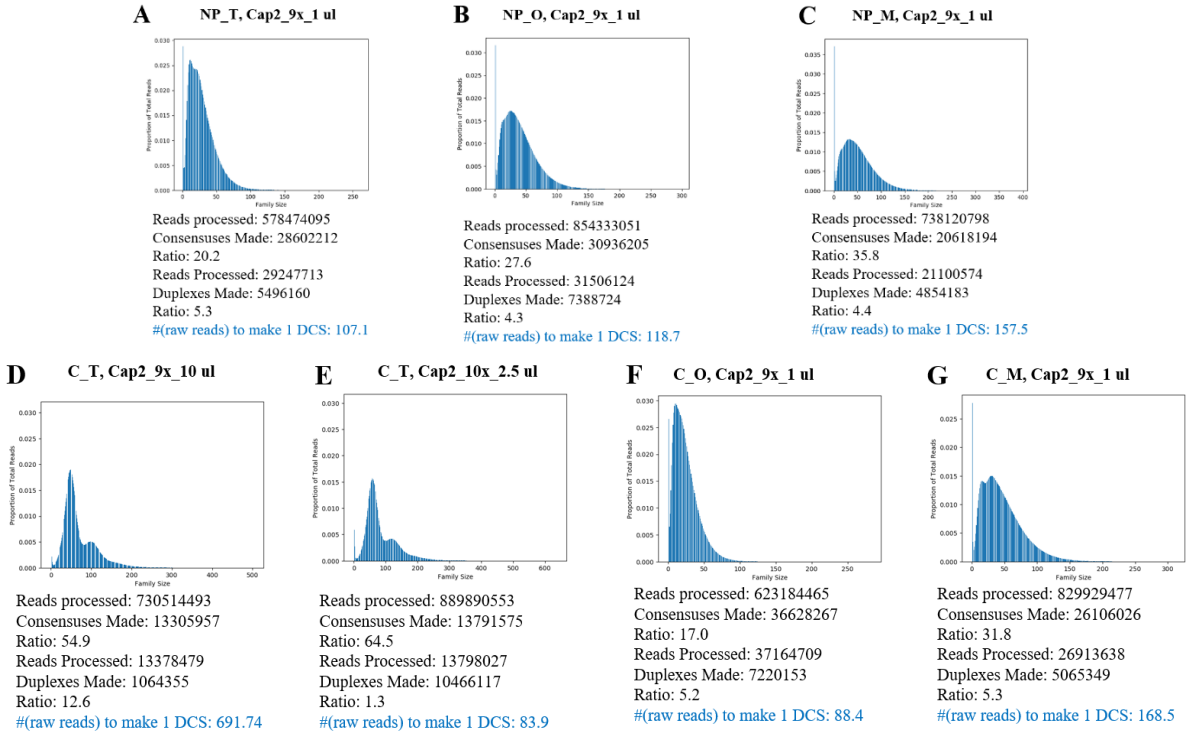


Figure 56: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Family size calculations.

FTD Disease samples: (a) NP16.163 Temp Cap 2 (1 µl, 9 cycles) (b) NP16.163 Occ Cap 2 (1 µl, 9 cycles) (c) NP16.163 Med Cap 2 (1 µl, 9 cycles); and, *FTD Control samples*: (d) C576 Temp Cap 2 (10 µl, 9 cycles) (e) C576 Temp Cap 2 (2.5 µl, 10 cycles) (f) C576 Occ Cap 2 (1 µl, 9 cycles) (g) C576 Med Cap 2 (1 µl, 9 cycles).
All the 'Cap 2: 1 µl, 9 cycles' samples gave a good peak family size with ~100 raw reads to make a DCS which indicated successful library preparation. (Optimal: peak family size = 6 with ~ 40-50 raw reads to make a DCS). The 'Cap 2: 2.5 µl, 10 cycles' sample gave comparable peak family size and raw read numbers and indicated successful library preparation as well. However, the 'Cap 2: 10 µl, 9 cycles' a very high peak family size and ~700 raw reads to make a DCS which indicated inefficient duplex library preparation and was discarded from further analyses.
C576: Age-matched control; C_M: C576 Med; C_O: C576 Occ; C_T: C576 Temp; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; Med: Medulla; NP16.163: FTD-Tau disease case; NP_M: NP16.163 Temp; NP_O: NP16.163 Occ, NP_T: NP16.163 Temp; Occ: Occipital lobe; Temp: Temporal lobe.

For the three FTD disease samples – Temp, Occ, and Med – (Capture 2: 1 µl DNA input for 9 cycles) (**Figs. 56a-c**), the peak family size was found to be 25-30 and the number of raw reads present to make a DCS was above 100 for all three regions, indicating that the duplex sequencing library preparation was a success for these samples. The greater number of raw reads to make a DCS than optimum indicates that the library amplification parameters, while successful, were sub-optimal.

Figs. 56d & e show the family size results for FTD Control Temp which had two conditions tested: (1) Capture 2: 10 µl DNA input for 9 cycles (**Fig. 56d**) – the peak family size was above 50 and had 700 raw reads to make a DCS, indicating the duplex library was highly inefficient and this sample was discarded for future analyses; and (2) Capture 2: 2.5 µl DNA input for 10 cycles (**Fig. 56e**) – the peak family size was close to 50 and had ~85 raw reads to make a DCS, indicating a slightly more favourable duplex library preparation, and was used for further analyses along with the other FTD samples.

The FTD Control Occ (**Fig. 56f**) and Med (**Fig. 56g**) using 1 µl DNA input for 9 cycles for the second capture also had a peak family size of 25-30 and the number of raw reads present to make a DCS was above 100 for all three regions, indicating that the duplex sequencing library preparation was a success for these samples as well.

2) Check for the "spacer" sequence

The first quality check done on the DCS reads was to eliminate all the reads that did not contain the complete spacer sequence "TGACT" that is a fixed sequence for all successful duplex adapters in the system (**Fig. 57a**). Any read without this spacer sequence or with an incomplete spacer sequence were eliminated from further analyses.

From **Fig. 57b**, it is evident that only about 1/10th of all the reads that had issues within this spacer sequence were eliminated, which showed that our duplex libraries were constructed correctly.

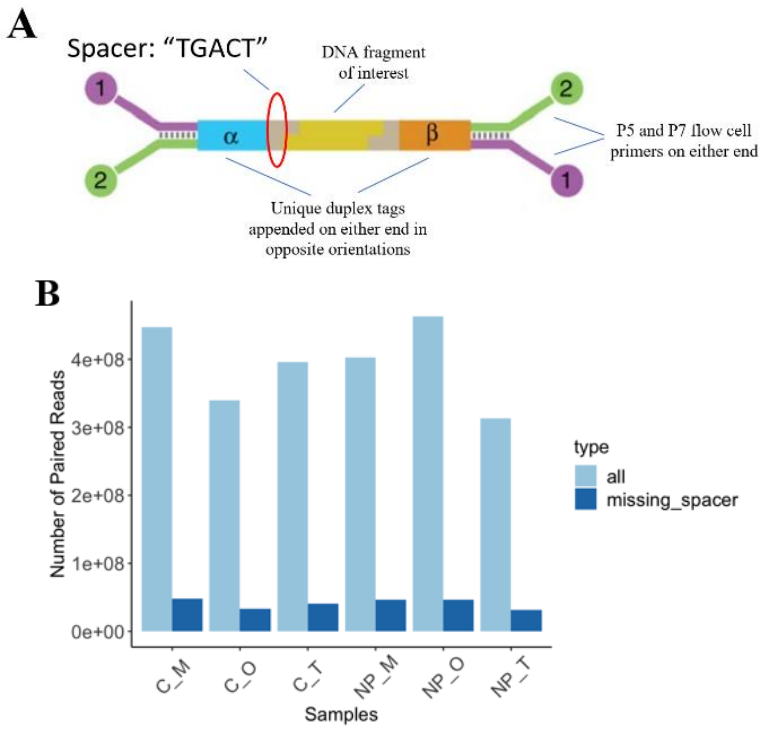


Figure 57: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Check for the "spacer" sequence.

The complete spacer sequence "TGACT" present in a read indicated successful duplex library preparation and any reads with missing/incomplete spacer sequence in our six FTD-Tau samples were eliminated from our data-set.

A: Adenine; C576: Age-matched control; C_M: C576 Med; C_O: C576 Occ; C_T: C576 Temp; C: Cytosine; DNA: Deoxyribonucleic acid; FTD: Frontotemporal dementia; G: Guanine; MAPT: Microtubule-associated protein tau; Med: Medulla; NP16.163: FTD-Tau disease case; NP_M: NP16.163 Temp; NP_O: NP16.163 Occ, NP_T: NP16.163 Temp; Occ: Occipital lobe; Temp: Temporal lobe; T: Thymine.

3) On-target ratio

The next analysis for the six FTD samples is to look at the on-target ratio, i.e., the number of raw reads found within the region of interest i.e., the MAPT gene on chromosome 17.

Fig. 58 is the calculation of the on-target ratio of a double-capture sample [x-axis: FTD samples, y-axis: number of reads] and the high number of reads corresponding to the MAPT gene on the reference genome shows that there is a highly significant enrichment of our regions of interest, indicating that this double capture-based duplex library preparation was successful.

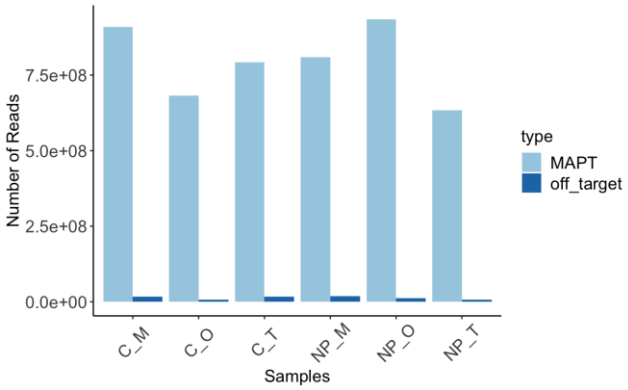


Figure 58: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – On-target ratio calculations.

The double capture protocol with custom-designed IDT probes for the whole MAPT gene was successful in capturing duplex libraries with our targeted MAPT gene sequences present on chromosome 17 for all six FTD-Tau samples.

C576: Age-matched control; C_M: C576 Med; C_O: C576 Occ; C_T: C576 Temp; FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; Med: Medulla; NP16.163: FTD-Tau disease case; NP_M: NP16.163 Temp; NP_O: NP16.163 Occ, NP_T: NP16.163 Temp; Occ: Occipital lobe; Temp: Temporal lobe.

4) Depth of Sequencing

Another parameter calculated to test the quality of the reads was to measure the depth of sequencing, i.e., number of times that a given nucleotide in the genome has been read while sequencing so that base calls may be made with a higher degree of confidence with higher depth.

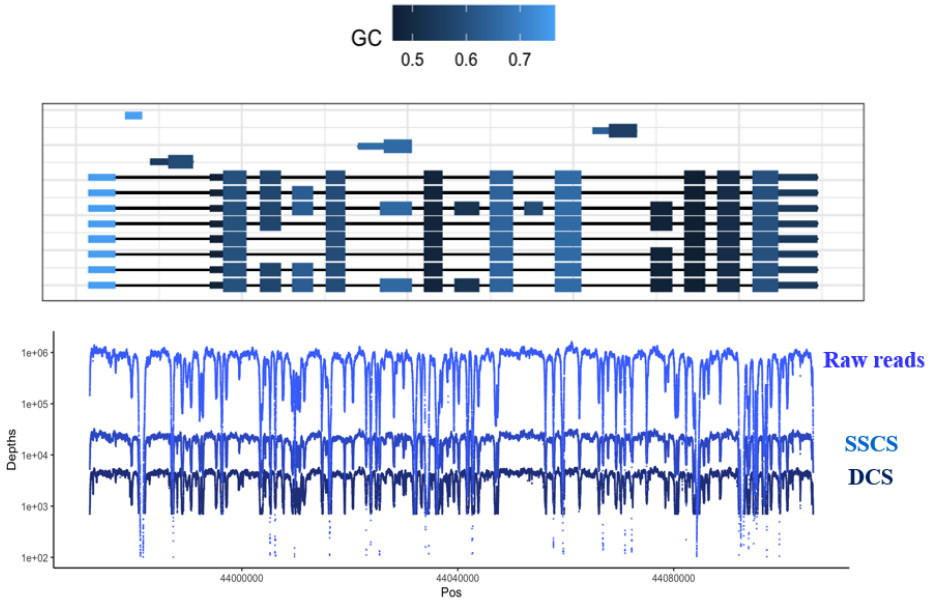


Figure 59: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Depth of sequencing.

A higher depth of sequencing increases the confidence of any base-call variants made. The raw reads had an excellent depth of ~1.0 x 10^6 X across the whole MAPT gene which underwent single-strand consensus and duplex consensus to achieve a high depth of ~1,500X across the whole MAPT gene indicating a highly successful duplex sequencing run.

DCS: Double consensus sequences; GC: Guanine-cytosine content ratio; MAPT: Microtubule-associated protein tau gene; SSCS: Single-strand consensus sequences.

To this end, a graph of the depth of reads made with raw reads, single consensus and duplex consensus corresponding to exons 1-12 of the MAPT gene was generated (**Fig. 59**) [x-axis: chromosomal position, y-axis: depth]. The raw reads had an average depth of ~ $1x10^6$ X across the entire MAPT gene which was much higher than the average depth of single consensus at above 10,000X while the final depth with duplex consensus was excellent with the average depth being higher than the cut-off of 1000X, all of which indicated that our duplex libraries are of very good quality for further analyses.

5) Mutational burden

The mutational burden and frequency calculations done in in this sub-section were performed by the PhD candidate under the guidance of Dr Claudia Calabrese from Prof Patrick Chinnery's lab at the University of Cambridge, UK.

- ***Number and rates of mutations detected:***

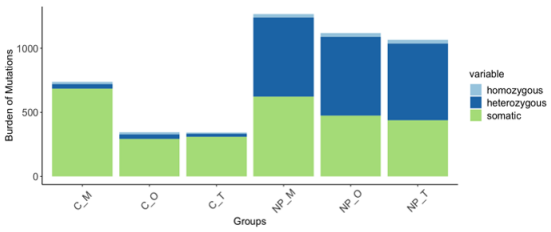
The VAF of the true mutations detected after duplex consensus and quality check were determined for each of the six DNA samples and the number of mutations detected as somatic SNVs (VAF < 0.25), germline heterozygous (VAF > 0.25 and VAF < 1.00), and germline homozygous (VAF ~ 1.00) were tabulated in (**Fig. 60a**) and compared in a bar graph (**Fig. 60b**) [x-axis: FTD samples, y-axis: number of mutations], the total per-bp mutational rate per sample was tabulated in **Fig. 60c**, and finally, the distribution of these mutations were represented in density plots (**Fig. 60d**) [x-axes for each brain region: VAF (0.00 to 1.00), y-axis: density].

Overall, the total mutational burden in each brain region of temporal (1066 SNVs) and occipital (1118 SNVs) lobes and the medulla (1268 SNVs) of the NP16.163 FTD disease case was over a thousand SNVs whereas the total mutational burden in each brain region of FTD control was ~ 340 SNVs in the temporal and occipital lobes and 739 SNVs in the medulla, indicating that the disease state was associated with higher mutational accumulation over time (**Figs. 60 a & b**). Surprisingly, the number of mutations was slightly higher in the Med samples of both, FTD-tau disease case and control.

A

Sample	No. of somatic mutations	No. of germline mutations			Total mutational burden
		Heterozygous	Homozygous	Total	
C576 Med	685	28	17	54	739
C576 Occ	293	29	17	52	345
C576 Temp	309	28	9	34	343
NP16.163 Med	623	617	37	645	1268
NP16.163 Occ	474	615	35	644	1118
NP16.163 Temp	439	599	25	627	1066

B



C

Sample	Mutation rate per sample
C576 Med	1.11×10^{-2}
C576 Occ	1.95×10^{-2}
C576 Temp	1.38×10^{-2}
NP16.163 Med	6.08×10^{-2}
NP16.163 Occ	6.82×10^{-2}
NP16.163 Temp	6.96×10^{-2}

D

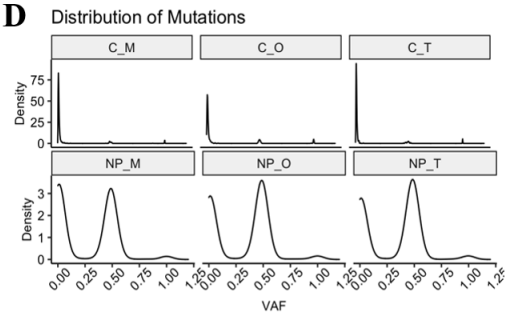


Figure 60: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Mutational burden detection.

Mutational burden detection in the MAPT gene for the three FTD-Tau samples and age-matched controls:
(a) *Mutational count:* Table describing the number of mutations detected with VAF < 0.25 (somatic), VAF > 0.25 (germline heterozygous), VAF ~ 1.00 (germline homozygous), and total mutational burden (somatic + total germline mutations) in each sample.
(b) *Graphical representation*: [x-axis: brain samples, y-axis: burden of mutations (0 to1000)] Germline homozygous SNVs with VAF ~ 1.00 (light blue) were constant across all the six samples as expected, germline heterozygous SNVs with VAF > 0.25 (dark blue) were much higher in the FTD-disease samples than age-matched control, and the somatic mutational burden was slightly higher in the FTD-disease samples than in control. Also, the number of somatic SNVs was higher in number than the germline SNVs within the patient as well as the control. These results indicate that somatic and germline heterozygous SNVs may be associated with FTD pathology in the patient.
(c) *Mutation rates:* Table describing the total per-bp mutational rate per sample where the overall rate of mutations in control was lower than that of FTD-Tau disease case. The overall per-bp mutation rate in the control increased from Med < Temp < Occ while in the disease, it increased from Med < Occ < Temp, with the temporal lobe having the highest per-bp mutation rate of all brain regions, suggesting that this region may be more susceptible to mutation accumulation in FTD-Tau pathology.
(d) *Density plot*: [x-axes for each brain region: VAF (0.00 to 1.00), y-axis: density (0-75 for control brain regions, 0-3 for diseased brain regions)] The relative peak density of somatic SNVs (VAF < 0.25) is much higher (at an average density of 60) than that of germline heterozygous (VAF > 0.25) (at an average density of 3-4) and germline homozygous (VAF ~ 1.00) (at an average density of 1-2) across each brain region of the age-matched control indicating that the somatic SNVs had a much greater difference in the distribution and a sharper density curve than that of germline SNVs in this case (**top row**) while the relative peak density of somatic SNVs (VAF < 0.25) was comparable (at an average density of 3) with that of germline heterozygous SNVs (VAF > 0.25) (at an average density of 3), but greater than that of germline homozygous SNVs (VAF ~ 1.00) (at an average density of 0.5) across each brain region of the FTD-Tau disease case, indicating that the former two mutation types had comparable distributions and density curves (**bottom row**). These results suggest that somatic and germline heterozygous SNVs may be associated with FTD pathology in the patient.
*A Kernel density plot is used to plot the graph on a continuous interval/time period whereby the values from a selected column are charted into equally binned distributions and the background noise is smoothened out using kernel smoothing. The peaks of a density plot display where the values are concentrated over an interval.
C576: Age-matched control; C_M: C576 Med; C_O: C576 Occ; C_T: C576 Temp; FTD: Frontotemporal dementia; Med: Medulla; NP16.163: FTD-Tau disease case; NP_M: NP16.163 Temp; NP_O: NP16.163 Occ; NP_T: NP16.163 Temp; Occ: Occipital lobe; SNVs: Single nucleotide variations; Temp: Temporal lobe; VAF:

However, the overall point mutation frequency per sample, which was calculated as the sum of all point mutations (A>T + A>C + A>G + T>A + T>C +

T>G + C>A + C>T + C>G + C>A + C>T + G>C) per sample divided by the total number of nucleotides sequenced (total number of A's sequenced + total number of C' sequenced + total number of G's sequenced + total number of T's sequenced) per sample (*Appendix VIIA*), was greater in the FTD-tau disease case compared to control, with the highest rates of mutation present in the temporal lobe of the patient (per-bp mutation rate = 6.96×10^{-2}), suggesting that this specific brain region may be slightly more susceptible to accumulating mutations over time (**Fig. 60c**).

Germline homozygous mutations play no significant role as evidenced by their low numbers in disease and control; however, germline heterozygous and somatic SNVs that are pathogenic could play a major role in FTD pathogenesis as seen in their comparable distributions in FTD-tau disease *vs* the sharper and higher relative peak density of somatic SNVs than germline SNVs in control (**Fig. 60d)**.

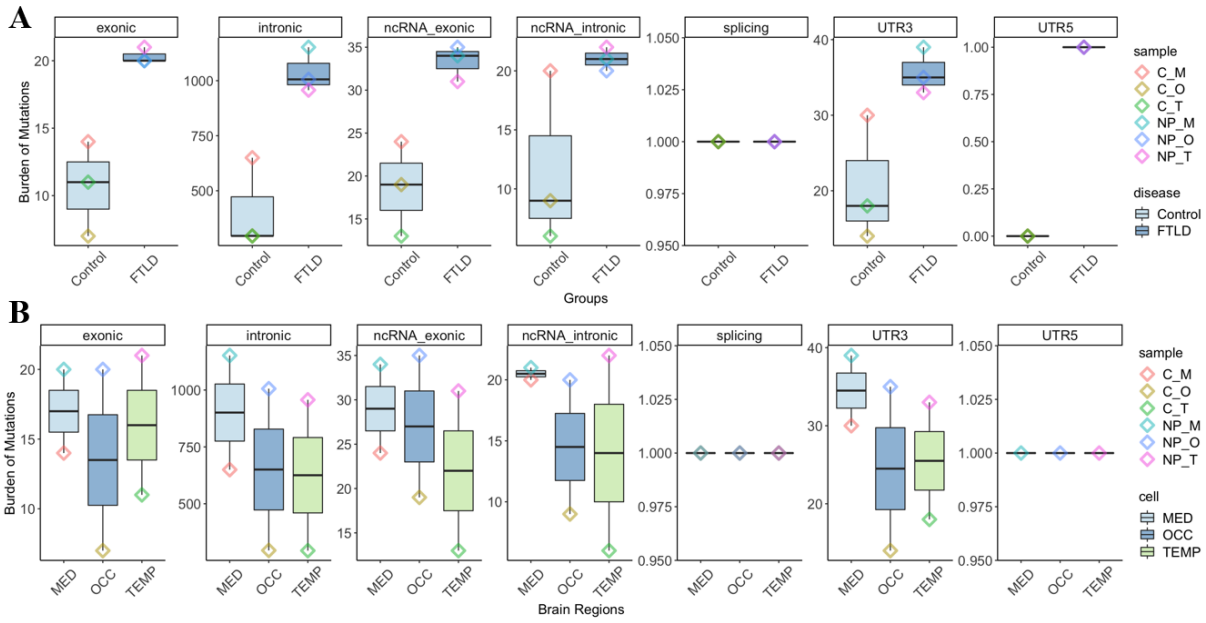
- ***Number and rates of somatic mutations detected in each genomic region of MAPT:***

Since the strength of DS is the ability to detect lower frequency mutations more accurately, the samples were analysed to look at the somatic mutational burden across different genomic regions in all the brain regions of the FTD-tau disease case and control (**Fig. 61**).

When the somatic mutational burden was explored within different genomic regions with respect to the control *versus* the patient (**Fig. 61a**) [x-axis: FTD control vs disease, y-axis: number of mutations], the number of somatic SNVs detected in the control was lower than that detected in the FTD disease patient across all genomic regions and the brain regions of the control showcased stretched mutation distributions compared to the squeezed distributions of the patient: in the *exonic region* (***column 1***, y-axis range: 10-20), the distribution of mutations was slightly negatively skewed in the control [median = 11 SNVs] while that of the FTD-Tau disease case was slightly positively skewed [median = 20 SNVs]; in the *intronic region* (***column 2***, y-axis range: 500-1000), the distribution of mutations was positively skewed in the control [median = 263 SNVs] as well as that of the FTD-Tau disease case [median = 412 SNVs]; in the *ncRNA exonic region* (***column 3***, y-axis range: 15-35), the distribution of mutations was only slightly negatively

skewed in the control [median = 19] while that of the FTD-Tau disease case was negatively skewed [median = 15]; in the *ncRNA intronic region* (***column 4***, y-axis range: 10-20), the distribution of mutations was positively skewed in the control [median = 9] while the FTD-Tau disease case had a normal distribution [median = 18]; in the *splicing region* (***column 5***, y-axis range: 0.950-1.050), the mutations had zero skewness in the control as well as the FTD-Tau disease case [median = 1 SNV for both cases]; in the *UTR3 region* (***column 6***, y-axis range: 20-40), the distribution of mutations was positively skewed in the control [median =16 SNVs] as well as that of the FTD-Tau disease case [median = 35 SNVs]; and, in the *UTR5 region* (***column 7***, y-axis range: 0.00-1.00), the distribution of mutations had zero skewness in the control [median = 0] as well as in the FTD-Tau disease case [median = 1 SNV]. This set of graphs indicated that the somatic mutational burden in the FTD-Tau disease case was greater than control for nearly all the genomic regions, except for splicing and UTR5 regions, with the highest mutational burdens present in the intronic region of the MAPT gene, followed by exonic, ncRNA, UTRs, and splicing regions, which suggests that genomic regions other than those identified to be directly involved in transcription, translation, and their regulation of the tau gene in brain cells may also be contributing to FTD pathology.

When comparing the total somatic mutational burden within different genomic regions of each brain region, there is a normal distribution of mutations within each brain region across FTD control and disease, where the median is equal to the mean of the number of SNVs of DTD-tau disease case *versus* the control in each brain region (**Fig. 61b**) [x-axis: Brain regions, y-axis: number of mutations]. This set of graphs reflect those in **Fig. 61a**, but provide additional information in that, in general, the medulla (the internal brain control within the patient as well as the control) had a higher number of mutations across all the genomic regions of the patient and the control than the temporal and occipital lobes which had more comparable numbers of mutations with similar distributions within each genomic region of the patient as well as the control which may be indicative of early developmental SNVs in both the brain samples; however, unlike other genomic regions, the number somatic SNVs is higher in the temporal lobe than in the occipital lobe and the medulla of the exonic region of the FTD-tau patient, which



C

Genomic region	C576 Med	C576 Occ	C576 Temp	NP16.163 Med	NP16.163 Occ	NP16.163 Temp
Exonic	1.14×10^{-3}	2.48×10^{-3}	1.93×10^{-3}	5.44×10^{-3}	6.33×10^{-3}	6.41×10^{-3}
Intronic	4.69×10^{-2}	9.25×10^{-2}	6.29×10^{-2}	2.27×10^{-1}	2.53×10^{-1}	2.58×10^{-1}
ncRNA exonic	4.37×10^{-4}	5.69×10^{-4}	3.16×10^{-4}	1.19×10^{-4}	1.99×10^{-4}	1.54×10^{-4}
ncRNA intronic	2.94×10^{-4}	1.19×10^{-4}	1.32×10^{-4}	1.15×10^{-4}	1.45×10^{-4}	1.60×10^{-4}
Splicing	2.01×10^{-5}	1.19×10^{-5}	2.46×10^{-5}	1.17×10^{-5}	1.35×10^{-5}	1.44×10^{-5}
UTR3	3.00×10^{-3}	6.27×10^{-3}	5.06×10^{-3}	9.52×10^{-3}	1.07×10^{-2}	1.08×10^{-2}
UTR5	-	-	-	5.84×10^{-4}	6.66×10^{-4}	6.55×10^{-4}
	Somatic rate of mutation per genomic region per sample					

Figure 61: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Somatic mutational burden within different genomic regions.

Somatic mutational burden detection within different regions of the MAPT gene for the three brain regions of FTD-Tau patient and the age-matched control:

(a) *Controls vs disease cases*: The somatic mutational burden in the brain regions of the control is lower and the data distribution more dispersed than that of the FTD-tau disease patient across all the genomic regions, except in the genomic regions of splicing and UTR5. Somatic mutational burden was the highest in the intronic region, followed by ncRNA, exonic, the UTRs, and splicing regions respectively, suggesting that the role of somatic SNVs in ageing and neurodegeneration may not be limited to alternative splicing, and possibly, that FTD pathogenesis is perhaps associated with processes other than mutant protein accumulation in brain cells.

(b) *Med vs Occ vs Temp*: Early developmental SNVs may be associated with both brain samples as the medulla has the highest somatic mutational burden with squeezed distributions in all genomic regions while the occipital and temporal lobes show stretched distributions in general, the exceptions being splicing and UTR5 regions. However, the somatic mutational burden in the exonic regions of the temporal lobe is especially higher than the occipital lobe and medulla, which might indicate that a slight increase in somatic mutational burden within the exonic region could be associated with the disease state.

(c) *Rate of somatic mutations*: Table describing the total per-bp mutational rate per genomic region per sample. The per-bp mutation rate was the lowest in the splicing region for all the brain regions while the intronic regions had the highest rates, with mutational rates generally being lower in the brain regions of the control than the FTD patient (lowest per-bp mutation rate: 1.19×10^{-5} in C576 Occ, splicing region; highest per-bp mutation rate: 2.58×10^{-1} in NP.163 Temp, intronic region; no SNVs were detected in the UTR5 region of the C576 brain regions). Of all the brain regions, the temporal lobe of the FTD patient had the highest per-bp mutation rate in every gene region which could suggest that this brain region may be more susceptible to mutation accumulation in FTD-Tau pathology.

C576: Age-matched control; C_M: C576 Med; C_O: C576 Occ; C_T: C576 Temp; FTLD/FTD/NP16.163: Frontotemporal dementia disease case; Med: Medulla; NP_M: NP16.163 Temp; NP_O: NP16.163 Occ; NP_T: NP16.163 Temp; ncRNA: non-coding RNA; Occ: Occipital lobe; RNA: Ribonucleic acid; Temp: Temporal lobe; UTRs: Untranslated regions.

suggests that a higher number of somatic mutations in the exonic region of the temporal lobe specifically may be associated with FTD pathology in the patient.

Fig. 61c tabulates the per-bp mutation rate of these SNVs detected per genomic region per sample. The overall point mutation frequency per genomic region per sample was calculated as the sum of all point mutations (A>T + A>C + A>G + T>A + T>C + T>G + C>A + C>T + C>G + C>A + C>T + G>C) per genomic region per sample divided by the total number of nucleotides sequenced (total number of A's sequenced + total number of C' sequenced + total number of G's sequenced + total number of T's sequenced) per sample (*Appendix VIIB*). The highest per-bp mutation rates were found in the intronic regions of the gene for each brain sample, which were followed by UTR3, exonic, ncRNA exonic, ncRNA intronic, and UTR5 regions, with the lowest rates being in the splicing regions of the MAPT gene for each brain sample. In general, the mutation rates of the control brain regions were lower than those of the FTD-Tau patient. Interestingly, the per-bp mutation rates in the splicing and both ncRNA regions were higher in the control brain regions than that of the patient, which could either be indicative of these SNVs being a part of an ageing signature or of individual-specific variations seen in the age-matched control. On the other hand, the per-bp mutation rates in the intronic, exonic, and both UTR regions are higher in the FTD-Tau patient than the control which could indicate that alterations in these genomic regions could be contributing to FTD pathology in the patient. While comparable, the per-bp mutation rates were higher in the temporal lobe of the patient than the occipital lobe and the medulla of the patient, which could suggest that this brain region may be susceptible to the accumulation of mutations that may be contributing to FTD pathology in the patient.

6) Somatic mutational frequency

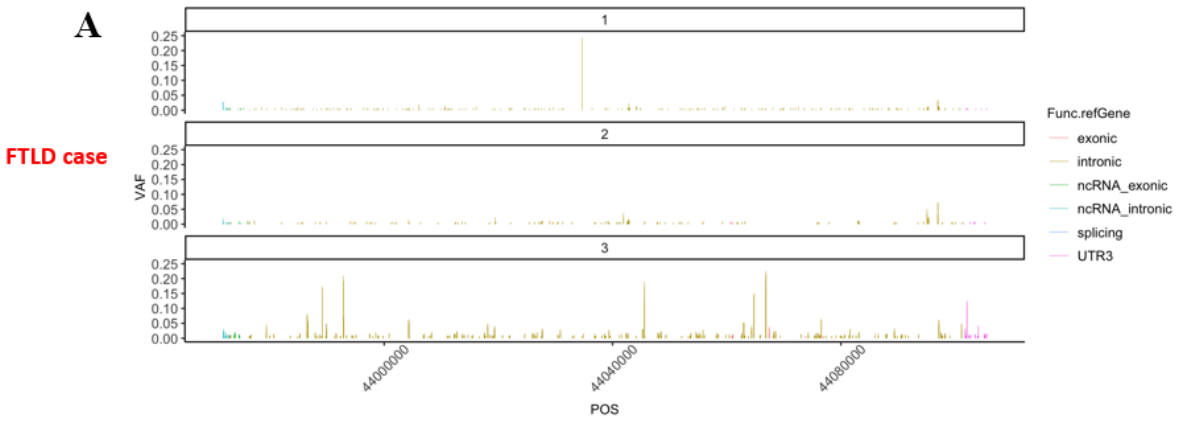
The frequency of somatic mutations (VAF < 0.25) detected were then analysed to determine if they were in 'only one' brain region (single-focal), in 'any two' brain regions, or in 'all three' brain regions (multi-focal) (**Fig. 62**).

Fig. 62a [x-axis: chromosomal position, y-axis: VAF] shows the VAF each nucleotide across the MAPT chromosome that is present in either 'only one' brain region (top row), in 'any two' brain regions (middle row), or in 'all three' brain regions

(bottom row) of the **FTD-Tau disease patient**. The VAF of the SNVs increased from 'only one' < 'any two' < 'all three' brain regions as expected since the SNVs present in all three regions were more likely to represent early developmental SNVs which would therefore have higher VAFs. The number of somatic SNVs detected that were present in 'any two brain regions' was slightly less (~280 SNVs) than those present in 'only one' brain region (~380 SNVs), and both were much lesser in number than those present in 'all three' brain regions (~650 SNVs), with the intronic regions showing the highest number of variations, followed by comparable numbers of UTR3, ncRNA intronic, ncRNA exonic, and the exonic regions, while UTR5 and the splicing regions had the least number of SNVs present in all three conditions.

The accumulative frequency calculations along 'only one' *vs* 'any two' *vs* 'all three' brain regions in **Fig. 62b** [x-axis: VAF, y-axis: accumulative frequency] also reflects this trend, wherein the accumulative frequency is highest in 'all three' brain regions (light green) where the slope curves at an accumulative frequency of ~ 0.75 compared to the other two planes of comparison (light and dark blue) where the slopes curve at an accumulative frequency of ~1.00.

Meanwhile, there is no difference in the accumulative frequency amongst different brain regions of FTD disease wherein the slopes curve at an accumulative frequency of ~0.75 for all three brain regions as seen in **Fig. 62c** [x-axis: VAF, y-axis: accumulative frequency], suggesting that somatic mutational load in each of the brain regions of the FTD-tau patient comparable.



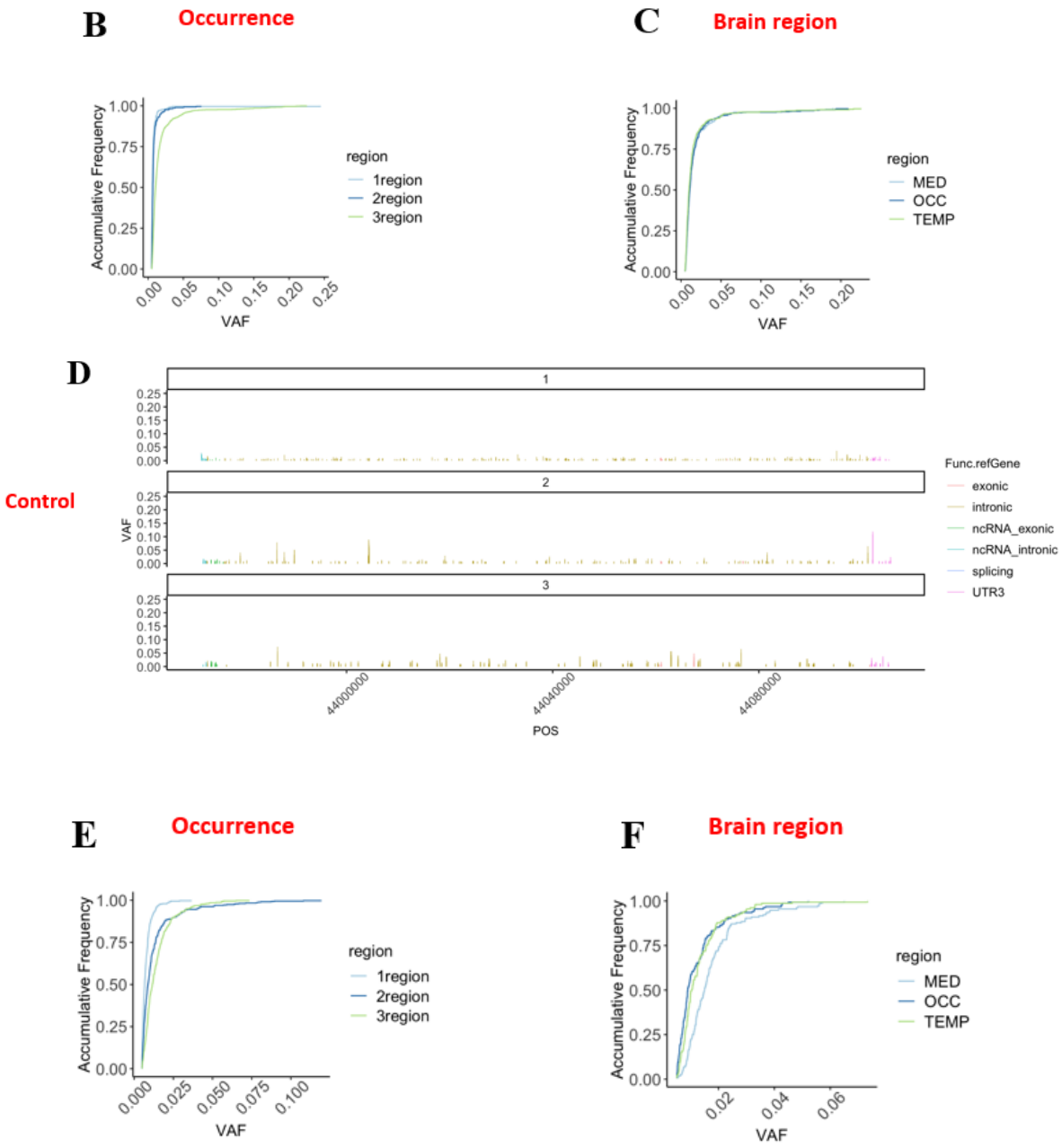


Figure 62: Double capture-based protocol 2 (with IDT probes for whole MAPT gene) – Mutational frequency calculations.

Mutational frequencies in:

FTD disease – (a) *'Only one' vs 'any two' vs 'all three' brain regions* – Each bar represents a somatic SNV detected in the chromosomal position along MAPT; the greatest number of bars with the higher VAFs seen in the plane of 'all three' brain regions (b) *Occurrence of accumulative frequency* – Higher accumulative frequency in 'all three' brain regions (light green) (at ~0.75) compared to 'any two' (dark blue) and 'only one' (light blue) brain regions (at ~1.00 for both planes of brain regions) (c) *Accumulative frequency within each brain region* – No difference in accumulative frequency seen in Temp (pale green) *vs* Occ (dark blue) *vs* Med (light blue) brain regions, suggesting comparable somatic mutational loads in the three brain regions of the patient.

FTD control – (d) *'Only one' vs 'any two' vs 'all three' brain regions* – Each bar represents a somatic SNV detected in the chromosomal position along MAPT; comparable number of bars with not much variations in the VAFs seen in all the planes of comparison of brain regions (e) *Occurrence of accumulative frequency* – Higher accumulative frequency in 'all three' (light green) and 'any two' (dark blue) brain regions (at ~0.80 for both planes of brain regions) compared to 'only one' (light blue) brain region (at ~1.00), (f) *Accumulative frequency withing each brain region* – Highest accumulative frequency seen in Med (light blue) (at ~ 0.60) compared to Temp (pale green) and Occ (dark blue) brain regions (at ~0.75 for both planes of brain regions).

FTD: Frontotemporal dementia; MAPT: Microtubule-associated protein tau; Med: Medulla; Occ: Occipital lobe; SNVs: Single nucleotide variations; VAF: Variable allele frequency; Temp: Temporal lobe.

Fig. 62d [x-axis: chromosomal position, y-axis: VAF] shows that shows the VAF each nucleotide across the MAPT chromosome that is present in either 'only one' brain region (top row), in 'any two' brain regions (middle row), or in 'all three' brain regions (bottom row) of the **age-matched control**. The VAF of the SNVs was mostly comparable in 'only one' ~ 'any two' ~ 'all three' brain regions, indicating that the SNVs present in all three regions were more likely to represent late-arising SNVs which would therefore have lower VAFs. The number of somatic SNVs detected that were present in 'any two brain regions' was slightly higher (~440 SNVs) than those present in 'all three' brain region (~390 SNVs), and both were much higher in number than those present in 'only one' brain region (~270 SNVs), with the intronic regions showing the highest number of variations, followed by comparable numbers of UTR3, ncRNA intronic, ncRNA exonic, and exonic regions, while the splicing regions had the least number of SNVs present in all three conditions.

This is also reflected in the accumulative frequency calculations along 'only one' *vs* 'any two' *vs* 'all three' brain regions in **Fig. 62e** [x-axis: VAF, y-axis: accumulative frequency], wherein the accumulative frequency is higher in 'all three' (light green) and 'any two' (dark blue) brain regions where the slopes curve at an accumulative frequency of ~ 0.80 compared to 'only one' brain region (light blue) where the slope curves at an accumulative frequency of ~ 1.00.

The medulla (light blue) shows a higher accumulative frequency where the slope curves at an accumulative frequency of ~ 0.60 compared to temporal (pale green) and occipital (dark blue) lobes of the control where the slopes curve at an accumulative frequency of ~ 0.75 as seen in **Fig. 62f** [x-axis: VAF, y-axis: accumulative frequency], indicating that the medulla of the control has the highest somatic mutational load of the three brain regions which could either be an individual-specific characteristic or may be related any unknown causes or conditions, and further investigations would be needed to determine this.

Thus, the IDT double capture protocol was successfully used to prepare duplex libraries for all the three FTD-Tau brain samples and age-matched control by using a large pool of custom-designed probes to capture libraries corresponding to both the '+' and '-' strands of the whole MAPT gene, and the computational data analyses gave us an insight into the origin and role of MAPT mutations in FTD-Tau pathology.

7.2.4 Summary of results for Double capture Protocol 2 DS (with IDT probes for whole MAPT gene)

For the second protocol adopter for double capture-based DS, duplex libraries for all six samples were made as per the optimized IDT protocol; however, further optimization of library amplification after second capture was necessary due to the increased number of probes used in the protocol described in this chapter. The final HS D5000 ScreenTape results for all six samples showed a single large peak at ~300 bp as expected and were submitted for Illumina NovaSeq. The quality check on the sequencing reads proved that our duplex libraries were constructed correctly and sufficient reads were present to make a duplex consensus and detect true mutations.

The investigations on mutational burden and frequency revealed that germline and somatic SNVs arising earlier during neurodevelopment were associated with FTD-Tau disease, and that these seem to be concentrated around intronic regions and a few UTR3, ncRNA intronic, ncRNA exonic, exonic, UTR5, and splicing regions of the MAPT gene. The somatic mutational burden in the three brain regions investigated in the FTD-Tau patient were also comparable, further indicating that the selective frontotemporal degeneration in FTD-Tau may be attributed to other risk factors or downstream processes in these neurons in addition to increased accumulation of mutations in the temporal lobe of the patient (compared to the control), and further analyses in the future will be needed to shed some more light on the role of these mutations in FTD.

7.3 Discussion

This chapter presented the wet-lab and computational results derived from the previously optimized protocols of the amplicon-based duplex library preparations of NP16.163 disease and C576 age-matched control based on amplicons prepared with Platinum SuperFi II DNA polymerase as well as the results of the duplex library preparation and subsequent capture using 1,696 custom-designed probes for both strands of the MAPT gene done on all six brain regions of the FTD-Tau disease patient and the control and the preliminary results obtained after sequencing data analyses.

The amplicon-based approach for duplex sequencing was successful in that duplex libraries were constructed and analysed successfully. It is a highly useful protocol to detect ‘true mutations’ within a shorter time period and can concentrate on the nuclear genome with great

specificity (<10 kb length). However, a major drawback to this approach is the fact that it requires a pre-library amplification stage which could introduce untraceable early-stage PCR errors into the system.

Therefore, we explored the results of the duplex library preparation and subsequent capture using 1,696 custom-designed probes for both strands of the MAPT gene done on all the three FTD-Tau and the age-matched control samples as well as the preliminary results obtained after sequencing data analyses to overcome the drawbacks of the amplicon-based approach. This method gave a higher quality of sequencing and detected mutations with greater accuracy for larger nuclear genomic regions (> 10 kb length). However, the duplex library preparation process is more laborious and the need for larger genomic region to be analysed can increase the cost of sequencing significantly compared to SuperFi-mediated amplicon-based DS.

Chapter VIII: Conclusions and Future Directions

Neurodegenerative disease (NDD) is an umbrella term referring to the debilitating and currently incurable set of conditions resulting in the progressive degeneration and/or death of neurons that can cause problems in movement (e.g.: ataxias, chorea, dystonia, etc.) or cognition (e.g.: dementias, attention deficits, aphasias, etc.) (JNPD Research 2019). Ageing is a natural phenomenon associated with an accumulation of mutations with time (Cagan et al. 2022), and some of these mutations are associated with disease signatures that can result in any one of the NDD (Lodato et al. 2018). The increasing mutational burden may be attributed to several endogenous and exogenous factors resulting in the formation of SNVs, CNVs, Indels, etc. (Leija-Salazar, Piette, and Proukakis 2018). Although the relationship between somatic mutational burden and ageing and disease has long been theorized (Curtis 1965), it has only been through recent advances in sequencing technologies, particularly that of error-corrected NGS and scDNA seq capable of pushing past the limits of detection, that has made it possible to attempt to determine the true mutation rates of sub-clonal variations (Risques and Kennedy 2018).

Frontotemporal dementia (FTD), an NDD that accounts for 10-20% of the global dementia cases, is a heterogenous disease that is associated with the selective degeneration of the frontal and temporal lobes of the brain (Lebert et al. 1998; Finger 2016). About 43% of the global FTD cases are inherited with an autosomal dominant pattern, and mutations in three genes – MAPT, GRN, and the C9orf72 hexanucleotide repeat expansion – are the most prevalent (Moore et al. 2020). Several pathologies have been linked to various clinical phenotypes of FTD, indicating a complex interplay between mutations, protein aggregations, and clinical manifestation of FTD (Dickerson, Ducharme, and Onyike 2016), and determining the role and origin of somatic mutations in each FTD pathology may lead to a better understanding of how the disease is expressed in patients.

This chapter includes an overview of the primary findings of this thesis along with an assessment of the strengths and limitations of the results and research approaches used to detect ultralow-frequency mutations in the temporal, occipital, and medulla brain regions of FTD-Tau patient and age-matched control via Duplex Sequencing. Potential strategies to further

understand NDD and ageing in the future have also been expounded upon, and the chapter ends with concluding remarks on this research project.

8.1 Primary Findings and Discussions

8.1.1 Comparison between the different duplex sequencing approaches

Duplex sequencing (DS) is an error-corrected NGS technique that has an extremely high accuracy, capable of detecting mutation frequencies as low as (5×10^{-8}) due to its duplex tagging system (Kennedy et al. 2014), and can be used to detect sub-clonal variations via bulk sequencing. Unfortunately, the very nature of this technique also requires a very high sequencing depth which necessitates high expenses, so WGS is very difficult using this technique.

To overcome these limitations, two approaches to duplex sequencing were explored in this thesis – the amplicon-based method and the double capture-based method – and the duplex libraries constructed using both these methods were analysed with appropriate bioinformatic pipelines (Kennedy et al. 2014) to detect mutations in the MAPT gene of three brain regions of FTD-Tau patient and control.

Two sets of duplex libraries were further made using the amplicon-based method: pre-library amplifications done with (a) Takara PrimeSTAR GXL DNA polymerase (Standard polymerase), and (b) Platinum SuperFi II DNA polymerase (Test polymerase) with higher fidelity. We found this approach to be easier to perform with less than two days of library preparation, useful in analysing exact, but small regions of interest (<10 kb length), requiring an initial DNA input of only 50 ng of DNA. However, the possibility of introducing untraceable early-stage PCR errors into the system meant that the detection of ‘true mutations’ could be compromised.

To circumvent this, the double capture-based method of DS was also investigated. Our first attempt included the capture protocol described by Maricic, Whitten, and Pääbo but modified for nuclear genes instead of mitochondrial DNA (Maricic, Whitten, and Pääbo 2010), which proved to be unsuccessful due to the highly reduced initial number of gene copies (2 copies per nuclear gene in a cell) in the system that were difficult to capture with only six MAPT probes. Our second attempt using the IDT Hybridization and Capture protocol as mentioned in (Schmitt et al. 2015) was successful since large capture pool panels with 150+ MAPT probes in a trial

attempt and 16+ probes in the final attempt were used and we were able to determine the mutation frequencies of germline and somatic SNVs in all three brain regions of FTD-Tau patient and control. Although this approach needed a longer preparation process (3.5-4 days) requiring an initial DNA input of ~500-1000 ng of DNA, we found that the results of sequencing larger regions of interest (10-100 kb length) were more accurate due to the reduced possibility of introducing PCR-errors into the system.

Our results show that using the capture-based method is the best approach to Duplex Sequencing for nuclear gene regions instead of either of the amplicon-based methods that used a pre-library amplification step (that may introduce untraceable PCR-errors into the system) as indicated by the optimal average peak family size obtained as well as the number of mutations detected by the double capture-based method. However, the SuperFi-mediated amplicon-based method also has advantages in that it can estimate the trends of mutation variations quite accurately and can be done on smaller genomic regions within lesser time which could be useful when limited resources are at hand.

8.1.2 Mutational burden and role of somatic and germline SNVs in the FTD-Tau patient

The genetic underpinnings of adult-onset NDD such as FTD can be extremely complex – in some cases, the disease can be inherited in a Mendelian fashion as an autosomal dominant trait in families with a positive history while in other cases the disease is a sporadic occurrence due to a post-zygotic or a late-somatic mutation within the brain (Nicolas and Veltman 2018). FTD-Tau, most commonly manifesting as bvFTD, accounts for ~30% of familial FTD and 10-30% of sporadic FTD cases wherein the most prevalent gene mutations are present in the MAPT gene on chromosome 17 (Van Swieten and Spillantini 2007).

In our study, we found that the overall mutational burden within the MAPT gene of the FTD-Tau patient was higher than the age-matched control as observed in the data of the number of mutations detected and per-bp mutation rates of all the brain regions of the patient and the control. While the germline homozygous SNVs (VAF ~ 1.00) remained constant in both the individuals across all three brain regions under investigation, germline heterozygous SNVs (VAF > 0.25 and VAF < 1.00) were much higher in number and the somatic SNVs (VAF < 0.25) were slightly higher in the brain regions of the patient than in the corresponding control brain regions. Moreover, the germline SNVs (VAF > 0.25 and VAF < 1.00) and the ultralow

frequency somatic SNVs (VAF < 0.01) detected in the patient was statistically significant compared to control, indicating that both somatic and germline SNVs were contributing to FTD pathology in the patient.

Additionally, the number of somatic SNVs detected (VAF ≥ 0.01 and VAF < 0.05) in FTD-Tau temporal lobe *versus* occipital lobe as well as the number of somatic (VAF ≥ 0.01 and VAF < 0.05) and ultralow frequency somatic (VAF < 0.01) SNVs detected in the MAPT-B amplicon (with the most number of established pathogenic site on the MAPT gene as per literature) of the FTD-Tau patient suggested that the exons 9-10 of the MAPT gene in the temporal lobe may be particularly susceptible to accumulation of mutations that are associated with FTD-Tau pathology.

We also found that per-bp mutation rates were the highest in the intronic regions of the patient, followed by the exonic, UTR3, ncRNA exonic, and the ncRNA intronic regions of the patient, which were all higher than those of the control. Splicing SNVs had the lowest mutation rates of all the genomic regions and were comparable in all the brain regions of patient and control, indicating that they played no significant role in FTD pathogenesis in the patient, although further analyses will be needed to confirm this statement. No mutations were detected in the UTR5 region of the control, suggesting that the UTR5 SNVs in the patient could be contributing to the disease state. Thus, genomic regions other than those identified to be directly involved in transcription, translation, and their processes of regulation in brain cells may also be contributing to FTD pathology

8.1.3 Origin of somatic mutations in FTD disease progression

For a long time, sequencing technologies had severe limits of detection that made it difficult to directly determine mutational frequency of somatic SNVs in any NDD. While attempts have been made to indirectly measure the somatic rates in different NDD (Wei et al. 2019), the emergence of single-cell sequencing and error-corrected NGS technologies has changed this status quo in recent years.

The use of duplex sequencing on different brain regions of the FTD patient and the control in our study allowed us to observe post-zygotic and late-somatic mutations by determining all the somatic mutational frequencies and predicting when the mutations might have arisen in the patient's lifetime.

Our results indicate that the VAF of the SNVs detected as well as the accumulative frequency of somatic SNVs is higher in 'all three' brain regions when compared to 'any two' or 'only one' brain region of the FTD patient. Moreover, the accumulative frequency of temporal lobe, occipital lobe, and the medulla of the patient were all comparable, suggesting that somatic mutations arising earlier along neurodevelopment were likely to contribute to FTD pathogenesis in the patient.

Being a clinically sporadic case, the FTD-Tau patient is interesting in that the DNA analyses results convey that ultralow frequency somatic SNVs (VAF < 0.01) are significantly present in the patient, and that early-arising somatic SNVs (VAF ≥ 0.05 and VAF < 0.25) and germline heterozygous SNVs (VAF > 0.25 and VAF < 1.00) were associated with the ageing disease state, and this sheds some light on the nature of FTD pathogenesis and disease progression in FTD-Tau pathology observed in the patient.

8.2 Strengths of the research project

This thesis project offers an extended overview of the process of optimizing and establishing DS in a lab. DS is often difficult to adapt in that successful library construction can only be done after careful regulation of various parameters at optimal levels. A very good understanding of both – the library construction process on the bench as well as an understanding of the computational workflow – proved to be necessary to obtain usable sequencing data. This thesis evaluates the pros and cons of the different approaches to DS, particularly in the context of detecting low-frequency mutations in nuclear genes.

DS, being an error-corrected NGS technique, has a higher accuracy rate than most conventional bulk NGS, which cannot distinguish between sequencing errors and somatic mutations with < 0.5% mosaicism (Sloan et al. 2018). Unpublished data in the lab using conventional NGS Illumina and SureSelect library preparation kits for the same three MAPT amplicons described in this thesis sequenced on a MiSeq platform showed that while heterogeneous mutations were detected well, only a negligible number of somatic mutations (VAF < 0.25) were detected (**Fig. 63**), further demonstrating the strength of DS in its higher sensitivity in detecting lower frequency mutations.

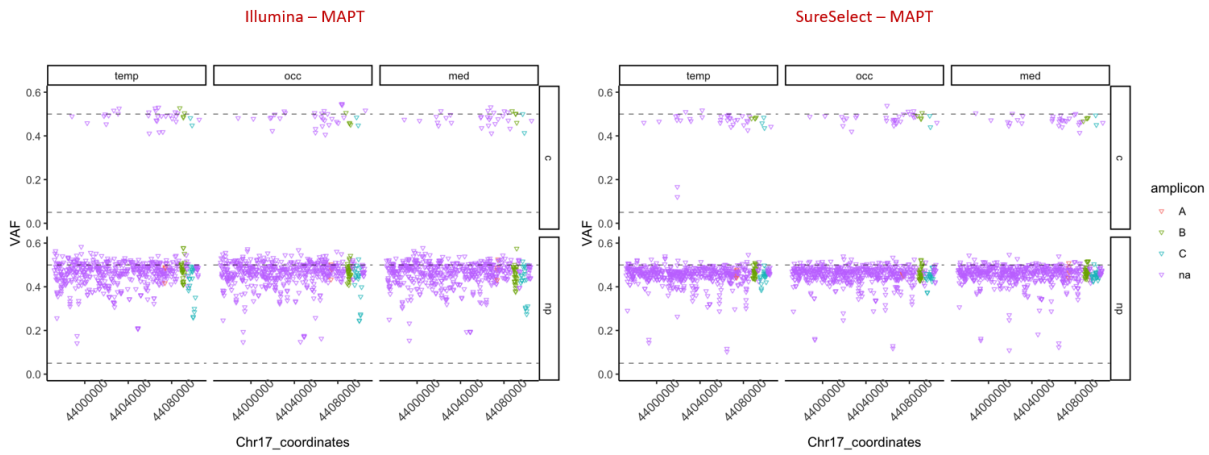


Figure 63: Mutation detection on FTD-Tau patient and control with conventional NGS.
Conventional sequencing libraries were prepared using the same amplicons used in the present thesis. The red spots correspond to germline mutations (VAF ~ 0.50) detected in the MAPT-A amplicon, green spots to those MAPT-B amplicon, blue spots to those in MAPT-C amplicon, and purple spots are unknown germline mutations (VAF > 0.25) detected. The extremely high number of unknown germline mutations (purple spots) as well as the lack of somatic mutation detection (VAF < 0.25) are indicative of the low accuracy of conventional NGS done on libraries prepared with Illumina and SureSelect library preparation kits (Y Nie and A Murley, unpublished data).
MAPT: Microtubule-associated protein tau; NGS: Next generation sequencing; VAF: Variable allele frequency.

The literature suggests that it is often easier to perform WES on a panel of genes suspected to be associated with the disease or perform WGS at lower read depths (Chapter 1, section 4.3.1). Due to the nature of the DS and the need for a large probe panel covering greater genomic lengths, we were able to sequence the entire MAPT gene in FTD-Tau pathology at very high read depths (> 1,000X), a novel feature in our work. In addition to detecting exonic mutations, determining the mutational burden of the intronic, ncRNA, UTR, and splicing regions, and discovering their role may play a key role in understanding disease progression.

The frozen brain samples used in the study contained tissues from three brain regions – the temporal lobe, occipital lobe, and the medulla. This provided an internal control within an individual to discern the levels of mutations in each region and determine when the mutations might have occurred, with the medulla having developed separately from the cerebrum and the temporal and occipital lobes differentiating a little later during neurodevelopment in the both the individuals. Moreover, the ability to compare these results from the patient with the corresponding regions of an age-matched control belonging to the same gender made it easier to easily ascertain the nature of the mutational burden in the FTD patient compared to an external control without gender-based confounding factors.

8.3 Limitations of the research project

A significant drawback of the optimized DS protocol is the high amounts of starting material as well as the high sequencing depth required, i.e., the high fidelity of the technique comes at the cost of the high throughput levels of the sequencing run. It is difficult to obtain usable sequencing data with more than eight samples on the Illumina MiSeq and more than twelve on the Illumina NovaSeq platforms, which would limit the number of genes/samples being considered for sequencing.

While DS overcomes some of the disadvantages of conventional bulk sequencing with respect to accuracy and sensitivity, the starting material is native DNA extracted from frozen brain tissue, thus making it impossible to determine the cellular source of the DNA and any data associated with the genomic heterogeneity of the brain cells cannot be recovered or retraced.

Another major limitation in our work is the sample size: we had only one FTD-Tau patient and an age-matched control, which makes it difficult to determine if our results are representative of the general trends, and if not, how unique the case is. Additionally, although FTD-Tau affects men and women equally (Ghetti et al. 2015), our sample size of two individuals represented only the female gender, which could occlude any data on how physiological differences may potentially affect disease progression.

8.4 Future Investigations

8.4.1 Applying optimized DS protocol to other FTD pathologies

In addition to the FTD-Tau pathology, the frozen brain samples described in Chapter 3, section 3.1.1, **Table 7**, also contain the three brain regions from 15 other patients that show FTD-TDP43, FTD-UPS, and FTD-ALS pathologies, along with their age-matched controls. Applying our optimized IDT double capture-based duplex library preparation will no doubt provide greater understanding of the ‘true’ mutational burden and frequency in other FTD pathologies; however, this may prove to be more difficult to perform as the genes involved in these other pathologies are more ambiguous.

For instance, FTD-TDP43 pathology, or the FTD-TDP Type B, is exhibited in ~50% of the FTD cases, but the TARDBP gene, which accounts for the formation of the TDP-43 protein, is only mutated in less than 2% FTD cases till date, while C9orf72 hexanucleotide repeat

expansions are the most common mutation type in the Type B pathology (Diana A. Olszewska et al. 2016; Jo et al. 2020). Mutations in VCP, DCTN1, GRN have also been reported to be associated with this pathology, but this is not always the case (Benussi, Padovani, and Borroni 2015). Similar ambiguity exists in the correlation between genes involved, FTD pathology, and clinical phenotype for the other pathologies (see Chapter 1, section 1.4.3.2).

Therefore, developing a gene panel to perform a sequencing run with higher sensitivity to detect somatic mutations and then performing DS on the genes with a somatic mutational burden in a FTD patient, irrespective of the pathology displayed, would be an ideal expansion of this project. However, the characteristic feature of the higher throughput levels that DS typically needs has to be considered since a large gene panel may not be feasible.

8.4.2 Single-cell DNA sequencing

While error-corrected NGS and single-cell DNA sequencing have both breached through the limits of detection and are capable of detecting even late-arising and unselected SNVs and small indels across the genome (Risques and Kennedy 2018), they still are vastly different technologies and sequencing data generated from either techniques offer different insights.

Duplex sequencing uses the complex molecular duplex tags to uniquely identify reads coming from the same library via specialized computational workflows to detect 'true' mutations (Kennedy et al. 2014). This means that while duplex sequencing increases the accuracy of sequencing, it essentially remains a bulk sequencing technique and has similar drawbacks – the DNA content of thousands-millions of cells has been homogenised before duplex libraries can be created. It is impossible to determine exactly which types of brain cells contributed to the detection of any one mutation and therefore, we cannot predict the origin of mutations with VAF < 0.01 and thus, important information could be lost.

scDNA seq is characterized by its three core characteristics: *fidelity*, which allows the resolution of DNA features present at lower levels of mosaicism; *co-presence*, to ascertain whether these low-level mosaic variants are present in the same cells or mutually exclusive subset of cells; and, *phenotypic association*, to simultaneously identify the cell type(s) or cell state(s) these specific DNA features are present (Evrony, Hinch, and Luo 2021). The capabilities of co-presence and phenotypic association in particular are important as they can differentiate between genetically distinct heterogeneous cells, making it possible to reconstruct cell lineage trees by tracking the spontaneous development of somatic mutations within a

cellular lifetime through retrospective lineage tracing (Woodworth, Girskis, and Walsh 2017). Therefore, performing single-cell DNA sequencing on all the brain samples in addition to the duplex sequencing will be an essential aspect to this research project to obtain greater understanding of the origin of somatic mutations and their effect on FTD pathogenesis over time.

8.5 Summary

This thesis summarizes the various attempts made to optimize two different approaches to duplex sequencing. This technique was applied to DNA extracted from three brain regions of an FTD-Tau patient and age matched control – from the temporal and occipital lobes, and the medulla. The results of the sequencing data after appropriate analyses have been presented for a comparison of both the sequencing approaches. Double capture-based duplex library preparation was found to have a higher accuracy than the amplicon-based method likely due to fewer rounds of amplification in the former that reduces the introduction of untraceable early PCR-errors. Results from the sequencing data of all the six brain regions showed that the overall mutational burden in the FTD-Tau patient is higher than the control, along with the novel discovery that somatic SNVs arising earlier along neurodevelopment ($VAF \geq 0.05$ and $VAF < 0.25$) as well as germline heterozygous variations ($VAF > 0.25$ and $VAF < 1.00$) were associated with FTD pathogenesis in the patient, and that exons 9-10 of the MAPT gene in the temporal lobe was particularly susceptible to a slightly increased mutation accumulation compared to the occipital lobe and the medulla in FTD-Tau pathology. Somatic SNVs ($VAF < 0.25$) found within the intronic, exonic, ncRNA, and the UTR regions have a higher presence in the patient with respect to the control, indicating a relationship between somatic mutational burden in genomic regions not directly involved in the transcription and translation of the tau gene and disease state in this sporadic case of FTD. Future research should focus on: (i) identifying any mutational signatures correlated with FTD-Tau, (ii) applying the DS protocol on samples with other FTD pathologies, and (iii) performing scDNA seq to directly determine somatic mutation rates and constructing cell lineage trees to better understand ageing and disease in the brain.

References

Allen, Nicola J., and Ben A. Barres. 2009. “Neuroscience: Glia - More than Just Brain Glue.” *Nature* 457 (7230): 675–77. https://doi.org/10.1038/457675a.

Andrews, S. 2010. “FASTQC: A Quality Control Tool for High Throughput Sequence Data.” (Online). 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anoar, Sharifah, Nathaniel S. Woodling, and Teresa Niccoli. 2021. “Mitochondria Dysfunction in Frontotemporal Dementia/Amyotrophic Lateral Sclerosis: Lessons From Drosophila Models.” *Frontiers in Neuroscience* 15 (November): 1–22. https://doi.org/10.3389/fnins.2021.786076.

Ansorge, Wilhelm J. 2009. “Next-Generation DNA Sequencing Techniques.” *New Biotechnology* 25 (4): 195–203. https://doi.org/10.1016/j.nbt.2008.12.009.

Armstrong, Melissa J., and Michael S. Okun. 2020. “Diagnosis and Treatment of Parkinson Disease: A Review.” *JAMA - Journal of the American Medical Association* 323 (6): 548–60. https://doi.org/10.1001/jama.2019.22360.

Arrant, Andrew E., Vincent C. Onyilo, Daniel E. Unger, and Erik D. Roberson. 2018. “Progranulin Gene Therapy Improves Lysosomal Dysfunction and Microglial Pathology Associated with Frontotemporal Dementia and Neuronal Ceroid Lipofuscinosis.” *Journal of Neuroscience* 38 (9): 2341–58. https://doi.org/10.1523/JNEUROSCI.3081-17.2018.

Arranz, Amaia M., and Bart De Strooper. 2019. “The Role of Astroglia in Alzheimer’s Disease: Pathophysiology and Clinical Implications.” *The Lancet Neurology* 18 (4): 406–14. https://doi.org/10.1016/S1474-4422(18)30490-3.

Artan, Sevilhan, Ebru Erzurumluoglu Gokalp, Bedia Samanci, Demet Ozbabalik Adapinar, Hasan Bas, Fatih Tepgec, Emilia Qomi Ekenel, et al. 2021. “Frequency of Frontotemporal Dementia-Related Gene Variants in Turkey.” *Neurobiology of Aging* 106: 332.e1-332.e11. https://doi.org/10.1016/j.neurobiolaging.2021.05.007.

Baker, Matt, Ian R Mackenzie, Stuart M Pickering-brown, Jennifer Gass, Rosa Rademakers, Caroline Lindholm, Julie Snowden, et al. 2006. “Mutations in Progranulin

Cause Tau-Negative Frontotemporal Dementia Linked to Chromosome 17." *Nature* 442 (August): 13–16. https://doi.org/10.1038/nature05016.

Bakulski, Kelly M., Dana C. Dolinoy, Maureen A. Sartor, Henry L. Paulson, John R. Konen, Andrew P. Lieberman, Roger L. Albin, Howard Hu, and Laura S. Rozek. 2012. "Genome-Wide DNA Methylation Differences between Late-Onset Alzheimer's Disease and Cognitively Normal Controls in Human Frontal Cortex." *Journal of Alzheimer's Disease* 29 (3): 571–88. https://doi.org/10.3233/JAD-2012-111223.

Bannwarth, Sylvie, Samira Ait-El-Mkadem, Annabelle Chaussenot, Emmanuelle C. Genin, Sandra Lacas-Gervais, Konstantina Fragaki, Laetitia Berg-Alonso, et al. 2014. "A Mitochondrial Origin for Frontotemporal Dementia and Amyotrophic Lateral Sclerosis through CHCHD10 Involvement." *Brain* 137 (8): 2329–45. https://doi.org/10.1093/brain/awu138.

Banzhaf-Strathmann, Julia, Rainer Claus, Oliver Mücke, Kristin Rentzsch, Julie van der Zee, Sebastiaan Engelborghs, Peter P De Deyn, et al. 2013. "Promoter DNA Methylation Regulates Progranulin Expression and Is Altered in FTLD." *Acta Neuropathologica Communications* 1 (1): 1–15. https://doi.org/10.1186/2051-5960-1-16.

Barber, Robert C. 2012. "The Genetics of Alzheimer's Disease." *Scientifica*, 1–14. https://doi.org/https://doi.org/10.6064/2012/246210.

Bartolome, Fernando, Noemi Esteras, Angeles Martin-Requero, Claire Boutoleau-Bretonniere, Martine Vercelletto, Audrey Gabelle, Isabelle Le Ber, et al. 2017. "Pathogenic P62/SQSTM1 Mutations Impair Energy Metabolism through Limitation of Mitochondrial Substrates." *Scientific Reports* 7 (1): 1–14. https://doi.org/10.1038/s41598-017-01678-4.

Beart, Philip, Michael Robinson, Marcus Rattray, and Nicholas J Maragakis Editors. 2017. *Advances in Neurobiology (Vol. 15): Neurodegenerative Diseases Pathology, Mechanisms, and Potential Therapeutic Targets*. *Book*.

Belzil, Veronique V., Peter O. Bauer, Mercedes Prudencio, Tania F. Gendron, Caroline T. Stetler, Irene K. Yan, Luc Pregent, et al. 2013. "Reduced C9orf72 Gene Expression in C9FTD/ALS Is Caused by Histone Trimethylation, an Epigenetic Event Detectable in

Blood." *Acta Neuropathologica* 126 (6): 895–905. https://doi.org/10.1007/s00401-013-1199-1.

Benraiss, Abdellatif, Su Wang, Stephanie Herrlinger, Xiaojie Li, Devin Chandler-Militello, Joseph Mauceri, Hayley B. Burm, et al. 2016. "Human Glia Can Both Induce and Rescue Aspects of Disease Phenotype in Huntington Disease." *Nature Communications* 7. https://doi.org/10.1038/ncomms11758.

Benussi, Alberto, Alessandro Padovani, and Barbara Borroni. 2015. "Phenotypic Heterogeneity of Monogenic Frontotemporal Dementia." *Frontiers in Aging Neuroscience* 7 (SEP): 1–19. https://doi.org/10.3389/fnagi.2015.00171.

Ber, Isabelle Le, Agnès Camuzat, Rita Guerreiro, and Kawtar Bouya-. 2013. "SQSTM1 Mutations in French Patients With Frontotemporal Dementia or Frontotemporal Dementia With Amyotrophic Lateral Sclerosis." *JAMA Neurol.* 70 (11): 1403–10. https://doi.org/10.1001/jamaneurol.2013.3849.SQSTM1.

Blauwendraat, Cornelis, Carlo Wilke, Javier Simón-Sánchez, Iris E. Jansen, Anika Reifschneider, Anja Capell, Christian Haass, et al. 2018. "The Wide Genetic Landscape of Clinical Frontotemporal Dementia: Systematic Combined Sequencing of 121 Consecutive Subjects." *Genetics in Medicine* 20 (2): 240–49. https://doi.org/10.1038/gim.2017.102.

Bleidorn, Christoph. 2016. "Third Generation Sequencing: Technology and Its Potential Impact on Evolutionary Biodiversity Research." *Systematics and Biodiversity* 14 (1): 1–8. https://doi.org/10.1080/14772000.2015.1099575.

Boeve, Bradley F., Ivo W. Tremont-Lukats, Andrew J. Waclawik, Jill R. Murrell, Bruce Hermann, Clifford R. Jack, Maria M. Shiung, et al. 2005. "Longitudinal Characterization of Two Siblings with Frontotemporal Dementia and Parkinsonism Linked to Chromosome 17 Associated with the S305N Tau Mutation." *Brain* 128 (4): 752–72. https://doi.org/10.1093/brain/awh356.

Borroni, Barbara, Francesca Ferrari, Daniela Galimberti, Benedetta Nacmias, Cinzia Barone, Silvia Bagnoli, Chiara Fenoglio, et al. 2014. "Heterozygous TREM2 Mutations in Frontotemporal Dementia." *Neurobiology of Aging* 35 (4): 934.e7-934.e10. https://doi.org/10.1016/j.neurobiolaging.2013.09.017.

Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, et al. 2018. “Ten Things You Should Know about Transposable Elements.” *Genome Biology* 19 (1): 199. https://doi.org/10.1186/s13059-018-1577-z.

Brás, Inês C., and Tiago F. Outeiro. 2021. “Alpha-Synuclein: Mechanisms of Release and Pathology Progression in Synucleinopathies.” *Cells* 10 (2): 1–19. https://doi.org/10.3390/cells10020375.

Breijyeh, Zeinab, and Rafik Karaman. 2020. “Comprehensive Review on Alzheimer ’ s Disease :” *Molecules* 25 (24): 57–89.

Broe, Melissa, Julian Kril, and Glenda M. Halliday. 2004. “Astrocytic Degeneration Relates to the Severity of Disease in Frontotemporal Dementia.” *Brain* 127 (10): 2214–20. https://doi.org/10.1093/brain/awh250.

Bruni, Amalia C., Junko Takahashi-Fujigasaki, Francesca Maltecca, Jean Francois Foncin, Antonio Servadio, Giorgio Casari, Pio D’Adamo, et al. 2004. “Behavioral Disorder, Dementia, Ataxia, and Rigidity in a Large Family with TATA Box-Binding Protein Mutation.” *Archives of Neurology* 61 (8): 1314–20. https://doi.org/10.1001/archneur.61.8.1314.

Buratti, Emanuele, and Francisco E. Baralle. 2010. “The Multiple Roles of TDP-43 in Pre-MRNA Processing and Gene Expression Regulation.” *RNA Biology* 7 (4): 420–29. https://doi.org/10.4161/rna.7.4.12205.

Byrne J H. 1997. “Synaptic Plasticity.” In *Neuroscience Online*, Neuroscien. http://neuroscience.uth.tmc.edu/s1/chapter07.html.

Cagan, Alex, Adrian Baez-Ortega, Natalia Brzozowska, Federico Abascal, Tim H. H. Coorens, Mathijs A. Sanders, Andrew R. J. Lawson, et al. 2022. “Somatic Mutation Rates Scale with Lifespan across Mammals.” *Nature* 604 (April). https://doi.org/10.1038/s41586-022-04618-z.

Capouch, Samuel D., Martin R. Farlow, and Jared R. Brosch. 2018. “A Review of Dementia with Lewy Bodies’ Impact, Diagnostic Criteria and Treatment.” *Neurology and Therapy* 7 (2): 249–63. https://doi.org/10.1007/s40120-018-0104-1.

Cardarelli, Roberto, Andrew Kertesz, and Janice A Knebl. 2010. “Frontotemporal Dementia: A Review for Primary Care Physicians.” *American Family Physician* 82 (11): 1372–77. http://www.ncbi.nlm.nih.gov/pubmed/21121521.

Cattaneo, Elena, Chiara Zuccato, and Marzia Tartari. 2005. “Normal Huntingtin Function: An Alternative Approach to Huntington’s Disease.” *Nature Reviews Neuroscience* 6 (12): 919–30. https://doi.org/10.1038/nrn1806.

Chen, Gary K., Qi Yan, Kimberly C. Paul, Cynthia D.J. Kusters, Aline Duarte Folle, Melissa Furlong, Adrienne Keener, Jeff Bronstein, Steve Horvath, and Beate Ritz. 2022. “Stochastic Epigenetic Mutations Influence Parkinson’s Disease Risk, Progression, and Mortality.” *Journal of Parkinson’s Disease* 12 (2): 545–56. https://doi.org/10.3233/JPD-212834.

Chin, Chen Shan, David H. Alexander, Patrick Marks, Aaron A. Klammer, James Drake, Cheryl Heiner, Alicia Clum, et al. 2013. “Nonhybrid, Finished Microbial Genome Assemblies from Long-Read SMRT Sequencing Data.” *Nature Methods* 10 (6): 563–69. https://doi.org/10.1038/nmeth.2474.

Chiò, Adriano, Andrea Calvo, Cristina Moglia, Gabriella Restagno, Irene Ossola, Maura Brunetti, Anna Montuschi, et al. 2010. “Amyotrophic Lateral Sclerosis-Frontotemporal Lobar Dementia in 3 Families with p.Ala382Thr TARDBP Mutations.” *Archives of Neurology* 67 (8): 1002–9. https://doi.org/10.1001/archneurol.2010.173.

Chun, Jerold J. M., Marjorie A Oettinger, Rudolf Jaenisch, and David Baltimore. 1991. “The Recombination Activating Gene-l (RAG-I) Transcript Is Present in the Murine Central Nervous System.” *Cell* 64 (January): 189–200.

Ciani, Miriam, Luisa Benussi, Cristian Bonvicini, and Roberta Ghidoni. 2019. “Genome Wide Association Study and next Generation Sequencing: A Glimmer of Light toward New Possible Horizons in Frontotemporal Dementia Research.” *Frontiers in Neuroscience* 13 (MAY): 1–8. https://doi.org/10.3389/fnins.2019.00506.

Corcia, P., P. Couratier, H. Blasco, C. R. Andres, S. Beltran, V. Meininger, and P. Vourc’h. 2017. “Genetics of Amyotrophic Lateral Sclerosis.” *Revue Neurologique* 173 (5): 254–62. https://doi.org/10.1016/j.neurol.2017.03.030.

CoreGenomics. 2012. “How Do SPRI Beads Work?” 2012. http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.html.

Couratier, P., P. Corcia, G. Lautrette, M. Nicol, and B. Marin. 2017. “ALS and Frontotemporal Dementia Belong to a Common Disease Spectrum.” *Revue Neurologique* 173 (5): 273–79. https://doi.org/10.1016/j.neurol.2017.04.001.

Coyle-Gilchrist, Ian T.S., Katrina M. Dick, Karalyn Patterson, Patricia Vázquez Rodríquez, Eileen Wehmann, Alicia Wilcox, Claire J. Lansdall, et al. 2016. “Prevalence, Characteristics, and Survival of Frontotemporal Lobar Degeneration Syndromes.” *Neurology* 86 (18): 1736–43. https://doi.org/10.1212/WNL.0000000000002638.

Cruts, Marc, and Christine Van Broeckhoven. 2008. “Loss of Progranulin Function in Frontotemporal Lobar Degeneration.” *Trends in Genetics* 24 (4): 186–94.

Curtis, Howard J. 1965. “The Somatic Mutation Theory of Aging.” *Contributions to the Psychobiology of Aging*, 69–80. https://doi.org/10.1007/978-3-662-39849-4_6.

David, Della C. 2012. “Aging and the Aggregating Proteome.” *Frontiers in Genetics* 3 (NOV): 1–6. https://doi.org/10.3389/fgene.2012.00247.

Davis, Miriam, and Clare Stroud. 2013. *Neurodegeneration: Exploring Commonalities Across Diseases*.

Deerlin, Vivianna M Van, Patrick M A Sleiman, Maria Martinez-lage, Alice Chen-, Li-san Wang, Neill R Graff-radford, Dennis W Dickson, et al. 2010. “Common Variants at 7p21 Are Associated with with TDP-43 Inclusions.” *Nature Genetics* 42 (3): 234–39. https://doi.org/10.1038/ng.536.Common.

DeJesus-Hernandez, Mariely, Ian R. Mackenzie, Bradley F. Boeve, Adam L. Boxer, Matt Baker, Nicola J. Rutherford, Alexandra M. Nicholson, et al. 2011. “Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS.” *Neuron* 72 (2): 245–56. https://doi.org/10.1016/j.neuron.2011.09.011.

Dejesus-hernandez, Mariely, Ian R Mackenzie, Bradley F Boeve, Adam L Boxer, Matt Baker, Nicola J Rutherford, Alexandra M Nicholson, et al. 2011. “Expanded GGGGCC Hexanucleotide Repeat in Non-Coding Region of C9ORF72 Causes Chromosome 9p-Linked

Frontotemporal Dementia and Amyotrophic Lateral Sclerosis." *Neuron* 72 (2): 245–56. https://doi.org/10.1016/j.neuron.2011.09.011.Expanded.

Deng, Hao, Peng Wang, and Joseph Jankovic. 2018. "The Genetics of Parkinson Disease." *Ageing Research Reviews* 42 (December 2017): 72–85. https://doi.org/10.1016/j.arr.2017.12.007.

Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, et al. 2012. "The GENCODE v7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression." *Genome Research* 22 (9): 1775–89. https://doi.org/10.1101/gr.132159.111.

Dickerson, Bradford, Simon Ducharme, and Chiadi Onyike. 2016. *Overview of Clinical Assessment of Frontotemporal Dementia Syndromes*. *Hodges' Frontotemporal Dementia*. Vol. 3. https://doi.org/10.1212/wnl.0b013e3181a56016.

Dickey, Audrey S, and Albert R La Spada. 2018. "Therapy Development in Huntington's Disease: From Current Strategies to Emerging Oppotunities." *American Journal of Medical Genetics* 176 (4): 842–61. https://doi.org/10.1002/ajmg.a.38494.Therapy.

Do-Ha, Dzung, Yossi Buskila, and Lezanne Ooi. 2018. "Impairments in Motor Neurons, Interneurons and Astrocytes Contribute to Hyperexcitability in ALS: Underlying Mechanisms and Paths to Therapy." *Molecular Neurobiology* 55 (2): 1410–18. https://doi.org/10.1007/s12035-017-0392-y.

Domenico, Angelique di, Giulia Carola, Carles Calatayud, Meritxell Pons-Espinal, Juan Pablo Muñoz, Yvonne Richaud-Patin, Irene Fernandez-Carasa, et al. 2019. "Patient-Specific IPSC-Derived Astrocytes Contribute to Non-Cell-Autonomous Neurodegeneration in Parkinson's Disease." *Stem Cell Reports* 12 (2): 213–29. https://doi.org/10.1016/j.stemcr.2018.12.011.

Dugger, Brittany N., and Dennis W. Dickson. 2017. "Pathology of Neurodegenerative Diseases." *Cold Spring Harbor Perspectives in Biology* 9 (7): 1–22. https://doi.org/10.1101/cshperspect.a028035.

Dyken, Paul, and Nicolas Krawiecki. 1983. "Neurodegenerative Diseases of Infancy and Childhood." *Annals of Neurology* 13 (4): 351–64. https://doi.org/10.1002/ana.410130402.

Eberwine, James, Jai Yoon Sul, Tamas Bartfai, and Junhyong Kim. 2014. “The Promise of Single-Cell Sequencing.” *Nature Methods* 11 (1): 25–27. https://doi.org/10.1038/nmeth.2769.

Emmady, Prabhu D, and Jeffrey Bodle. 2022. “Werdnig Hoffmann Disease - PubMed.” *NCBI Bookshelf*, 1–5. https://pubmed.ncbi.nlm.nih.gov/32644359/.

Evans, Dana A P, J Peter H Burbach, and Fred W Van Leeuwen. 1995. “Somatic Mutations in the Brain: Relationship to Aging?” *Mutation Research* 338: 173–82.

Evrony, Gilad D., Anjali Gupta Hinch, and Chongyuan Luo. 2021. “Applications of Single-Cell DNA Sequencing.” *Annual Review of Genomics and Human Genetics* 22: 171–97. https://doi.org/10.1146/annurev-genom-111320-090436.

Fenoglio, Chiara, Elio Scarpini, and Daniela Galimberti. 2018. “Epigenetic Regulatory Modifications in Genetic and Sporadic Frontotemporal Dementia.” *Expert Review of Neurotherapeutics* 18 (6): 469–75. https://doi.org/10.1080/14737175.2018.1481389.

Filippini, Alice, Massimo Gennarelli, and Isabella Russo. 2019. “α-Synuclein and Glia in Parkinson’s Disease: A Beneficial or a Detrimental Duet for the Endo-Lysosomal System?” *Cellular and Molecular Neurobiology* 39 (2): 161–68. https://doi.org/10.1007/s10571-019-00649-9.

Finger, Elizabeth C. 2016. “Frontotemporal Dementias.” *Continuum (American Academy of Neurology)* 22 (2) (April): 464–89.

Forsberg, Lars A., Chiara Rasi, Hamid R. Razzaghian, Geeta Pakalapati, Lindsay Waite, Krista Stanton Thilbeault, Anna Ronowicz, et al. 2012. “Age-Related Somatic Structural Changes in the Nuclear Genome of Human Blood Cells.” *American Journal of Human Genetics* 90 (2): 217–28. https://doi.org/10.1016/j.ajhg.2011.12.009.

Fostinelli, Silvia, Miriam Ciani, Roberta Zanardini, Orazio Zanetti, Giuliano Binetti, Roberta Ghidoni, and Luisa Benussi. 2018. “The Heritability of Frontotemporal Lobar Degeneration : Validation of Pedigree Classification Criteria in a Northern Italy Cohort.” *Journal of Alzheimer’s Disease* 61 (2): 170661. https://doi.org/10.3233/JAD-170661.

Frank, Samuel. 2014. “Treatment of Huntington’s Disease.” *Neurotherapeutics* 11 (1): 153–60. https://doi.org/10.1007/s13311-013-0244-z.

Frank, Steven A. 2010. “Somatic Evolutionary Genomics: Mutations during Development Cause Highly Variable Genetic Mosaicism with Risk of Cancer and Neurodegeneration.” *PNAS* 107 (1): 1725–30. https://doi.org/10.1073/pnas.0909343106.

Frost B., Diamond M. I. 2010. “Prion-like Mechanisms in Neurodegenerative Diseases.” *Nat Rev Neurosci* 11 (3): 155–59. https://doi.org/10.1038/nrn2786.Prion-like.

Gabryelewicz, Tomasz, Mario Masellis, Mariusz Berdynski, Juan M. Bilbao, Ekaterina Rogaeva, Peter St. George-Hyslop, Anna Barczak, et al. 2010. “Intra-Familial Clinical Heterogeneity Due to FTLD-U with TDP-43 Proteinopathy Caused by a Novel Deletion in Progranulin Gene (PGRN).” *Journal of Alzheimer’s Disease* 22 (4): 1123–33. https://doi.org/10.3233/JAD-2010-101413.

Gawad, Charles, Winston Koh, and Stephen R. Quake. 2016. “Single-Cell Genome Sequencing: Current State of the Science.” *Nature Reviews Genetics* 17 (3): 175–88. https://doi.org/10.1038/nrg.2015.16.

Ghetti, Bernardino, Adrian L. Oblak, Bradley F. Boeve, Keith A. Johnson, Bradford C. Dickerson, and Michel Goedert. 2015. “Invited Review: Frontotemporal Dementia Caused by Microtubule-Associated Protein Tau Gene (MAPT) Mutations: A Chameleon for Neuropathology and Neuroimaging.” *Neuropathology and Applied Neurobiology* 41 (1): 24–46. https://doi.org/10.1111/nan.12213.

Gijselinck, Ilse, Marc Cruts, and Christine Van Broeckhoven. 2018. “The Genetics of C9orf72 Expansions.” *Cold Spring Harbor Perspectives in Medicine* 8 (4): 1–16. https://doi.org/10.1101/cshperspect.a026757.

Gleichman, Amy J., and S. Thomas Carmichael. 2020. “Glia in Neurodegeneration: Drivers of Disease or along for the Ride?” *Neurobiology of Disease* 142 (December 2019): 104957. https://doi.org/10.1016/j.nbd.2020.104957.

Goldman, Jill, and Vivianna Deerlin. 2018. “Alzheimer’s Disease and Frontotemporal Dementia: The Current State of Genetics and Genetic Testing Since the Advent of Next-

Generation Sequencing." *Mol Diagn Ther.* 22 (5): 503–13. https://doi.org/10.1007/s40291-018-0347-7.Alzheimer.

Grazina, Manuela, Filipe Silva, Isabel Santana, Beatriz Santiago, Cândida Mendes, Marta Simões, Miguel Oliveira, Luís Cunha, and Catarina Oliveira. 2004. "Frontotemporal Dementia and Mitochondrial DNA Transitions." *Neurobiology of Disease* 15 (2): 306–11. https://doi.org/10.1016/j.nbd.2003.11.004.

Greaves, Caroline V., and Jonathan D. Rohrer. 2019. "An Update on Genetic Frontotemporal Dementia." *Journal of Neurology* 266 (8): 2075–86. https://doi.org/10.1007/s00415-019-09363-4.

Guo, Hongshan, Ping Zhu, Xinglong Wu, Xianlong Li, Lu Wen, and Fuchou Tang. 2013. "Single-Cell Methylome Landscapes of Mouse Embryonic Stem Cells and Early Embryos Analyzed Using Reduced Representation Bisulfite Sequencing." *Genome Research* 23 (12): 2126–35. https://doi.org/10.1101/gr.161679.113.

Gupta, Pushpendra K. 2008. "Single-Molecule DNA Sequencing Technologies for Future Genomics Research." *Trends in Biotechnology* 26 (11): 602–11. https://doi.org/10.1016/j.tibtech.2008.07.003.

Hansen, D., H. Ling, T. Lashley, J. L. Holton, and T. T. Warner. 2019. "Review: Clinical, Neuropathological and Genetic Features of Lewy Body Dementias." *Neuropathology and Applied Neurobiology* 45 (7): 635–54. https://doi.org/10.1111/nan.12554.

Hardiman, Orla, Ammar Al-Chalabi, Carol Brayne, Ettore Beghi, Leonard H. Van Den Berg, Adriano Chio, Sarah Martin, Giancarlo Logroscino, and James Rooney. 2017. "The Changing Picture of Amyotrophic Lateral Sclerosis: Lessons from European Registers." *Journal of Neurology, Neurosurgery and Psychiatry* 88 (7): 557–63. https://doi.org/10.1136/jnnp-2016-314495.

Harding, Olivia, Chantell S. Evans, Junqiang Ye, Jonah Cheung, Tom Maniatis, and Erika L.F. Holzbaur. 2021. "ALS- And FTD-Associated Missense Mutations in TBK1 Differentially Disrupt Mitophagy." *Proceedings of the National Academy of Sciences of the United States of America* 118 (24). https://doi.org/10.1073/pnas.2025053118.

Hardy, John. 2005. “A Genetic Outline of the Pathways to Cell Death in Alzheimer’s Disease, Parkinson’s Disease, Frontal Dementias and Related Disorders.” *Neurodegenerative Diseases: Neurobiology, Pathogenesis and Therapeutics*, 222–26. https://doi.org/10.1017/CBO9780511544873.018.

Hartl, Daniel L, and M Ruvolo. 2012. *Genetic Analysis of Genes and Genomes*. Jones & Bartlett Learning, Burlington.

Hausmann, Stéphane, Dominique Garcin, Christophe Delenda, and Daniel Kolakofsky. 1999. “The Versatility of Paramyxovirus RNA Polymerase Stuttering.” *Journal of Virology* 73 (7): 5568–76. https://doi.org/10.1128/jvi.73.7.5568-5576.1999.

Heerema, Esther. 2020. “Factors That Increase Your Risk of Lewy Body Dementia.” VeryWell Health. 2020. https://www.verywellhealth.com/lewy-body-dementia-risks-98761.

Herpen, Esther Van, Sonia M. Rosso, Lies Anne Serverijnen, Hirotaka Yoshida, Guido Breedveld, Raoul Van De Graaf, Wouter Kamphorst, et al. 2003. “Variable Phenotypic Expression and Extensive Tau Pathology in Two Families with the Novel Tau Mutation L315R.” *Annals of Neurology* 54 (5): 573–81. https://doi.org/10.1002/ana.10721.

Hershey, Linda A., and Rhonda Coleman-Jackson. 2019. “Pharmacological Management of Dementia with Lewy Bodies.” *Drugs and Aging* 36 (4): 309–19. https://doi.org/10.1007/s40266-018-00636-7.

Hobson, Esther V., and Christopher J. McDermott. 2016. “Supportive and Symptomatic Management of Amyotrophic Lateral Sclerosis.” *Nature Reviews Neurology* 12 (9): 526–38. https://doi.org/10.1038/nrneurol.2016.111.

Holm, Ida Elisabeth, Adrian M. Isaacs, and Ian R.A. MacKenzie. 2009. “Absence of FUS-Immunoreactive Pathology in Frontotemporal Dementia Linked to Chromosome 3 (FTD-3) Caused by Mutation in the CHMP2B Gene.” *Acta Neuropathologica* 118 (5): 719–20. https://doi.org/10.1007/s00401-009-0593-1.

Hop, Paul J., Ramona A.J. Zwamborn, Eilis Hannon, Gemma L. Shireby, Marta F. Nabais, Emma M. Walker, Wouter van Rheenen, et al. 2022. “Genome-Wide Study of DNA Methylation Shows Alterations in Metabolic, Inflammatory, and Cholesterol Pathways in

ALS.” *Science Translational Medicine* 14 (633). https://doi.org/10.1126/scitranslmed.abj0264.

Huang, Yadong, and Lennart Mucke. 2012. “Alzheimer Mechanisms and Therapeutic Strategies.” *Cell* 148 (6): 1204–22. https://doi.org/10.1016/j.cell.2012.02.040.

Huey, Edward D., Raffaele Ferrari, Jorge H. Moreno, Christopher Jensen, Christopher M. Morris, Felix Potocnik, Rajesh N. Kalaria, et al. 2012. “FUS and TDP43 Genetic Variability in FTD and CBS.” *Neurobiology of Aging* 33 (5): 1016.e9-1016.e17. https://doi.org/10.1016/j.neurobiolaging.2011.08.004.

Hugo, Peter, Lodewijk Krijger, and Wouter De Laat. 2016. “Regulation of Disease-Associated Gene Expression in the 3D Genome.” *Nature Publishing Group* 17 (12): 771–82. https://doi.org/10.1038/nrm.2016.138.

Hyeon, Jae Wook, Albert H Kim, and Hiroko Yano. 2022. “Epigenetic Regulation in Huntington’s Disease.” *Neurochem. Int.* 148: 1–38. https://doi.org/10.1016/j.neuint.2021.105074.Epigenetic.

Illarioshkin, S. N., S. A. Klyushnikov, V. A. Vigont, Yu A. Seliverstov, and E. V. Kaznacheyeva. 2018. “Molecular Pathogenesis in Huntington’s Disease.” *Biochemistry (Moscow)* 83 (9): 1030–39. https://doi.org/10.1134/S0006297918090043.

Isaacs, A M, P Johannsen, I Holm, J E Nielsen, and Freja Consortium. 2011. “Frontotemporal Dementia Caused by CHMP2B Mutations.” *Current Alzheimer Research* 8 (3): 246–51.

Jan H Lui, David V Hansen, and Arnold R Kriegstein. 2011. “Development and Evolution of the Human Neocortex.” *Cell* 146 (1): 18–36. https://doi.org/10.1016/j.cell.2011.06.030.Lui.

Jankovic, Joseph, and Eng King Tan. 2020. “Parkinson’s Disease: Etiopathogenesis and Treatment.” *Journal of Neurology, Neurosurgery and Psychiatry* 91 (8): 795–808. https://doi.org/10.1136/jnnp-2019-322338.

Jaunmuktane, Z., and S. Brandner. 2020. “Invited Review: The Role of Prion-like Mechanisms in Neurodegenerative Diseases.” *Neuropathology and Applied Neurobiology* 46 (6): 522–45. https://doi.org/10.1111/nan.12592.

Jiang, Yaling, Bin Jiao, Xuewen Xiao, and Lu Shen. 2021. “Genetics of Frontotemporal Dementia in China.” *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 22 (5): 19–22.

Jiao, Bin, Hui Liu, Lina Guo, Xuewen Xiao, Xinxin Liao, Yafang Zhou, Ling Weng, Lu Zhou, and Xin Wang. 2021. “The Role of Genetics in Neurodegenerative Dementia : A Large Cohort Study in South China.” *Npj Genomic Medicine* 6 (69). https://doi.org/10.1038/s41525-021-00235-3.

JNPD Research. 2019. “What Is Neurodegenerative Disease?” EU Joint Programme - Neurodegenerative Disease Research. 2019. https://www.neurodegenerationresearch.eu/what/.

Jo, Myungjin, Shinrye Lee, Yu Mi Jeon, Seyeon Kim, Younghwi Kwon, and Hyung Jun Kim. 2020. “The Role of TDP-43 Propagation in Neurodegenerative Diseases: Integrating Insights from Clinical and Experimental Studies.” *Experimental and Molecular Medicine* 52 (10): 1652–62. https://doi.org/10.1038/s12276-020-00513-7.

Johnson, Julene K, Janine Diehl, Mario F Mendez, John Neuhaus, Jill S Shapira, and Mark Forman. 2005. “Frontotemporal Lobar Degeneration: Demographic Characteristics of 353 Patients.” *Arch Neurol.* 62: 925–30. https://doi.org/10.1016/B978-0-323-85654-6.00041-1.

Kennedy, Scott R., Lawrence A. Loeb, and Alan J. Herr. 2012. “Somatic Mutations in Aging, Cancer and Neurodegeneration.” *Mechanisms of Ageing and Development* 133 (4): 118–26. https://doi.org/10.1016/j.mad.2011.10.009.

Kennedy, Scott R., Michael W. Schmitt, Edward J. Fox, Brendan F. Kohrn, Jesse J. Salk, Eun Hyun Ahn, Marc J. Prindle, et al. 2014. “Detecting Ultralow-Frequency Mutations by Duplex Sequencing (Loeb Protocol).” *Nature Protocols* 9 (11): 2586–2606. https://doi.org/10.1093/bib/bbx048.

Khalfallah, Yousra, Rachel Kuta, Camille Grasmuck, Alexandre Prat, Heather D. Durham, and Christine Vande Velde. 2018. “TDP-43 Regulation of Stress Granule Dynamics in Neurodegenerative Disease-Relevant Cell Types /631/80/304 /631/378/87 /13/1 /13/31 /13/51 /13/109 /13/106 /13/89 /14/19 /14/32 /82/80 Article.” *Scientific Reports* 8 (1): 1–13. https://doi.org/10.1038/s41598-018-25767-0.

Kirshner, Howard, and Jasvinder Chawla. 2018. “What Is the Global Prevalence of Frontotemporal Dementia ?” Medscape. 2018. https://www.medscape.com/answers/1135164-119293/what-is-the-global-prevalence-of-frontotemporal-dementia-ftd.

Koriath, C., J. Kenny, G. Adamson, R. Druyeh, W. Taylor, J. Beck, L. Quinn, et al. 2020. “Predictors for a Dementia Gene Mutation Based on Gene-Panel next-Generation Sequencing of a Large Dementia Referral Series.” *Molecular Psychiatry* 25 (12): 3399–3412. https://doi.org/10.1038/s41380-018-0224-0.

Kovačević, Nives. 2016. “Magnetic Beads Based Nucleic Acid Purification for Molecular Biology Applications,” 53–67. https://doi.org/10.1007/978-1-4939-3185-9_5.

Lake, Blue B, Rizi Ai, Gwendolyn E Kaeser, Neeraj S Salathia, Yun C Yung, Rui Liu, Andre Wildberg, et al. 2016. “Neuronal Subtypes and Diversity Revealed by Single-Nucleus RNA Sequencing of the Human Brain.” *Science Neurogenomics* 352 (6293): 352–57.

Landin-Romero, Ramon, Rachel Tan, John R. Hodges, and Fiona Kumfor. 2016. “An Update on Semantic Dementia: Genetics, Imaging, and Pathology.” *Alzheimer’s Research and Therapy* 8 (1): 1–9. https://doi.org/10.1186/s13195-016-0219-5.

Larner, Andrew J., and Mark Doran. 2009. “Genotype-Phenotype Relationships of Presenilin-1 Mutations in Alzheimer’s Disease: An Update.” *Journal of Alzheimer’s Disease* 17 (2): 259–65. https://doi.org/10.3233/JAD-2009-1042.

Lau, Lonneke M L de, and Monique M B Breteler. 2006. “Epidemiology of Parkinson’s Disease in Africa.” *Lancet Neurol* 5 (June): 525–35. https://doi.org/10.1016/j.neurol.2015.09.012.

Lebert, Florence, Florence Pasquier, Lydie Souliez, and Henri Petit. 1998. “Frontotemporal Behavioral Scale.” *Alzheimer Disease & Associated Disorders* 12 (4): 335–39.

Lee, Gloria, and Chad J. Leugers. 2012. “Tau and Tauopathies.” *Progress in Molecular Biology and Translational Science* 107: 263–93. https://doi.org/10.1016/B978-0-12-385883-2.00004-7.

Lee, V. M.Y., M. Goedert, and J. Q. Trojanowski. 2001. “Neurodegenerative Tauopathies.” *Annual Review of Neuroscience* 24: 1121–59. https://doi.org/10.1146/annurev.neuro.24.1.1121.

Leija-Salazar, M., C. Piette, and C. Proukakis. 2018. “Somatic Mutations in Neurodegeneration.” *Neuropathology and Applied Neurobiology* 44 (3): 267–85. https://doi.org/10.1111/nan.12465.

Lewis, Charles A., Jesse Crayle, Shuntai Zhou, Ronald Swanstrom, and Richard Wolfenden. 2016. “Cytosine Deamination and the Precipitous Decline of Spontaneous Mutation during Earth’s History.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (29): 8194–99. https://doi.org/10.1073/pnas.1607580113.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Yun, Jason A. Chen, Renee L. Sears, Fuying Gao, Eric D. Klein, Anna Karydas, Michael D. Geschwind, et al. 2014. “An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy.” *PLoS Genetics* 10 (3). https://doi.org/10.1371/journal.pgen.1004211.

Lin, Brian C., Trong H. Phung, Nicole R. Higgins, Jessie E. Greenslade, Miguel A. Prado, Daniel Finley, Mariusz Karbowski, Brian M. Polster, and Mervyn J. Monteiro. 2021. “ALS/FTD Mutations in UBQLN2 Are Linked to Mitochondrial Dysfunction through Loss-of-Function in Mitochondrial Protein Import.” *Human Molecular Genetics* 30 (13): 1230–46. https://doi.org/10.1093/hmg/ddab116.

Lin, Michael T., and M. Flint Beal. 2006. “Mitochondrial Dysfunction and Oxidative Stress in Neurodegenerative Diseases.” *Nature Review* 443 (10): 787–95.

Linard, Morgane, Alix Ravier, Louisa Mougué, Iris Grgurina, Anne Laurence Boutillier, Alexandra Foubert-Samier, Frédéric Blanc, and Catherine Helmer. 2022. “Infectious Agents as Potential Drivers of α-Synucleinopathies.” *Movement Disorders* 37 (3): 464–77. https://doi.org/10.1002/mds.28925.

Liscic, Rajka Maria, Antonella Alberici, Nigel John Cairns, Maurizio Romano, and Emanuele Buratti. 2020. “From Basic Research to the Clinic: Innovative Therapies for ALS and FTD in the Pipeline.” *Molecular Neurodegeneration* 15 (1): 1–17. https://doi.org/10.1186/s13024-020-00373-9.

Liu, Elaine Y., Jenny Russ, Kathryn Wu, Donald Neal, Eunran Suh, Anna G. McNally, David J. Irwin, Vivianna M. Van Deerlin, and Edward B. Lee. 2014. “C9orf72 Hypermethylation Protects against Repeat Expansion-Associated Pathology in ALS/FTD.” *Acta Neuropathologica* 128 (4): 525–41. https://doi.org/10.1007/s00401-014-1286-y.

Liu, Lin, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law. 2012. “Comparison of Next-Generation Sequencing Systems.” *Journal of Biomedicine and Biotechnology* 2012. https://doi.org/10.1155/2012/251364.

Liu, Xiaolei, Bin Jiao, and Lu Shen. 2018. “The Epigenetics of Alzheimer’s Disease: Factors and Therapeutic Implications.” *Frontiers in Genetics* 9 (November): 1–10. https://doi.org/10.3389/fgene.2018.00579.

Lodato, Michael A, Rachel E Rodin, Craig L Bohrson, Michael E Coulter, Alison R Barton, Minseok Kwon, Maxwell A Sherman, et al. 2018. “Aging and Neurodegeneration Are Associated with Increased Mutations in Single Human Neurons.” *Science* 359 (6375): 555–59. https://doi.org/10.1126/science.aao4426.Aging.

Lynch, T., M. Sano, K. S. Marder, K. L. Bell, N. L. Foster, R. F. Defending, A. A.F. Sima, et al. 1994. “Clinical Characteristics of a Family with Chromosome 17-Linked Disinhibition-Dementia-Parkinsonism-Amyotrophy Complex.” *Neurology* 44 (10): 1878–84. https://doi.org/10.1212/wnl.44.10.1878.

Mackenzie, Ian R.A., and Manuela Neumann. 2016. “Molecular Neuropathology of Frontotemporal Dementia: Insights into Disease Mechanisms from Postmortem Studies.” *Journal of Neurochemistry* 138: 54–70. https://doi.org/10.1111/jnc.13588.

Mackenzie, Ian R A, Manuela Neumann, Atik Baborie, Deepak M Sampathu, Daniel Du Plessis, Evelyn Jaros, Robert H Perry, John Q Trojanowski, David M A Mann, and Virginia M Y Lee. 2011. “A Harmonized Classification System for FTLD-TDP Pathology.” *Acta Neuropathologica* 122 (1): 111–13. https://doi.org/10.1038/jid.2014.371.

Magalhães, João Pedro de, Caleb E. Finch, and Georges Janssens. 2010. “Next-Generation Sequencing in Aging Research: Emerging Applications, Problems, Pitfalls and Possible Solutions.” *Ageing Research Reviews* 9 (3): 315–23. https://doi.org/10.1016/j.arr.2009.10.006.

Manno, Gioele La, Daniel Gyllborg, Simone Codeluppi, Roger A Barker, Ernest Arenas, and Sten Linnarsson. 2016. “Molecular Diversity of Midbrain Development in Resource Molecular Diversity of Midbrain Development in Mouse , Human , and Stem Cells.” *Cell* 167 (October): 566–80. https://doi.org/10.1016/j.cell.2016.09.027.

Manolio, Teri A, Francis S Collins, Nancy J Cox, David B Goldstein, Lucia A Hindorff, David J Hunter, Mark I McCarthy, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461 (7256): 747–53. https://doi.org/10.1038/nature08494.Finding.

Mariani, Louise Laure, Christelle Tesson, Perrine Charles, Cile Cazeneuve, Valie Hahn, Katia Youssov, Leorah Freeman, et al. 2016. “Expanding the Spectrum of Genes Involved in Huntington Disease Using a Combined Clinical and Genetic Approach.” *JAMA Neurology* 73 (9): 1105–14. https://doi.org/10.1001/jamaneurol.2016.2215.

Maricic, Tomislav, Mark Whitten, and Svante Pääbo. 2010. “Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products.” *PLoS ONE* 5 (11): 9–13. https://doi.org/10.1371/journal.pone.0014004.

Marisa, Eliana, Christos Koros, Deepika Reddy, Victoria Van Berlo, Christos Kroupis, Kevin Wojta, Qing Wang, et al. 2019. “Frontotemporal Dementia Spectrum : First Genetic Screen in a Greek Cohort.” *Neurobiology of Aging* 75: 224.e1-224.e8. https://doi.org/10.1016/j.neurobiolaging.2018.10.029.

Marrone, Lara, Hannes C.A. Drexler, Jie Wang, Priyanka Tripathi, Tania Distler, Patrick Heisterkamp, Eric Nathaniel Anderson, et al. 2019. “FUS Pathology in ALS Is Linked to Alterations in Multiple ALS-Associated Proteins and Rescued by Drugs Stimulating Autophagy.” *Acta Neuropathologica* 138 (1): 67–84. https://doi.org/10.1007/s00401-019-01998-x.

Masrori, P., and P. Van Damme. 2020. “Amyotrophic Lateral Sclerosis: A Clinical Review.” *European Journal of Neurology* 27 (10): 1918–29. https://doi.org/10.1111/ene.14393.

Maynard, Scott, Evandro Fei Fang, Morten Scheibye-Knudsen, Deborah L Croteau, and Vilhelm A Bohr. 2015. “DNA Damage, DNA Repair, Aging, and Neurodegeneration.” *CSH Perspectives in Medicine* September (5): a025130. 10.1101/cshperspect.a025130.

McDermott, C. J., K. White, K. Bushby, and P. J. Shaw. 2000. “Hereditary Spastic Paraparesis: A Review of New Developments.” *Journal of Neurology Neurosurgery and Psychiatry* 69 (2): 150–60. https://doi.org/10.1136/jnnp.69.2.150.

McKeith, Ian. 2004. “Dementia with Lewy Bodies.” *Dialogues in Clinical Neuroscience* 6 (34): 333–41. https://doi.org/110.31887/DCNS.2004.6.3/imckeith.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew M. Kernytsky, Kiran V. Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.20.

MedlinePlus Genetics. 2020. “TARDBP Gene.” National Institutes of Health (NIH). 2020. https://medlineplus.gov/genetics/gene/tardbp/.

Mendez, Mario F. 2009. “Frontotemporal Dementia: Therapeutic Interventions.” *Frontiers of Neurology and Neuroscience* 24: 168–78. https://doi.org/10.1159/000197896.

Miller, Robert G., J. D. Mitchell, M. Lyon, and D. H. Moore. 2007. “Riluzole for Amyotrophic Lateral Sclerosis (ALS)/Motor Neuron Disease (MND).” *Cochrane Database of Systematic Reviews*, no. 1. https://doi.org/10.1002/14651858.CD001447.pub2.

Miller, Ted. 2017. “Global, Regional, and National Burden of Neurological Disorders during 1990-2015: A Systematic Analysis for the Global Burden of Disease Study 2015.” *The Lancet Neurology* xx (xx): 459–80. https://doi.org/10.17863/CAM.20529.

Min, Young Gi, Seok Jin Choi, Yoon Ho Hong, Sung Min Kim, Je Young Shin, and Jung Joon Sung. 2020. “Dissociated Leg Muscle Atrophy in Amyotrophic Lateral Sclerosis/Motor Neuron Disease: The ‘Split-Leg’ Sign.” *Scientific Reports* 10 (1): 1–8. https://doi.org/10.1038/s41598-020-72887-7.

Mol, Merel O, Jeroen G J Van Rooij, Tsz H Wong, Shamiram Melhem, Annemieke J M H Verkerk, Anneke J A Kievit, Rick Van Minkelen, et al. 2021. “Underlying Genetic Variation in Familial Frontotemporal Dementia : Sequencing of 198 Patients.” *Neurobiology of Aging* 97: 148.e9-148.e16. https://doi.org/10.1016/j.neurobiolaging.2020.07.014.

Moore, Katrina M., Jennifer Nicholas, Murray Grossman, Corey T. McMillan, David J. Irwin, Lauren Massimo, Vivianna M. Van Deerlin, et al. 2020. “Age at Symptom Onset and Death and Disease Duration in Genetic Frontotemporal Dementia: An International Retrospective Cohort Study.” *The Lancet Neurology* 19 (2): 145–56. https://doi.org/10.1016/S1474-4422(19)30394-1.

Mossevelde, Sara Van, Julie Van Der Zee, Ilse Gijselinck, Sebastiaan Engelborghs, Anne Sieben, Tim Van Langenhove, Jan De Bleecker, et al. 2016. “Clinical Features of TBK1 Carriers Compared with C9orf72, GRN and Non-Mutation Carriers in a Belgian Cohort.” *Brain* 139 (2): 452–67. https://doi.org/10.1093/brain/awv358.

Mudher, Amrit, Morvane Colin, Simon Dujardin, Miguel Medina, Ilse Dewachter, Seyedeh Maryam Alavi Naini, Eva Maria Mandelkow, et al. 2017. “What Is the Evidence That Tau Pathology Spreads through Prion-like Propagation?” *Acta Neuropathologica Communications* 5 (1): 99. https://doi.org/10.1186/s40478-017-0488-7.

Murphy, Natalie A., Karissa C. Arthur, Pentti J. Tienari, Henry Houlden, Adriano Chiò, and Bryan J. Traynor. 2017. “Age-Related Penetrance of the C9orf72 Repeat Expansion.” *Scientific Reports* 7 (1): 1–7. https://doi.org/10.1038/s41598-017-02364-1.

Mustafin, R. N., and E. K. Khusnutdinova. 2020. “Involvement of Transposable Elements in Neurogenesis.” *Vavilovskii Zhurnal Genetiki i Selektsii* 24 (2): 209–18. https://doi.org/10.18699/VJ20.613.

Nagai, Makiko, Diane B. Re, Tetsuya Nagata, Alcmène Chalazonitis, Thomas M. Jessell, Hynek Wichterle, and Serge Przedborski. 2007. “Astrocytes Expressing ALS-Linked Mutated SOD1 Release Factors Selectively Toxic to Motor Neurons.” *Nature Neuroscience* 10 (5): 615–22. https://doi.org/10.1038/nn1876.

National Organization for Rare Disorders (NORD). 2019. “Rare Disease Information.” 2019. https://rarediseases.org/for-patients-and-families/information-resources/rare-disease-information/.

Navarro, Claire L., Pierre Cau, and Nicolas Lévy. 2006. “Molecular Bases of Progeroid Syndromes.” *Human Molecular Genetics* 15 (SUPPL. 2): 151–61. https://doi.org/10.1093/hmg/ddl214.

Nawy, Tal. 2014. “Single-Cell Sequencing.” *Nature Methods* 11 (1): 18. https://doi.org/10.1038/nmeth.2771.

Neumann, Manuela, Deepak M. Sampathu, Linda K. Kwong, Adam C. Truax, Matthew C. Micsenyi, Thomas T. Chou, Jennifer Bruce, et al. 2006. “Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis.” *Science* 314 (5796): 130–33. https://doi.org/10.1126/science.1134108.

Nguyen, Andrew D, Thi A Nguyen, Lauren Herl Martens, Laura L Mitic, and Robert V Farese Jr. 2013. “Progranulin: At the Interface of Neurodegenerative and Metabolic Diseases.” *Trends in Endocrinology & Metabolism* 24 (12): 597–606. https://doi.org/10.1016/j.tem.2013.08.003.

Nicolas, Gaël, and Joris A. Veltman. 2018. “The Role of de Novo Mutations in Adult-Onset Neurodegenerative Disorders (AOND).” *Acta Neuropathologica*, no. 0123456789. https://doi.org/10.1007/s00401-018-1939-3.

Nie, Yu, Alexander Murley, Zoe Golder, James B. Rowe, Kieren Allinson, and Patrick F. Chinnery. 2022. “Heteroplasmic Mitochondrial DNA Mutations in Frontotemporal Lobar

Degeneration." *Acta Neuropathologica* 143 (6): 687–95. https://doi.org/10.1007/s00401-022-02423-6.

Niedringhaus, Thomas P., Denitsa Milanova, Matthew B. Kerby, Michael P. Snyder, and Annelise E. Barron. 2011. "Landscape of Next-Generation Sequencing Technologies." *Analytical Chemistry* 83 (12): 4327–41. https://doi.org/10.1021/ac2010857.

Olszewska, Diana A., Roisin Lonergan, Emer M. Fallon, and Tim Lynch. 2016. "Genetics of Frontotemporal Dementia." *Current Neurology and Neuroscience Reports* 16 (12): 707 (15). https://doi.org/10.1007/s11910-016-0707-9.

Olszewska, Diana Angelika, E. M. Fallon, G. M. Pastores, K. Murphy, A. Blanco, T. Lynch, and S. M. Murphy. 2019. "Autosomal Dominant Gene Negative Frontotemporal Dementia-Think of SCA17." *Cerebellum* 18 (3): 654–58. https://doi.org/10.1007/s12311-018-0998-2.

Ozkan, Ekin, and Marcelo P. Lacerda. 2021. "Genetics, Cytogenetic Testing And Conventional Karyotype." *NCBI Bookshelf*, 6–11. http://www.ncbi.nlm.nih.gov/pubmed/33085440.

Pan, Xiao dong, and Xiao chun Chen. 2013. "Clinic, Neuropathology and Molecular Genetics of Frontotemporal Dementia: A Mini-Review." *Translational Neurodegeneration* 2 (1): 0–8. https://doi.org/10.1186/2047-9158-2-8.

Papapetropoulos, Spiridon, Nikhil Adi, John Ellul, Andreas A. Argyriou, and Elisabeth Chroni. 2007. "A Prospective Study of Familial versus Sporadic Parkinson's Disease." *Neurodegenerative Diseases* 4 (6): 424–27. https://doi.org/10.1159/000107702.

Pavone, Piero, Agata Polizzi, Simona Domenica Marino, Giovanni Corsello, Raffaele Falsaperla, Silvia Marino, and Martino Ruggieri. 2020. "West Syndrome: A Comprehensive Review." *Neurological Sciences* 41 (12): 3547–62. https://doi.org/10.1007/s10072-020-04600-5.

"PCR Primer Design Tips - Behind the Bench." 2019. Thermo Fisher Scientific. 2019. https://www.thermofisher.com/blog/behindthebench/pcr-primer-design-tips/.

Polymenidou, Magdalini, Clotilde Lagier-tourenne, Kasey R Hutt, C Stephanie, Jacqueline Moran, Tiffany Y Liang, Shuo-chien Ling, et al. 2011. “Long Pre-MRNA Depletion and RNA Missplicing Contribute to Neuronal Vulnerability from Loss of TDP-43.” *Nature Neuroscience* 14 (4): 459–68. https://doi.org/10.1038/nn.2779.Long.

Pottier, Cyril, Evadnie Rampersaud, Matt Baker, Gang Wu, Joanne Wuu, L Jacob, Stephan Zuchner, et al. 2018. “ALS-FTD Patient from the CReATe Consortium : A Case Report.” *Amyotroph Lateral Scler Frontotemporal Degener.* 19 (5): 469–71. https://doi.org/10.1080/21678421.2018.1452947.Identification.

Pottier, Cyril, Thomas A. Ravenscroft, Monica Sanchez-Contreras, and Rosa Rademakers. 2016. “Genetics of FTLD: Overview and What Else We Can Expect from Genetic Studies.” *Journal of Neurochemistry* 138: 32–53. https://doi.org/10.1111/jnc.13622.

Pozzi, Silvia, Sai Sampath Thammisetty, Philippe Codron, Reza Rahimian, Karine Valérie Plourde, Geneviève Soucy, Christine Bareil, et al. 2019. “Virus-Mediated Delivery of Antibody Targeting TAR DNA-Binding Protein-43 Mitigates Associated Neuropathology.” *Journal of Clinical Investigation* 129 (4): 1581–95. https://doi.org/10.1172/JCI123931.

Pray, Leslie A. 2008. “Transposons: The Jumping Genes.” Nature Education, Vol. 1 (1). https://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518.

Purves D, Augustine G J, Fitzpatrick D. 2001. *Excitatory and Inhibitory Synaptic Potentials*. Neuroscien. Sinauer Associates Inc. https://www.ncbi.nlm.nih.gov/books/NBK11117/.

QIAamp® Fast DNA Tissue Kit Handbook. 2015. “QIAamp® Fast DNA Tissue Kit Handbook.” *QIAamp® Fast DNA Tissue Kit Handbook*, no. July: 1–20.

QIAamp ® Fast DNA Tissue Kit Safety Data Sheet. 2021. *QIAamp ® Fast DNA Tissue Kit*.

Quail, Michael A., Miriam Smith, Paul Coupland, Thomas D. Otto, Simon R. Harris, Thomas R. Connor, Anna Bertoni, Harold P. Swerdlow, and Yong Gu. 2012. “A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers.” *BMC Genomics* 13 (1): 1. https://doi.org/10.1186/1471-2164-13-341.

Radke, Josefine, Werner Stenzel, and Hans Hilmar Goebel. 2018. *Neurometabolic and Neurodegenerative Diseases in Children*. *Handbook of Clinical Neurology*. 1st ed. Vol. 145. Elsevier B.V. https://doi.org/10.1016/B978-0-12-802395-2.00009-2.

Raffaele, Ferrari, Manzoni Claudia, and Hardy John. 2019. "Genetics and Molecular Mechanisms of Frontotemporal Lobar Degeneration: An Update and Future Avenues." *Neurobiology of Aging* 78: 98–110. https://doi.org/10.1016/j.neurobiolaging.2019.02.006.

Ramos-campoy, Oscar, Anna Antonell, Neus Falgàs, Mircea Balasa, Lluís Armengol, Guadalupe Fernández-villullas, Beatriz Bosch, et al. 2020. "Screening of Dementia Genes by Whole-Exome Sequencing in Spanish Patients with Early-Onset Dementia : Likely Pathogenic , Uncertain Signi Fi Cance and Risk Variants." *Neurobiology of Aging* 93: e1–9. https://doi.org/10.1016/j.neurobiolaging.2020.02.008.

Ramos, Eliana Marisa, Deepika Reddy Dokuru, Victoria Van Berlo, Kevin Wojta, Qing Wang, Alden Y. Huang, Sandeep Deverasetty, et al. 2020. "Genetic Screening of a Large Series of North American Sporadic and Familial Frontotemporal Dementia Cases." *Alzheimer's and Dementia* 16 (1): 118–30. https://doi.org/10.1002/alz.12011.

Ramsey, Marilyn J., Dan H. Moore, Jane F. Briner, Denise A. Lee, Letha A. Olsen, Jami R. Senft, and James D. Tucker. 1995. "The Effects of Age and Lifestyle Factors on the Accumulation of Cytogenetic Damage as Measured by Chromosome Painting." *Mutation Research DNAging* 338 (1–6): 95–106. https://doi.org/10.1016/0921-8734(95)00015-X.

Ray Dorsey, E., Alexis Elbaz, Emma Nichols, Foad Abd-Allah, Ahmed Abdelalim, Jose C. Adsuar, Mustafa Geleto Ansha, et al. 2018. "Global, Regional, and National Burden of Parkinson's Disease, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016." *The Lancet Neurology* 17 (11): 939–53. https://doi.org/10.1016/S1474-4422(18)30295-3.

Reith, R. Michelle, Sharon Way, James McKenna, Katherine Haines, and Michael J. Gambello. 2011. "Loss of the Tuberous Sclerosis Complex Protein Tuberin Causes Purkinje Cell Degeneration." *Neurobiology of Disease* 43 (1): 113–22. https://doi.org/10.1016/j.nbd.2011.02.014.

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, and Julie Gastier-Foster. 2015. “Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association of Molecular Pathology.” *Genet Med* 17 (5): 405–24. https://doi.org/10.1111/aos.13972_502.

Risques, Rosa Ana, and Scott R. Kennedy. 2018. “Aging and the Rise of Somatic Cancer-Associated Mutations in Normal Tissues.” *PLoS Genetics* 14 (1): 1–12. https://doi.org/10.1371/journal.pgen.1007108.

Rocke, Zoe, and Mariya Belyayeva. 2022. “Subacute Sclerosing Panencephalitis.” *NCBI Bookshelf*. https://doi.org/10.1001/jama.1969.03160290111045.

Rohrback, Suzanne, Benjamin Siddoway, Christine S. Liu, and Jerold Chun. 2018. “Genomic Mosaicism in the Developing and Adult Brain.” *Developmental Neurobiology*. https://doi.org/10.1002/dneu.22626.

Rohrer, Jonathan D., Adrian M. Isaacs, Sarah Mizlienska, Simon Mead, Tammaryn Lashley, Selina Wray, Katie Sidle, et al. 2015. “C9orf72 Expansions in Frontotemporal Dementia and Amyotrophic Lateral Sclerosis.” *The Lancet Neurology* 14 (3): 291–301. https://doi.org/10.1016/S1474-4422(14)70233-9.

Roos, Peter, Ida Holm, Joergen Nielsen, Troels Nielsen, Jeremy Brown, Peter Johannsen, and Adrian Isaacs. 2020. “CHMP2B Frontotemporal Dementia.” *Gene Reviews*, 1–16.

Roos, Raymund A.C. 2010. “Huntington’s Disease: A Clinical Review.” *Orphanet Journal of Rare Diseases* 5 (40). https://doi.org/10.1186/1750-1172-5-40.

Rossum, G van, and F L Drake. 2009. “Python 3 Reference Manual.” Scotts Valley, CA, US. 2009. http://citebay.com/how-to-cite/python/.

Rouaux, Caroline, Natasa Jokic, Corinne Mbebi, Stephanie Boutillier, Jean Philippe Loeffler, and Anne Laurence Boutillier. 2003. “Critical Loss of CBP/P300 Histone Acetylase Activity by Caspase-6 during Neurodegeneration.” *EMBO Journal* 22 (24): 6537–49. https://doi.org/10.1093/emboj/cdg615.

RStudio Team. 2020. “RStudio: Integrated Development for R.” RStudio, PBC, Boston, MA, US. 2020. http://www.rstudio.com/.

Rubinsztein, David. 2006. “The Roles of Intracellular Protein-Degradation Pathways in Neurodegeneration.” *Nature* 443 (10): 780–86.

Ryu, Hoon, Junghee Lee, Sean W Hagerty, Byoung Yul Soh, Sara E Mcalpin, Kerry A Cormier, Karen M Smith, and Robert J Ferrante. 2006. “ESET/SETDB1 Gene Expression and Histone H3(K9) Trimethylation in Huntington’ s Disease.” *Gene* 103 (50): 19176–81.

Sama, Reddy R.anjith Kumar, Catherine L. Ward, and Daryl A. Bosco. 2014. “Functions of FUS/TLS from DNA Repair to Stress Response: Implications for ALS.” *American Society for Neurochemistry* 6 (4): 1–18. https://doi.org/10.1177/1759091414544472.

Saxowsky, Tina T, and Paul W Doetsch. 2006. “RNA Polymerase Encounters with DNA Damage: Transcription-Coupled Repair or Transcriptional Mutagenesis?” *Chem. Rev.* 106 (2): 474–88. https://doi.org/10.1021/cr040466q.

Schmitt, Michael W., Edward J. Fox, Marc J. Prindle, Kate S. Reid-Bayliss, Lawrence D. True, Jerald P. Radich, and Lawrence A. Loeb. 2015. “Sequencing Small Genomic Targets with High Efficiency and Extreme Accuracy.” *Nature Methods* 12 (5): 423–25. https://doi.org/10.1038/nmeth.3351.

Schmitt, Michael W, Scott R Kennedy, Jesse J Salk, Edward J Fox, Joseph B Hiatt, and Lawrence A Loeb. 2012. “Detection of Ultra-Rare Mutations by NGS.Pdf.” *PNAS* 109 (36): 14508–13. https://doi.org/10.1073/pnas.1208715109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1208715109.

She, Angela, Iren Kurtser, Surya A Reis, Krista Hennig, Jenny Lai, Audrey Lang, Wen-Ning Zhao, et al. 2017. “Selectivity and Kinetic Requirements of HDAC Inhibitors as Progranulin Enhancers for Treating Frontotemporal Dementia.” *Cell Chem Biol.* 24 (7): 892–906. https://doi.org/10.1016/j.chembiol.2017.06.010.Selectivity.

Sieben, Anne, Tim Van Langenhove, Sebastiaan Engelborghs, Jean Jacques Martin, Paul Boon, Patrick Cras, Peter Paul De Deyn, Patrick Santens, Christine Van Broeckhoven, and Marc Cruts. 2012. “The Genetics and Neuropathology of Frontotemporal Lobar

Degeneration." *Acta Neuropathologica* 124 (3): 353–72. https://doi.org/10.1007/s00401-012-1029-x.

Simone, Roberto, Rubika Balendra, Thomas G Moens, Elisavet Preza, Katherine M Wilson, Amanda Heslegrave, Nathan S Woodling, et al. 2018. " G-quadruplex-binding Small Molecules Ameliorate C9orf72 FTD / ALS Pathology in Vitro and in Vivo ." *EMBO Molecular Medicine* 10 (1): 22–31. https://doi.org/10.15252/emmm.201707850.

Singh, Anju, Ritushree Kukreti, Luciano Saso, and Shrikant Kukreti. 2019. "Oxidative Stress: A Key Modulator in Neurodegenerative Diseases." *Molecules* 24 (8): 1–20. https://doi.org/10.3390/molecules24081583.

Sloan, Daniel B., Amanda K. Broz, Joel Sharbrough, and Zhiqiang Wu. 2018. "Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods." *Trends in Biotechnology* 36 (7): 729–40. https://doi.org/10.1016/j.tibtech.2018.02.009.

Smith, CD, M Badadani, A Nalbandian, E Dec, J Vesa, S Donkervoort, B Martin, GD Watts, V Caiozzo, and V Kimonis. 2011. "Valosin-Containing Protein (VCP) Disease and Familial Alzheimer's Disease: Contrasts and Overlaps." *The Clinical Spectrum of Alzheimer's Disease -The Charge Toward Comprehensive Diagnostic and Therapeutic Strategies* 43: 2–9. https://doi.org/10.5772/18811.

Snowden, Julie S, David Neary, and David M A Mann. 2002. "Frontotemporal Dementia." *British Journal Psychiatry* 180: 140–43.

Stein, Richard A. 2019. "Single-Cell Sequencing Sifts through Multiple Omics." *Genetic Engineering and Biotechnology News*, 2019. https://doi.org/10.1089/gen.39.07.10.

Stephenson, Jodie, Erik Nutma, Paul van der Valk, and Sandra Amor. 2018. "Inflammation in CNS Neurodegenerative Diseases." *Immunology* 154 (2): 204–19. https://doi.org/10.1111/imm.12922.

Stevenson, Rebecca, Evgeniia Samokhina, Ilaria Rossetti, John W. Morley, and Yossi Buskila. 2020. "Neuromodulation of Glial Function During Neurodegeneration." *Frontiers in Cellular Neuroscience* 14 (August): 1–23. https://doi.org/10.3389/fncel.2020.00278.

Su, Joseph H., Kathryn E. Nichol, Tom Sitch, Philip Sheu, Charlie Chubb, Bruce L. Miller, Kevin J. Tomaselli, Ronald C. Kim, and Carl W. Cotman. 2000. “DNA Damage and Activated Caspase-3 Expression in Neurons and Astrocytes: Evidence for Apoptosis in Frontotemporal Dementia.” *Experimental Neurology* 163 (1): 9–19. https://doi.org/10.1006/exnr.2000.7340.

Swartz, J. Randolph, Bruce L. Miller, Ira M. Lesser, and Amy L. Darby. 1997. “Frontotemporal Dementia: Treatment Response to Serotonin Selective Reuptake Inhibitors.” *Journal of Clinical Psychiatry* 58 (5): 212–16. https://doi.org/10.4088/jcp.v58n0506.

Swieten, John Van, and Maria Grazia Spillantini. 2007. “Hereditary Frontotemporal Dementia Caused by Tau Gene Mutations.” *Brain Pathology* 17 (1): 63–73. https://doi.org/10.1111/j.1750-3639.2007.00052.x.

Takada, Leonel T. 2015. “The Genetics of Monogenic Frontotemporal Dementia.” *Dementia e Neuropsychologia* 9 (3): 219–29. https://doi.org/10.1590/1980-57642015DN93000003.

Tammela, Tuomas, and Julien Sage. 2020. “Investigating Tumor Heterogeneity in Mouse Models.” *Annual Review of Cancer Biology* 4: 99–119. https://doi.org/10.1146/annurev-cancerbio-030419-033413.

Tarasoff-Conway, Jenna M., Roxana O. Carare, Ricardo S. Osorio, Lidia Glodzik, Tracy Butler, Els Fieremans, Leon Axel, et al. 2015. “Clearance Systems in the Brain - Implications for Alzheimer Disease.” *Nature Reviews Neurology* 11 (8): 457–70. https://doi.org/10.1038/nrneurol.2015.119.

Tariq, Amber, Jia Bei Lin, Megan M. Noll, Mariana P. Torrente, Korrie L. Mack, Oscar Hernandez Murillo, Meredith E. Jackrel, and James Shorter. 2018. “Potentiating Hsp104 Activity via Phosphomimetic Mutations in the Middle Domain.” *FEMS Yeast Research* 18 (5): 1–14. https://doi.org/10.1093/femsyr/foy042.

Thermo Fisher Scientific. 2020. “Performance Comparison of Platinum SuperFi II and Platinum SuperFi DNA Polymerases.” https://assets.thermofisher.com/TFS-Assets/BID/Reference-Materials/platinum-superfi-ii-and-superfi-benchmarking-data.pdf.

Thompson, Leslie. 2008. “Neurodegeneration - A Question of Balance.” *Nature* 452 (4): 707–8. https://doi.org/10.1177/0308575919887545.

Tollervey, James R, Tomaž Curk, Boris Rogelj, Michael Briese, Melis Kayikci, Tibor Hortobágyi, Agnes L Nishimura, et al. 2011. “Characterising the RNA Targets and Position-Dependent Splicing Regulation by TDP-43 ; Implications for Neurodegenerative Diseases.” *Nat Neurosci.* 14 (4): 452–58. https://doi.org/10.1038/nn.2778.Characterising.

Torrente, Mariana P., Edward Chuang, Megan M. Noll, Meredith E. Jackrel, Michelle S. Go, and James Shorter. 2016. “Mechanistic Insights into Hsp104 Potentiation.” *Journal of Biological Chemistry* 291 (10): 5101–15. https://doi.org/10.1074/jbc.M115.707976.

Tracy, Tara E., Jesus Madero-Pérez, Danielle L. Swaney, Timothy S. Chang, Michelle Moritz, Csaba Konrad, Michael E. Ward, et al. 2022. “Tau Interactome Maps Synaptic and Mitochondrial Processes Associated with Neurodegeneration.” *Cell* 185 (4): 712-728.e14. https://doi.org/10.1016/j.cell.2021.12.041.

Tsuji, Shoji. 2012. “Dentatorubral–Pallidoluysian Atrophy.” *Handbook of Clinical Neurology* 103 (January): 587–94. https://doi.org/10.1016/B978-0-444-51892-7.00041-3.

Urbizu, Aintzane, and Katrin Beyer. 2020. “Epigenetics in Lewy Body Diseases: Impact on Gene Expression, Utility as a Biomarker, and Possibilities for Therapy.” *International Journal of Molecular Sciences* 21 (13): 1–31. https://doi.org/10.3390/ijms21134718.

V.V., Belzil, Gendron T.F., and Petrucelli L. 2013. “RNA-Mediated Toxicity in Neurodegenerative Disease.” *Molecular and Cellular Neuroscience* 56: 406–19. https://doi.org/10.1016/j.mcn.2012.12.006.RNA-mediated.

Verheijen, Bert M., Marc Vermulst, and Fred W. van Leeuwen. 2018. “Somatic Mutations in Neurons during Aging and Neurodegeneration.” *Acta Neuropathologica* 135 (6): 811–26. https://doi.org/10.1007/s00401-018-1850-y.

Viguera, Enrique, Danielle Canceill, and S. Dusko Ehrlich. 2001. “Replication Slippage Involves DNA Polymerase Pausing and Dissociation.” *EMBO Journal* 20 (10): 2587–95. https://doi.org/10.1093/emboj/20.10.2587.

Vila, Miquel, and Serge Przedborski. 2003. “Targeting Programmed Cell Death in Neurodegenerative Diseases.” *Nature Reviews Neuroscience* 4 (5): 1–11. https://doi.org/10.1038/nrn1100.

Vinicius, M, C De Mello, LC Vieira, L Cruz de Souza, KB Gomes, and M Carvalho. 2019. “Alzheimer’s Disease: Risk Factors and Potentially Protective Measures.” *Journal of Biomedical Science* 26 (33): 1–11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6507104/.

Vldal, Ruben, Blas Franglone, Agueda Rostagno, Simon Mead, Tamas Révészt, Gordon Plant, and Jorge Ghiso. 1999. “A Stop-Codon Mutation in the BRI Gene Associated with Familial British Dementia.” *Nature* 399 (6738): 776–81. https://doi.org/10.1038/21637.

Watts, Katrina E., Nicholas J. Storr, Phoebe G. Barr, and Anto P. Rajkumar. 2022. “Systematic Review of Pharmacological Interventions for People with Lewy Body Dementia.” *Aging and Mental Health*, 1–14. https://doi.org/10.1080/13607863.2022.2032601.

Wei, Wei, Michael J. Keogh, Juvid Aryaman, Zoe Golder, Peter J. Kullar, Ian Wilson, Kevin Talbot, et al. 2019. “Frequency and Signature of Somatic Variants in 1461 Human Brain Exomes.” *Genetics in Medicine* 21 (4): 904–12. https://doi.org/10.1038/s41436-018-0274-3.

Wei, Wei, Michael J. Keogh, Ian Wilson, Jonathan Coxhead, Sarah Ryan, Sara Rollinson, Helen Griffin, et al. 2017. “Mitochondrial DNA Point Mutations and Relative Copy Number in 1363 Disease and Control Human Brains.” *Acta Neuropathologica Communications* 5 (1): 13. https://doi.org/10.1186/s40478-016-0404-6.

Wells, Jonathan N, and Cedric Feshotte. 2020. “A Field Guide to Eukaryotic Transposable Elements.” *Annual Review of Genetics* 54 (11): 539–61. https://doi.org/10.1146/annurev-genet-040620-022145.A.

Woodworth, Mollie B, Kelly M Girskis, and Christopher A Walsh. 2017. “Building a Lineage from Single Cells: Genetic Techniques for Cell Lineage Tracking.” *Nat Rev Genet.* 18 (4): 230–44. https://doi.org/10.1038/nrg.2016.159.Building.

World Health Organization. 2008. “Neurological Disorders: Public Health Challenges (Chapter 2).” *World Health Organization - Annual Report*, 27–39. https://doi.org/10.1111/j.1741-1130.2007.00143.x.

Xi, Zhengrui, Ming Zhang, Amalia C. Bruni, Raffaele G. Maletta, Rosanna Colao, Pietro Fratta, James M. Polke, et al. 2015. “The C9orf72 Repeat Expansion Itself Is Methylated in ALS and FTLD Patients.” *Acta Neuropathologica* 129 (5): 715–27. https://doi.org/10.1007/s00401-015-1401-8.

Xi, Zhengrui, Lorne Zinman, Yakov Grinberg, Danielle Moreno, Christien Sato, and Juan M BilBao. 2012. “Investigation of C9orf72 in 4 Neurodegenerative Disorders.” *Arch Neurol.* 69 (12): 1583–90. https://doi.org/10.1001/archneurol.2012.2016.Investigation.

Zarei, Sara, Karen Carr, Luz Reiley, Kelvin Diaz, Orleiquis Guerra, Pablo Fernandez Altamirano, Wilfredo Pagani, Daud Lodin, Gloria Orozco, and Angel Chinea. 2015. “A Comprehensive Review of Amyotrophic Lateral Sclerosis.” *Surgical Neurology International* 6 (1). https://doi.org/10.4103/2152-7806.169561.

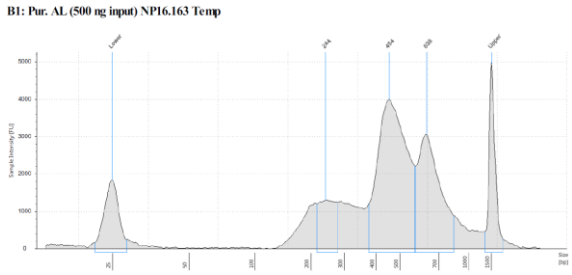
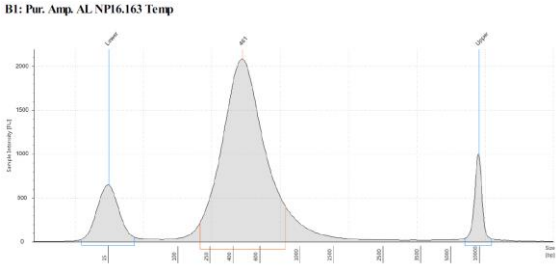
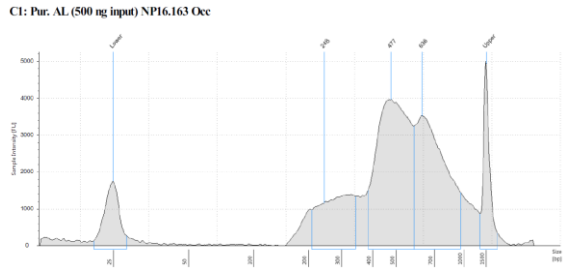
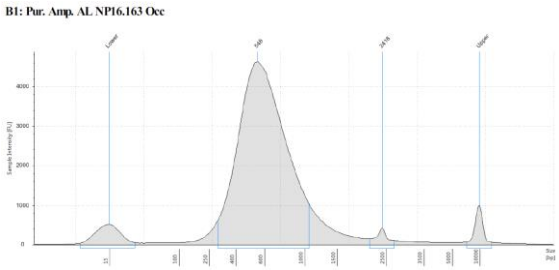
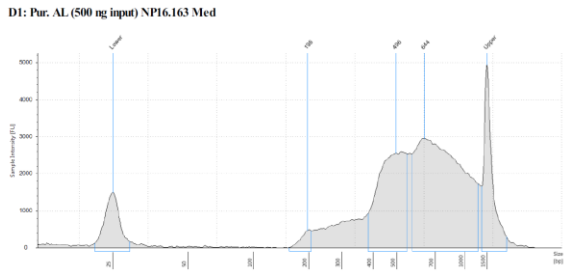
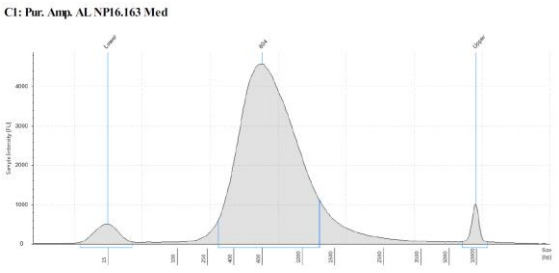
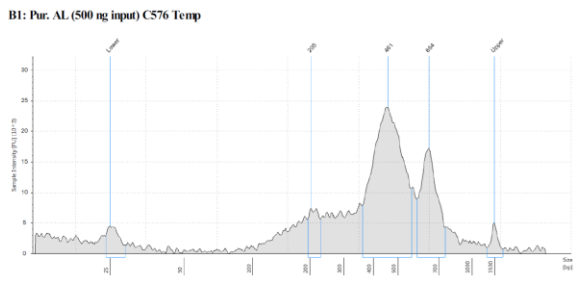
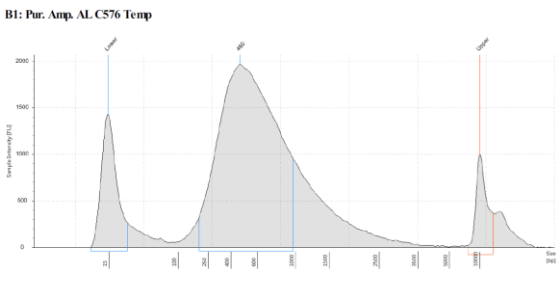
Zee, Julie Van Der, and Christine Van Broeckhoven. 2014. “Frontotemporal Lobar Degeneration — Building on Breakthroughs.” *Nature Rev. Neurol.* 10 (2): 70–72. https://doi.org/10.1038/nrneurol.2013.270.

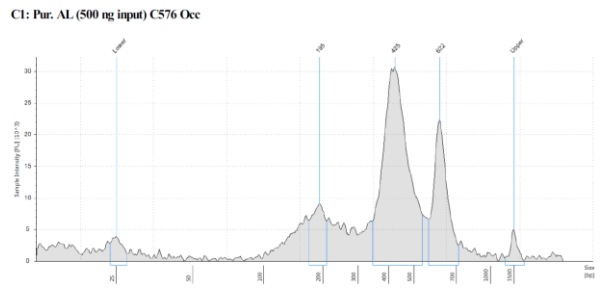
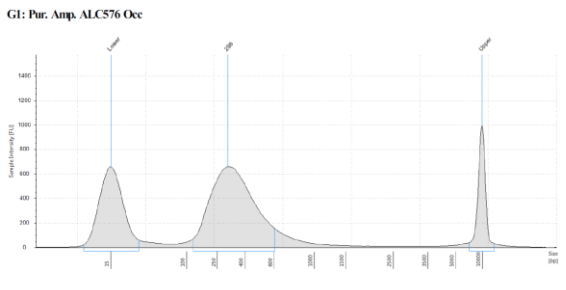
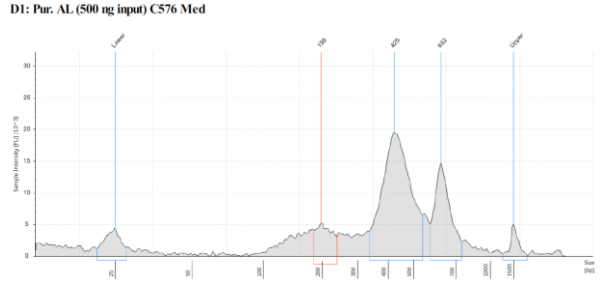
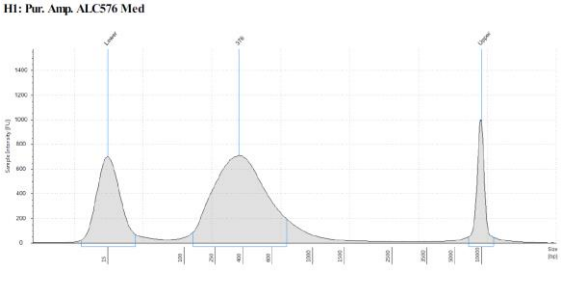
****

Appendix I

Double capture-based duplex library preparation Optimization ScreenTape results:

Protocol I (Pre-capture lib. prep.)

Post-ligation HS D1000 ScreenTape	Post-amplification HS D5000 ScreenTape (40 ng 9 cycles)
NP16.163 Temp	NP16.163 Temp
	
NP16.163 Occ	NP16.163 Occ
	
NP16.163 Med	NP16.163 Med
	
C576 Temp	C576 Temp
	

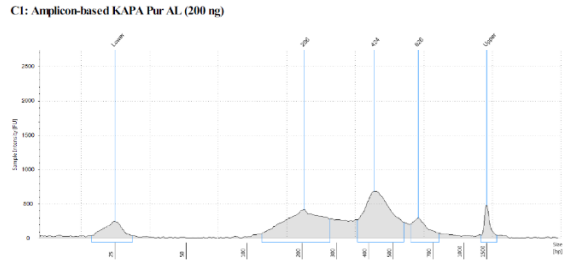
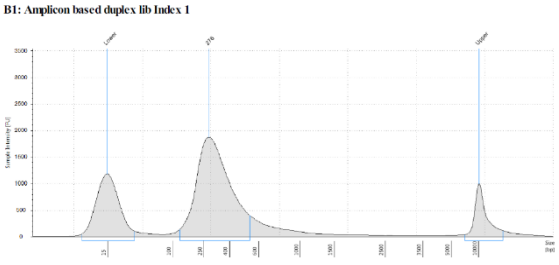
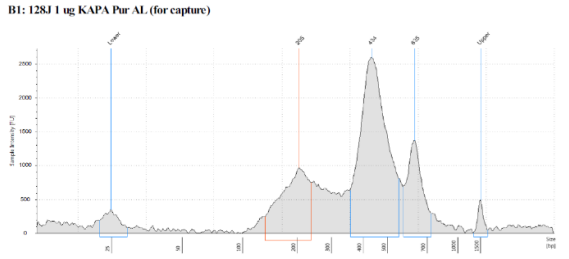
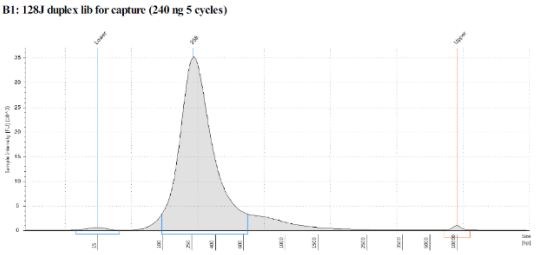
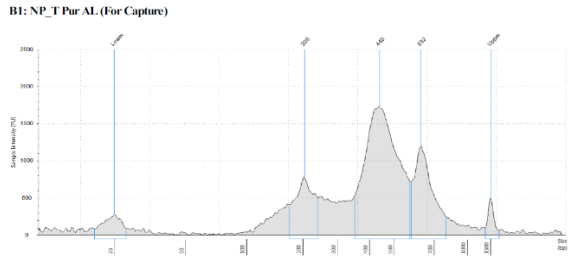
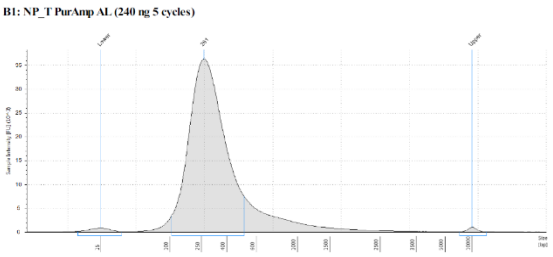
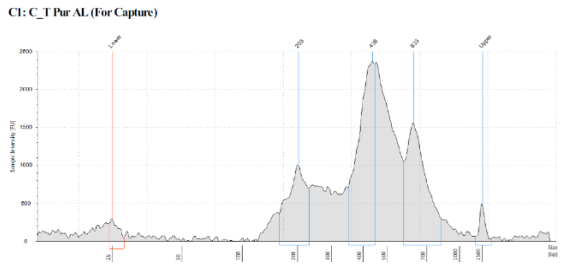
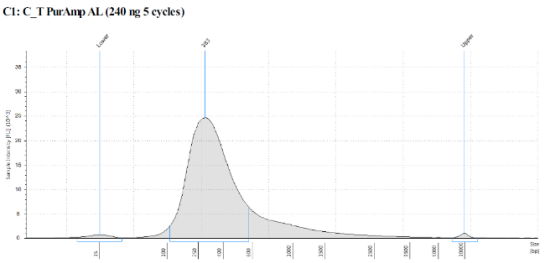
C576 Occ 	C576 Occ 
C576 Med 	C576 Med 

The results above are from Chapter 5, sections 5.2.1 (column 1) and 5.2.3, Attempt 3 (column 2): The initial duplex library construction of protocol 1 of double capture-based duplex library preparation was successfully done with High Sensitivity ScreenTape performed at two crucial stages of preparatory process. For all the six brain regions – three of NP16.163 (disease) and three of C576 (age-matched control) – all the adapter-ligated products had at least three main peaks recognized by the machine software as desired (~200, 400, and 600 bp; Column 1) on the HS D1000 ScreenTape while the amplified products with index primers gave a single sharp peak at ~300 bp (Column 2) as desired on the HS D5000 ScreenTape.

Appendix II

Double capture-based duplex library preparation Optimization ScreenTape results:

Protocol II (Pre-capture lib. prep.)

Post-ligation HS D1000 ScreenTape	**Post-amplification HS D5000 ScreenTape (240 ng, 5 cycles)**
Test DNA amplicon-based method	Test DNA amplicon-based method (4 amol 20 cycles)
	
Test DNA for (i) double capture (IDT protocol), (ii) double capture (Kennedy *et al* 2014), and (iii) double capture (Cap 1 – all DNA 16 cycles, Cap 2 – 4 amol 18 cycles)	Test DNA for (i) double capture (IDT protocol), (ii) double capture (Kennedy *et al* 2014), and (iii) double capture (Cap 1 – all DNA 16 cycles, Cap 2 – 4 amol 18 cycles)
	
NP16.163 Temp	NP16.163 Temp
	
C576 Temp	C576 Temp
	

The results above are from Chapter 6, section 6.2.2.1: The initial duplex library construction of protocol 2 of double capture-based duplex library preparation (IDT method) was successfully done with High Sensitivity ScreenTape performed at two crucial stages of preparatory process.

For all the four samples – Test DNA amplicon-based, Test DNA for double capture, NP16.163 (disease) Temporal lobe for double capture, and C576 (age-matched control) Temporal lobe for double capture – all the adapter-ligated products had at least three main peaks recognized by the machine software as desired (~200, 400, and 600 bp; Column 1) on the HS D1000 ScreenTape while the amplified products – with index primer for amplicon-based and P5 and test primers for the double capture (240 ng, 5 cycles) – gave a single sharp peak at ~300 bp (Column 2) as desired on the HS D5000 ScreenTape.

Appendix III

Double capture-based duplex library preparation Optimization: Protocol II – Probes

The following is the list of probes custom-designed by IDT for the '+' strand from GRCh38 reference genome used in Chapter 6, section 6.2.1.2 ('good' probes will capture DNA sequences that will yield good reads that can be properly mapped to the reference genome; 'risk' and 'remove' probes will capture repetitive sequences whose resultant reads will be hard to map correctly to their respective genomic location on the reference genome):

Chromosome	Start	Stop	Name	Score	Strand	Code	Len
chr17	45995919	45996039	776091_43057317_MAPT_1	1	+	**good**	120
chr17	45996039	45996159	776091_43057317_MAPT_2	1	+	**good**	120
chr17	45996159	45996279	776091_43057317_MAPT_3	1	+	**good**	120
chr17	45996279	45996399	776091_43057317_MAPT_4	1	+	**good**	120
chr17	45996399	45996519	776091_43057317_MAPT_5	1	+	**good**	120
chr17	45996519	45996639	776091_43057317_MAPT_6	1	+	**good**	120
chr17	45996639	45996759	776091_43057317_MAPT_7	1	+	**good**	120
chr17	45996759	45996879	776091_43057317_MAPT_8	1	+	**good**	120
chr17	45996879	45996999	776091_43057317_MAPT_9	1	+	**good**	120
chr17	45996999	45997119	776091_43057317_MAPT_10	1	+	**good**	120
chr17	45997119	45997239	776091_43057317_MAPT_11	1	+	**good**	120
chr17	45997239	45997359	776091_43057317_MAPT_12	1	+	**good**	120
chr17	45997359	45997479	776091_43057317_MAPT_13	1	+	**good**	120
chr17	45997479	45997599	776091_43057317_MAPT_14	10001	+	**remove**	120
chr17	45997599	45997719	776091_43057317_MAPT_15	9999	+	**remove**	120
chr17	45997719	45997839	776091_43057317_MAPT_16	889	+	**remove**	120
chr17	45997839	45997959	776091_43057317_MAPT_17	49	+	**good**	120
chr17	45997959	45998079	776091_43057317_MAPT_18	1	+	**good**	120
chr17	45998079	45998199	776091_43057317_MAPT_19	1	+	**good**	120
chr17	45998199	45998319	776091_43057317_MAPT_20	1	+	**good**	120
chr17	45998319	45998439	776091_43057317_MAPT_21	1	+	**good**	120
chr17	45998439	45998559	776091_43057317_MAPT_22	1	+	**good**	120
chr17	45998559	45998679	776091_43057317_MAPT_23	1	+	**good**	120
chr17	45998679	45998799	776091_43057317_MAPT_24	1	+	**good**	120
chr17	45998799	45998919	776091_43057317_MAPT_25	1	+	**good**	120
chr17	45998919	45999039	776091_43057317_MAPT_26	1	+	**good**	120
chr17	45999039	45999159	776091_43057317_MAPT_27	1	+	**good**	120
chr17	45999159	45999279	776091_43057317_MAPT_28	1	+	**good**	120
chr17	45999279	45999399	776091_43057317_MAPT_29	1	+	**good**	120
chr17	45999399	45999519	776091_43057317_MAPT_30	1	+	**good**	120
chr17	45999519	45999639	776091_43057317_MAPT_31	1	+	**good**	120
chr17	45999639	45999759	776091_43057317_MAPT_32	1	+	**good**	120
chr17	45999759	45999879	776091_43057317_MAPT_33	1	+	**good**	120
chr17	45999879	45999999	776091_43057317_MAPT_34	1	+	**good**	120
chr17	45999999	46000119	776091_43057317_MAPT_35	1	+	**good**	120
chr17	46000119	46000239	776091_43057317_MAPT_36	1	+	**good**	120
chr17	46000239	46000359	776091_43057317_MAPT_37	1	+	**good**	120
chr17	46000359	46000479	776091_43057317_MAPT_38	1	+	**good**	120
chr17	46000479	46000599	776091_43057317_MAPT_39	1	+	**good**	120
chr17	46000599	46000719	776091_43057317_MAPT_40	1	+	**good**	120

chr17	46000719	46000839	776091_43057317_MAPT_41	1	+	**good**	120
chr17	46000839	46000959	776091_43057317_MAPT_42	1	+	**good**	120
chr17	46000959	46001079	776091_43057317_MAPT_43	9850	+	**remove**	120
chr17	46001079	46001199	776091_43057317_MAPT_44	9177	+	**remove**	120
chr17	46001199	46001319	776091_43057317_MAPT_45	1	+	**good**	120
chr17	46001319	46001439	776091_43057317_MAPT_46	1	+	**good**	120
chr17	46001439	46001559	776091_43057317_MAPT_47	209	+	**risk**	120
chr17	46001559	46001679	776091_43057317_MAPT_48	8588	+	**remove**	120
chr17	46001679	46001799	776091_43057317_MAPT_49	1	+	**good**	120
chr17	46001799	46001919	776091_43057317_MAPT_50	1	+	**good**	120
chr17	46001919	46002039	776091_43057317_MAPT_51	1	+	**good**	120
chr17	46002039	46002159	776091_43057317_MAPT_52	1	+	**good**	120
chr17	46002159	46002279	776091_43057317_MAPT_53	1	+	**good**	120
chr17	46002279	46002399	776091_43057317_MAPT_54	1	+	**good**	120
chr17	46002399	46002519	776091_43057317_MAPT_55	1	+	**good**	120
chr17	46002519	46002639	776091_43057317_MAPT_56	1	+	**good**	120
chr17	46002639	46002759	776091_43057317_MAPT_57	1	+	**good**	120
chr17	46002759	46002879	776091_43057317_MAPT_58	8032	+	**remove**	120
chr17	46002879	46002999	776091_43057317_MAPT_59	9805	+	**remove**	120
chr17	46002999	46003119	776091_43057317_MAPT_60	7156	+	**remove**	120
chr17	46003119	46003239	776091_43057317_MAPT_61	1	+	**good**	120
chr17	46003239	46003359	776091_43057317_MAPT_62	9695	+	**remove**	120
chr17	46003359	46003479	776091_43057317_MAPT_63	9997	+	**remove**	120
chr17	46003479	46003599	776091_43057317_MAPT_64	9985	+	**remove**	120
chr17	46003599	46003719	776091_43057317_MAPT_65	1	+	**good**	120
chr17	46003719	46003839	776091_43057317_MAPT_66	1	+	**good**	120
chr17	46003839	46003959	776091_43057317_MAPT_67	1	+	**good**	120
chr17	46003959	46004079	776091_43057317_MAPT_68	1	+	**good**	120
chr17	46004079	46004199	776091_43057317_MAPT_69	1	+	**good**	120
chr17	46004199	46004319	776091_43057317_MAPT_70	1	+	**good**	120
chr17	46004319	46004439	776091_43057317_MAPT_71	1	+	**good**	120
chr17	46004439	46004559	776091_43057317_MAPT_72	1	+	**good**	120
chr17	46004559	46004679	776091_43057317_MAPT_73	1	+	**good**	120
chr17	46004679	46004799	776091_43057317_MAPT_74	1	+	**good**	120
chr17	46004799	46004919	776091_43057317_MAPT_75	10001	+	**remove**	120
chr17	46004919	46005039	776091_43057317_MAPT_76	10000	+	**remove**	120
chr17	46005039	46005159	776091_43057317_MAPT_77	666	+	**remove**	120
chr17	46005159	46005279	776091_43057317_MAPT_78	1	+	**good**	120
chr17	46005279	46005399	776091_43057317_MAPT_79	1	+	**good**	120
chr17	46005399	46005519	776091_43057317_MAPT_80	1	+	**good**	120
chr17	46005519	46005639	776091_43057317_MAPT_81	27	+	**good**	120
chr17	46005639	46005759	776091_43057317_MAPT_82	1	+	**good**	120
chr17	46005759	46005879	776091_43057317_MAPT_83	1	+	**good**	120
chr17	46005879	46005999	776091_43057317_MAPT_84	1	+	**good**	120
chr17	46005999	46006119	776091_43057317_MAPT_85	1	+	**good**	120
chr17	46006119	46006239	776091_43057317_MAPT_86	8772	+	**remove**	120
chr17	46006239	46006359	776091_43057317_MAPT_87	9476	+	**remove**	120
chr17	46006359	46006479	776091_43057317_MAPT_88	4483	+	**remove**	120
chr17	46006479	46006599	776091_43057317_MAPT_89	338	+	**risk**	120
chr17	46006599	46006719	776091_43057317_MAPT_90	9999	+	**remove**	120
chr17	46006719	46006839	776091_43057317_MAPT_91	10001	+	**remove**	120
chr17	46006839	46006959	776091_43057317_MAPT_92	9993	+	**remove**	120
chr17	46006959	46007079	776091_43057317_MAPT_93	583	+	**remove**	120
chr17	46007079	46007199	776091_43057317_MAPT_94	1573	+	**remove**	120
chr17	46007199	46007319	776091_43057317_MAPT_95	7277	+	**remove**	120
chr17	46007319	46007439	776091_43057317_MAPT_96	9810	+	**remove**	120
chr17	46007439	46007559	776091_43057317_MAPT_97	3561	+	**remove**	120
chr17	46007559	46007679	776091_43057317_MAPT_98	1	+	**good**	120

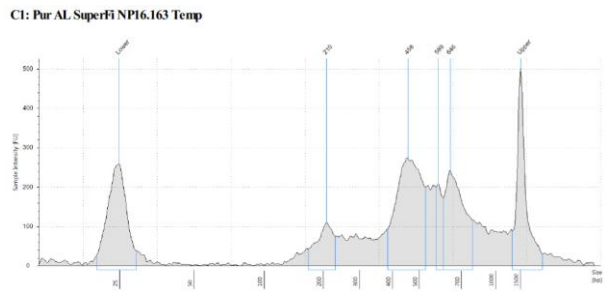
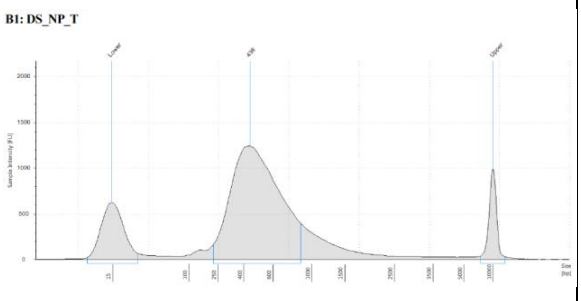
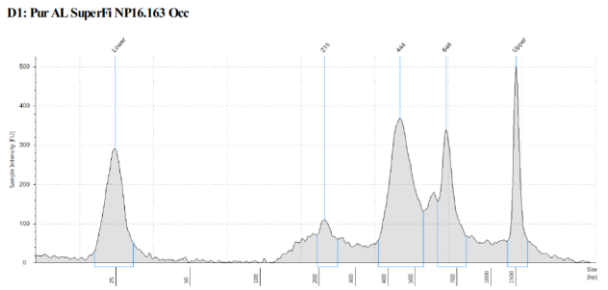
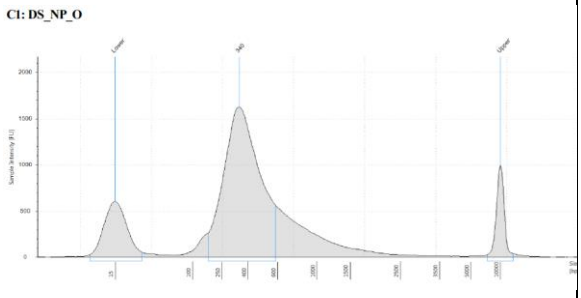
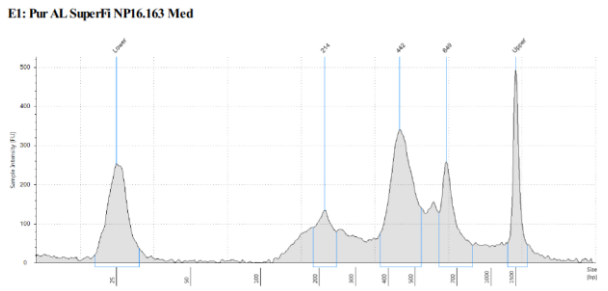
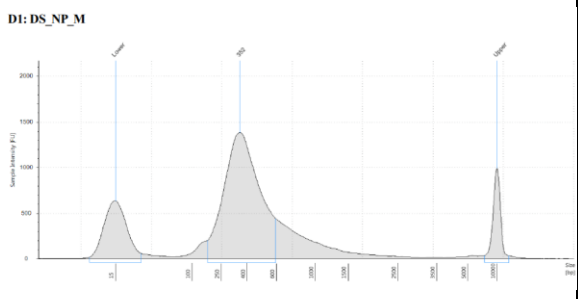
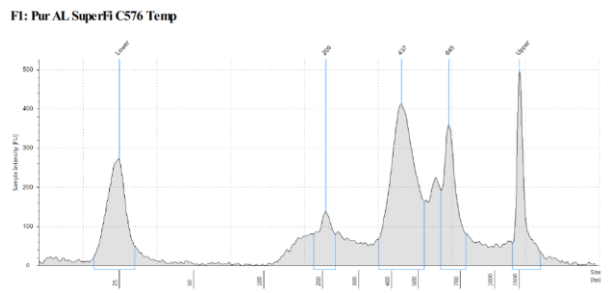
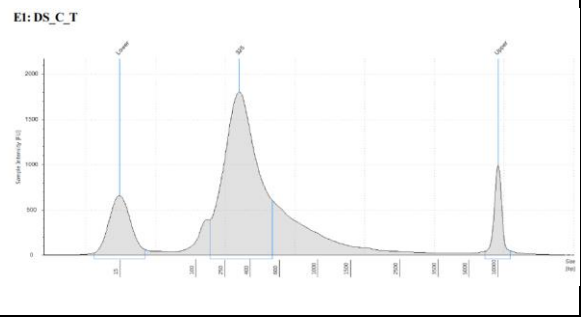
chr17	46007679	46007799	776091_43057317_MAPT_99	1	+	**good**	120
chr17	46007799	46007919	776091_43057317_MAPT_100	1	+	**good**	120
chr17	46007919	46008039	776091_43057317_MAPT_101	1	+	**good**	120
chr17	46008039	46008159	776091_43057317_MAPT_102	1244	+	**remove**	120
chr17	46008159	46008279	776091_43057317_MAPT_103	9987	+	**remove**	120
chr17	46008279	46008399	776091_43057317_MAPT_104	9958	+	**remove**	120
chr17	46008399	46008519	776091_43057317_MAPT_105	1	+	**good**	120
chr17	46008519	46008639	776091_43057317_MAPT_106	1	+	**good**	120
chr17	46008639	46008759	776091_43057317_MAPT_107	1	+	**good**	120
chr17	46008759	46008879	776091_43057317_MAPT_108	1	+	**good**	120
chr17	46008879	46008999	776091_43057317_MAPT_109	397	+	**risk**	120
chr17	46008999	46009119	776091_43057317_MAPT_110	9854	+	**remove**	120
chr17	46009119	46009239	776091_43057317_MAPT_111	8888	+	**remove**	120
chr17	46009239	46009359	776091_43057317_MAPT_112	1	+	**good**	120
chr17	46009359	46009479	776091_43057317_MAPT_113	1	+	**good**	120
chr17	46009479	46009599	776091_43057317_MAPT_114	1	+	**good**	120
chr17	46009599	46009719	776091_43057317_MAPT_115	1	+	**good**	120
chr17	46009719	46009839	776091_43057317_MAPT_116	1	+	**good**	120
chr17	46009839	46009959	776091_43057317_MAPT_117	1	+	**good**	120
chr17	46009959	46010079	776091_43057317_MAPT_118	1	+	**good**	120
chr17	46010079	46010199	776091_43057317_MAPT_119	1	+	**good**	120
chr17	46010199	46010319	776091_43057317_MAPT_120	2	+	**good**	120
chr17	46010319	46010439	776091_43057317_MAPT_121	1	+	**good**	120
chr17	46010439	46010559	776091_43057317_MAPT_122	1	+	**good**	120
chr17	46010559	46010679	776091_43057317_MAPT_123	1	+	**good**	120
chr17	46010679	46010799	776091_43057317_MAPT_124	1	+	**good**	120
chr17	46010799	46010919	776091_43057317_MAPT_125	1	+	**good**	120
chr17	46010919	46011039	776091_43057317_MAPT_126	1	+	**good**	120
chr17	46011039	46011159	776091_43057317_MAPT_127	2	+	**good**	120
chr17	46011159	46011279	776091_43057317_MAPT_128	9714	+	**remove**	120
chr17	46011279	46011399	776091_43057317_MAPT_129	9966	+	**remove**	120
chr17	46011399	46011519	776091_43057317_MAPT_130	1	+	**good**	120
chr17	46011519	46011639	776091_43057317_MAPT_131	1	+	**good**	120
chr17	46011639	46011759	776091_43057317_MAPT_132	1	+	**good**	120
chr17	46011759	46011879	776091_43057317_MAPT_133	1	+	**good**	120
chr17	46011879	46011999	776091_43057317_MAPT_134	1	+	**good**	120
chr17	46011999	46012119	776091_43057317_MAPT_135	1	+	**good**	120
chr17	46012119	46012239	776091_43057317_MAPT_136	1	+	**good**	120
chr17	46012239	46012359	776091_43057317_MAPT_137	1	+	**good**	120
chr17	46012359	46012479	776091_43057317_MAPT_138	1	+	**good**	120
chr17	46012479	46012599	776091_43057317_MAPT_139	1	+	**good**	120
chr17	46012599	46012719	776091_43057317_MAPT_140	1	+	**good**	120
chr17	46012719	46012839	776091_43057317_MAPT_141	1	+	**good**	120
chr17	46012839	46012959	776091_43057317_MAPT_142	1	+	**good**	120
chr17	46012959	46013079	776091_43057317_MAPT_143	1	+	**good**	120
chr17	46013079	46013199	776091_43057317_MAPT_144	1	+	**good**	120
chr17	46013199	46013319	776091_43057317_MAPT_145	1	+	**good**	120
chr17	46013319	46013439	776091_43057317_MAPT_146	1	+	**good**	120
chr17	46013439	46013559	776091_43057317_MAPT_147	1	+	**good**	120
chr17	46013559	46013679	776091_43057317_MAPT_148	1	+	**good**	120
chr17	46013679	46013799	776091_43057317_MAPT_149	1	+	**good**	120
chr17	46013799	46013919	776091_43057317_MAPT_150	1	+	**good**	120
chr17	46013919	46014039	776091_43057317_MAPT_151	1	+	**good**	120
chr17	46014039	46014159	776091_43057317_MAPT_152	1	+	**good**	120
chr17	46014159	46014279	776091_43057317_MAPT_153	1	+	**good**	120
chr17	46014279	46014399	776091_43057317_MAPT_154	1	+	**good**	120
chr17	46014399	46014519	776091_43057317_MAPT_155	1	+	**good**	120
chr17	46014519	46014639	776091_43057317_MAPT_156	312	+	**risk**	120

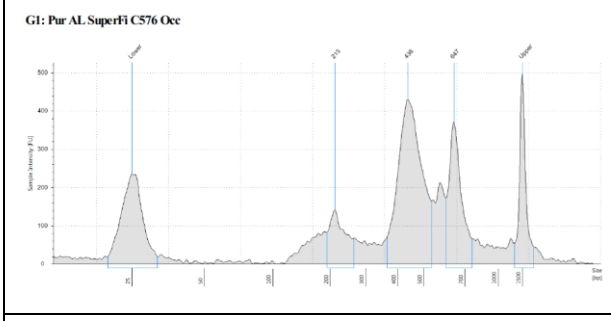
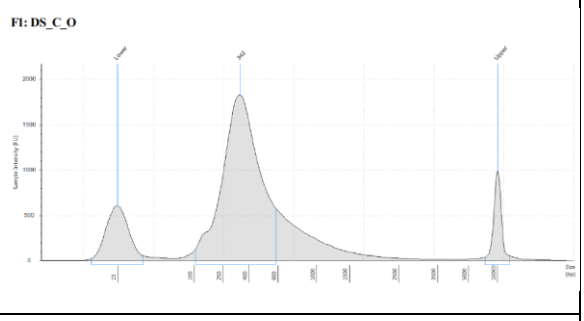
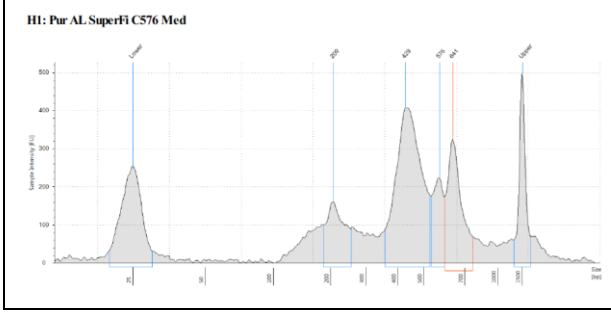
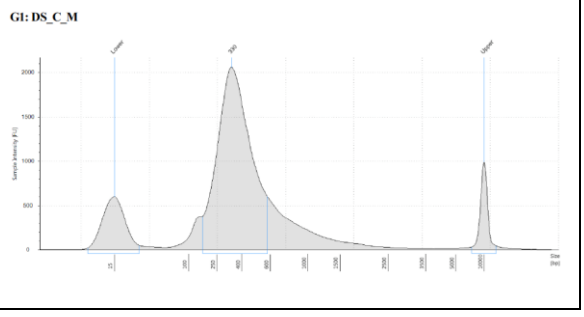
chr17	46014639	46014759	776091_43057317_MAPT_157	10000	+	**remove**	120
chr17	46014759	46014879	776091_43057317_MAPT_158	10001	+	**remove**	120
chr17	46014879	46014999	776091_43057317_MAPT_159	216	+	**risk**	120
chr17	46014999	46015119	776091_43057317_MAPT_160	3980	+	**remove**	120
chr17	46015119	46015239	776091_43057317_MAPT_161	9983	+	**remove**	120
chr17	46015239	46015359	776091_43057317_MAPT_162	9999	+	**remove**	120
chr17	46015359	46015479	776091_43057317_MAPT_163	10000	+	**remove**	120
chr17	46015479	46015599	776091_43057317_MAPT_164	10000	+	**remove**	120
chr17	46015599	46015719	776091_43057317_MAPT_165	8612	+	**remove**	120
chr17	46015719	46015839	776091_43057317_MAPT_166	5422	+	**remove**	120
chr17	46015839	46015959	776091_43057317_MAPT_167	1	+	**good**	120
chr17	46015959	46016079	776091_43057317_MAPT_168	1	+	**good**	120
chr17	46016079	46016199	776091_43057317_MAPT_169	9162	+	**remove**	120
chr17	46016199	46016319	776091_43057317_MAPT_170	9996	+	**remove**	120
chr17	46016319	46016439	776091_43057317_MAPT_171	9532	+	**remove**	120
chr17	46016439	46016559	776091_43057317_MAPT_172	6745	+	**remove**	120
chr17	46016559	46016679	776091_43057317_MAPT_173	9999	+	**remove**	120
chr17	46016679	46016799	776091_43057317_MAPT_174	9950	+	**remove**	120
chr17	46016799	46016919	776091_43057317_MAPT_175	37	+	**good**	120
chr17	46016919	46017039	776091_43057317_MAPT_176	1	+	**good**	120
chr17	46017039	46017159	776091_43057317_MAPT_177	1	+	**good**	120
chr17	46017159	46017279	776091_43057317_MAPT_178	1	+	**good**	120
chr17	46017279	46017399	776091_43057317_MAPT_179	9999	+	**remove**	120
chr17	46017399	46017519	776091_43057317_MAPT_180	4103	+	**remove**	120
chr17	46017519	46017639	776091_43057317_MAPT_181	9999	+	**remove**	120
chr17	46017639	46017759	776091_43057317_MAPT_182	10000	+	**remove**	120
chr17	46017759	46017879	776091_43057317_MAPT_183	1	+	**good**	120
chr17	46017879	46017999	776091_43057317_MAPT_184	10001	+	**remove**	120
chr17	46017999	46018119	776091_43057317_MAPT_185	10001	+	**remove**	120
chr17	46018119	46018239	776091_43057317_MAPT_186	64	+	**risk**	120
chr17	46018239	46018359	776091_43057317_MAPT_187	1	+	**good**	120
chr17	46018359	46018479	776091_43057317_MAPT_188	1	+	**good**	120
chr17	46018479	46018599	776091_43057317_MAPT_189	1	+	**good**	120
chr17	46018599	46018719	776091_43057317_MAPT_190	1	+	**good**	120
chr17	46018719	46018839	776091_43057317_MAPT_191	1	+	**good**	120
chr17	46018839	46018959	776091_43057317_MAPT_192	1	+	**good**	120
chr17	46018959	46019079	776091_43057317_MAPT_193	61	+	**risk**	120
chr17	46019079	46019199	776091_43057317_MAPT_194	9870	+	**remove**	120
chr17	46019199	46019319	776091_43057317_MAPT_195	9847	+	**remove**	120
chr17	46019319	46019439	776091_43057317_MAPT_196	7222	+	**remove**	120
chr17	46019439	46019559	776091_43057317_MAPT_197	9191	+	**remove**	120
chr17	46019559	46019679	776091_43057317_MAPT_198	9975	+	**remove**	120
chr17	46019679	46019799	776091_43057317_MAPT_199	4861	+	**remove**	120
chr17	46019799	46019919	776091_43057317_MAPT_200	9997	+	**remove**	120
chr17	46019919	46020039	776091_43057317_MAPT_201	9994	+	**remove**	120
chr17	46020039	46020159	776091_43057317_MAPT_202	1	+	**good**	120
chr17	46020159	46020279	776091_43057317_MAPT_203	1	+	**good**	120
chr17	46020279	46020399	776091_43057317_MAPT_204	1	+	**good**	120
chr17	46020399	46020519	776091_43057317_MAPT_205	1	+	**good**	120
chr17	46020519	46020639	776091_43057317_MAPT_206	7217	+	**remove**	120
chr17	46020639	46020759	776091_43057317_MAPT_207	9550	+	**remove**	120
chr17	46020759	46020879	776091_43057317_MAPT_208	9913	+	**remove**	120
chr17	46020879	46020999	776091_43057317_MAPT_209	1	+	**good**	120
chr17	46020999	46021119	776091_43057317_MAPT_210	1	+	**good**	120
chr17	46021119	46021239	776091_43057317_MAPT_211	1	+	**good**	120
chr17	46021239	46021359	776091_43057317_MAPT_212	1	+	**good**	120
chr17	46021359	46021479	776091_43057317_MAPT_213	1	+	**good**	120
chr17	46021479	46021599	776091_43057317_MAPT_214	1	+	**good**	120

chr17	46021599	46021719	776091_43057317_MAPT_215	1	+	**good**	120
chr17	46021719	46021839	776091_43057317_MAPT_216	1	+	**good**	120
chr17	46021839	46021959	776091_43057317_MAPT_217	1	+	**good**	120
chr17	46021959	46022079	776091_43057317_MAPT_218	1	+	**good**	120
chr17	46022079	46022199	776091_43057317_MAPT_219	9994	+	**remove**	120
chr17	46022199	46022319	776091_43057317_MAPT_220	10000	+	**remove**	120
chr17	46022319	46022439	776091_43057317_MAPT_221	327	+	**risk**	120
chr17	46022439	46022559	776091_43057317_MAPT_222	1	+	**good**	120
chr17	46022559	46022679	776091_43057317_MAPT_223	1	+	**good**	120
chr17	46022679	46022799	776091_43057317_MAPT_224	1	+	**good**	120
chr17	46022799	46022919	776091_43057317_MAPT_225	1391	+	**remove**	120
chr17	46022919	46023039	776091_43057317_MAPT_226	1	+	**good**	120
chr17	46023039	46023159	776091_43057317_MAPT_227	1	+	**good**	120
chr17	46023159	46023279	776091_43057317_MAPT_228	1	+	**good**	120
chr17	46023279	46023399	776091_43057317_MAPT_229	1	+	**good**	120
chr17	46023399	46023519	776091_43057317_MAPT_230	1	+	**good**	120
chr17	46023519	46023639	776091_43057317_MAPT_231	6917	+	**remove**	120
chr17	46023639	46023759	776091_43057317_MAPT_232	9976	+	**remove**	120
chr17	46023759	46023879	776091_43057317_MAPT_233	7235	+	**remove**	120
chr17	46023879	46023999	776091_43057317_MAPT_234	1	+	**good**	120
chr17	46023999	46024119	776091_43057317_MAPT_235	1	+	**good**	120
chr17	46024119	46024239	776091_43057317_MAPT_236	1	+	**good**	120
chr17	46024239	46024359	776091_43057317_MAPT_237	1	+	**good**	120
chr17	46024359	46024479	776091_43057317_MAPT_238	1	+	**good**	120
chr17	46024479	46024599	776091_43057317_MAPT_239	1	+	**good**	120
chr17	46024599	46024719	776091_43057317_MAPT_240	1	+	**good**	120

Appendix IV

SuperFi-mediated amplicon-based duplex library preparation ScreenTape results

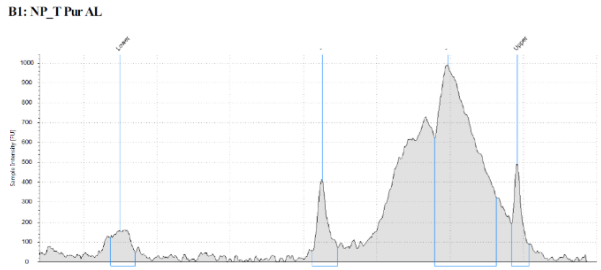
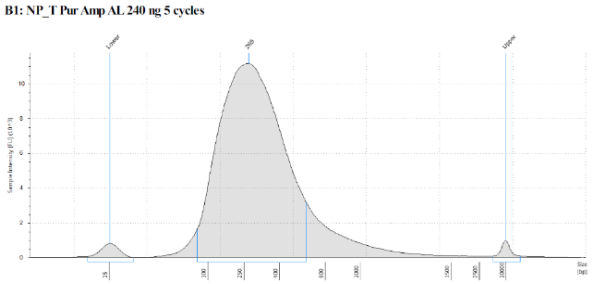
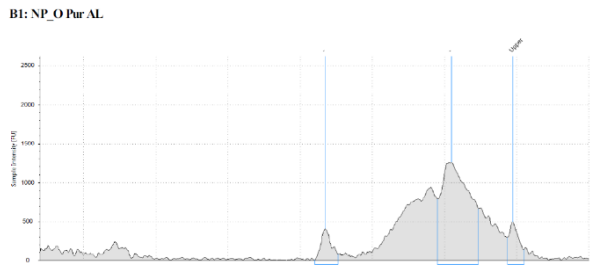
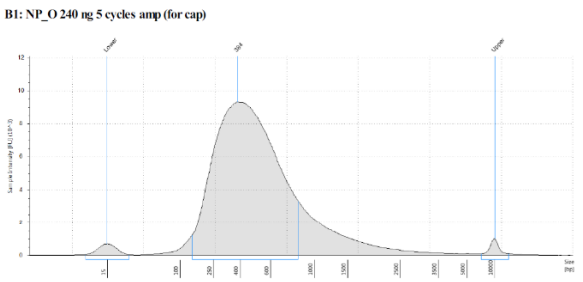
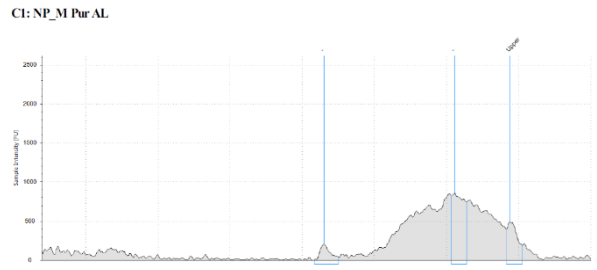
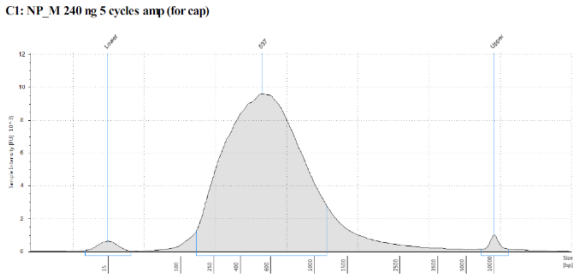
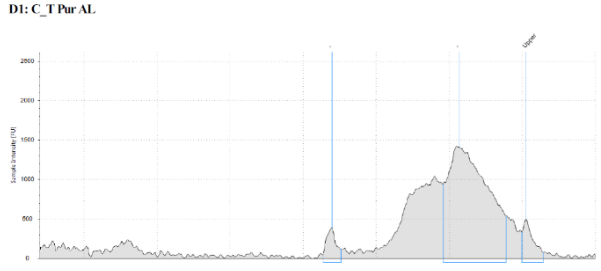
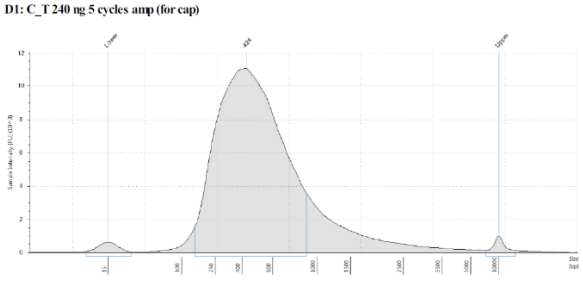
Post-ligation HS D1000 ScreenTape	Post-amplification HS D5000 ScreenTape (4 amol 20 cycles)
NP16.163 Temp 	NP16.163 Temp 
NP16.163 Occ 	NP16.163 Occ 
NP16.163 Med 	NP16.163 Med 
C576 Temp 	C576 Temp 

C576 Occ 	C576 Occ 
C576 Med 	C576 Med 

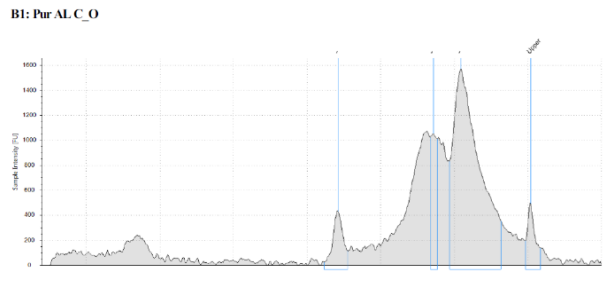
The results above are from Chapter 7, section 7.2.1.3: The SuperFi-mediated amplicon-based duplex library preparation was successfully done with High Sensitivity ScreenTape performed at two crucial stages of preparatory process. For all the six brain regions – three of NP16.163 (disease) and three of C576 (age-matched control) – all the adapter-ligated products had at least three main peaks recognized by the machine software as desired (~200, 400, and 600 bp; Column 1) on the HS D1000 ScreenTape while the amplified products with index primers gave a single sharp peak at ~300 bp (Column 2) as desired on the HS D5000 ScreenTape.

Appendix V

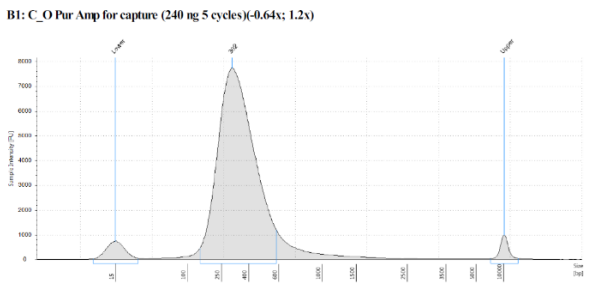
Double capture-based duplex library preparation ScreenTape results: Protocol II (Pre-capture lib. prep.)

Post-ligation HS D1000 ScreenTape	Post-amplification HS D5000 ScreenTape (240 ng 5 cycles)
NP16.163 Temp	NP16.163 Temp
	
NP16.163 Occ	NP16.163 Occ
	
NP16.163 Med	NP16.163 Med
	
C576 Temp	C576 Temp
	

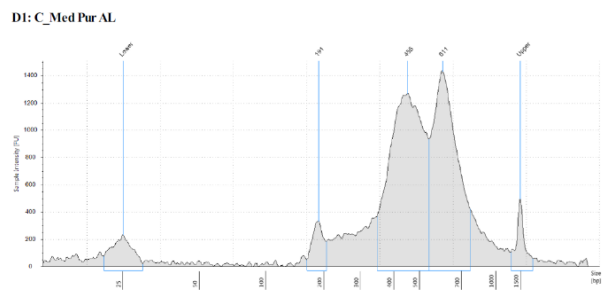
C576 Occ



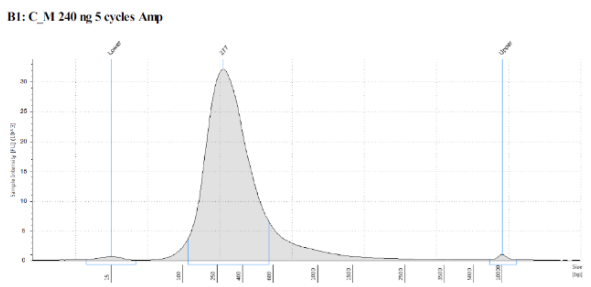
C576 Occ



C576 Med



C576 Med



The results above are from Chapter 7, section 7.2.2.1: The initial duplex library construction of protocol 2 of double capture-based duplex library preparation (IDT method) was successfully done with High Sensitivity ScreenTape performed at two crucial stages of preparatory process. Unfortunately, the HS D1000 ladder was completely consumed at this stage of performing experiments, the HS D1000 buffer had very few microlitres left and the ScreenTapes had passed the expiry date at the time of this experiment and the waiting time to obtain new reagents from the company more than a year because of Brexit and the COVID-19 pandemic, so we settled for poor quality results for NP16.163 Occ, Med, and C576 Temp at the adapter-ligation stage.

For all the six brain regions – three of NP16.163 (disease) and three of C576 (age-matched control) – all the adapter-ligated products had at least approximately three main recognizable peaks (~200, 400, and 600 bp; Column 1) on the HS D1000 ScreenTape while the amplified products with index primers gave a single peak at ~300-500 bp (Column 2) as desired on the HS D5000 ScreenTape. The HS D5000 ScreenTape results of C576 Occ initially had a smaller second peak with larger fragments that was selected against in a double-tail beads selection.

Appendix VI

Double capture-based duplex library preparation: Protocol II – Probes

1. The following is the list of "good" probes custom-designed by IDT for the '+' strand from GRCh38 reference genome used in Chapter 7, section 7.2.2.2:

Chromosome	Start	Stop	Name	Score	Strand	Code	Len
chr17	45894544	45894664	781615_44389461_MAPT_1	1	+	**good**	120
chr17	45894664	45894784	781615_44389461_MAPT_2	1	+	**good**	120
chr17	45894784	45894904	781615_44389461_MAPT_3	1	+	**good**	120
chr17	45894904	45895024	781615_44389461_MAPT_4	1	+	**good**	120
chr17	45895024	45895144	781615_44389461_MAPT_5	30	+	**good**	120
chr17	45895144	45895264	781615_44389461_MAPT_6	1	+	**good**	120
chr17	45895264	45895384	781615_44389461_MAPT_7	1	+	**good**	120
chr17	45895384	45895504	781615_44389461_MAPT_8	1	+	**good**	120
chr17	45895504	45895624	781615_44389461_MAPT_9	1	+	**good**	120
chr17	45895624	45895744	781615_44389461_MAPT_10	1	+	**good**	120
chr17	45895744	45895864	781615_44389461_MAPT_11	1	+	**good**	120
chr17	45895864	45895984	781615_44389461_MAPT_12	1	+	**good**	120
chr17	45895984	45896104	781615_44389461_MAPT_13	1	+	**good**	120
chr17	45896104	45896224	781615_44389461_MAPT_14	1	+	**good**	120
chr17	45896224	45896344	781615_44389461_MAPT_15	1	+	**good**	120
chr17	45896344	45896464	781615_44389461_MAPT_16	1	+	**good**	120
chr17	45896464	45896584	781615_44389461_MAPT_17	1	+	**good**	120
chr17	45896584	45896704	781615_44389461_MAPT_18	1	+	**good**	120
chr17	45896704	45896824	781615_44389461_MAPT_19	1	+	**good**	120
chr17	45896824	45896944	781615_44389461_MAPT_20	1	+	**good**	120
chr17	45896944	45897064	781615_44389461_MAPT_21	1	+	**good**	120
chr17	45897064	45897184	781615_44389461_MAPT_22	1	+	**good**	120
chr17	45897184	45897304	781615_44389461_MAPT_23	1	+	**good**	120
chr17	45897304	45897424	781615_44389461_MAPT_24	1	+	**good**	120
chr17	45897424	45897544	781615_44389461_MAPT_25	1	+	**good**	120
chr17	45897544	45897664	781615_44389461_MAPT_26	1	+	**good**	120
chr17	45897664	45897784	781615_44389461_MAPT_27	1	+	**good**	120
chr17	45897784	45897904	781615_44389461_MAPT_28	1	+	**good**	120
chr17	45897904	45898024	781615_44389461_MAPT_29	1	+	**good**	120
chr17	45898024	45898144	781615_44389461_MAPT_30	1	+	**good**	120
chr17	45898144	45898264	781615_44389461_MAPT_31	1	+	**good**	120
chr17	45898264	45898384	781615_44389461_MAPT_32	1	+	**good**	120
chr17	45898384	45898504	781615_44389461_MAPT_33	1	+	**good**	120
chr17	45898504	45898624	781615_44389461_MAPT_34	1	+	**good**	120
chr17	45898624	45898744	781615_44389461_MAPT_35	1	+	**good**	120
chr17	45898744	45898864	781615_44389461_MAPT_36	1	+	**good**	120

chr17	45898864	45898984	781615_44389461_MAPT_37	1	+	**good**	120
chr17	45898984	45899104	781615_44389461_MAPT_38	1	+	**good**	120
chr17	45899104	45899224	781615_44389461_MAPT_39	1	+	**good**	120
chr17	45899344	45899464	781615_44389461_MAPT_41	2	+	**good**	120
chr17	45899464	45899584	781615_44389461_MAPT_42	1	+	**good**	120
chr17	45899584	45899704	781615_44389461_MAPT_43	1	+	**good**	120
chr17	45899704	45899824	781615_44389461_MAPT_44	1	+	**good**	120
chr17	45899824	45899944	781615_44389461_MAPT_45	1	+	**good**	120
chr17	45899944	45900064	781615_44389461_MAPT_46	1	+	**good**	120
chr17	45900064	45900184	781615_44389461_MAPT_47	1	+	**good**	120
chr17	45900184	45900304	781615_44389461_MAPT_48	1	+	**good**	120
chr17	45900304	45900424	781615_44389461_MAPT_49	1	+	**good**	120
chr17	45900424	45900544	781615_44389461_MAPT_50	1	+	**good**	120
chr17	45900544	45900664	781615_44389461_MAPT_51	1	+	**good**	120
chr17	45900664	45900784	781615_44389461_MAPT_52	1	+	**good**	120
chr17	45900784	45900904	781615_44389461_MAPT_53	1	+	**good**	120
chr17	45900904	45901024	781615_44389461_MAPT_54	1	+	**good**	120
chr17	45901024	45901144	781615_44389461_MAPT_55	1	+	**good**	120
chr17	45901144	45901264	781615_44389461_MAPT_56	1	+	**good**	120
chr17	45901264	45901384	781615_44389461_MAPT_57	1	+	**good**	120
chr17	45901384	45901504	781615_44389461_MAPT_58	1	+	**good**	120
chr17	45901504	45901624	781615_44389461_MAPT_59	1	+	**good**	120
chr17	45901624	45901744	781615_44389461_MAPT_60	1	+	**good**	120
chr17	45901744	45901864	781615_44389461_MAPT_61	1	+	**good**	120
chr17	45901864	45901984	781615_44389461_MAPT_62	1	+	**good**	120
chr17	45901984	45902104	781615_44389461_MAPT_63	1	+	**good**	120
chr17	45902344	45902464	781615_44389461_MAPT_66	2	+	**good**	120
chr17	45902464	45902584	781615_44389461_MAPT_67	1	+	**good**	120
chr17	45902584	45902704	781615_44389461_MAPT_68	1	+	**good**	120
chr17	45902704	45902824	781615_44389461_MAPT_69	1	+	**good**	120
chr17	45902824	45902944	781615_44389461_MAPT_70	1	+	**good**	120
chr17	45902944	45903064	781615_44389461_MAPT_71	1	+	**good**	120
chr17	45903064	45903184	781615_44389461_MAPT_72	1	+	**good**	120
chr17	45903184	45903304	781615_44389461_MAPT_73	1	+	**good**	120
chr17	45903304	45903424	781615_44389461_MAPT_74	1	+	**good**	120
chr17	45904744	45904864	781615_44389461_MAPT_86	1	+	**good**	120
chr17	45904864	45904984	781615_44389461_MAPT_87	1	+	**good**	120
chr17	45904984	45905104	781615_44389461_MAPT_88	1	+	**good**	120
chr17	45905104	45905224	781615_44389461_MAPT_89	1	+	**good**	120
chr17	45905224	45905344	781615_44389461_MAPT_90	1	+	**good**	120
chr17	45905344	45905464	781615_44389461_MAPT_91	1	+	**good**	120
chr17	45905464	45905584	781615_44389461_MAPT_92	1	+	**good**	120
chr17	45905584	45905704	781615_44389461_MAPT_93	1	+	**good**	120
chr17	45905704	45905824	781615_44389461_MAPT_94	1	+	**good**	120
chr17	45905824	45905944	781615_44389461_MAPT_95	1	+	**good**	120
chr17	45905944	45906064	781615_44389461_MAPT_96	1	+	**good**	120

chr17	45906064	45906184	781615_44389461_MAPT_97	1	+	**good**	120
chr17	45906184	45906304	781615_44389461_MAPT_98	1	+	**good**	120
chr17	45906304	45906424	781615_44389461_MAPT_99	1	+	**good**	120
chr17	45906424	45906544	781615_44389461_MAPT_100	1	+	**good**	120
chr17	45906544	45906664	781615_44389461_MAPT_101	1	+	**good**	120
chr17	45906664	45906784	781615_44389461_MAPT_102	1	+	**good**	120
chr17	45906784	45906904	781615_44389461_MAPT_103	1	+	**good**	120
chr17	45906904	45907024	781615_44389461_MAPT_104	1	+	**good**	120
chr17	45907024	45907144	781615_44389461_MAPT_105	1	+	**good**	120
chr17	45907144	45907264	781615_44389461_MAPT_106	1	+	**good**	120
chr17	45907264	45907384	781615_44389461_MAPT_107	1	+	**good**	120
chr17	45907384	45907504	781615_44389461_MAPT_108	1	+	**good**	120
chr17	45907504	45907624	781615_44389461_MAPT_109	1	+	**good**	120
chr17	45907624	45907744	781615_44389461_MAPT_110	1	+	**good**	120
chr17	45907744	45907864	781615_44389461_MAPT_111	1	+	**good**	120
chr17	45907864	45907984	781615_44389461_MAPT_112	1	+	**good**	120
chr17	45907984	45908104	781615_44389461_MAPT_113	1	+	**good**	120
chr17	45908104	45908224	781615_44389461_MAPT_114	1	+	**good**	120
chr17	45908224	45908344	781615_44389461_MAPT_115	1	+	**good**	120
chr17	45908344	45908464	781615_44389461_MAPT_116	1	+	**good**	120
chr17	45908464	45908584	781615_44389461_MAPT_117	1	+	**good**	120
chr17	45908584	45908704	781615_44389461_MAPT_118	1	+	**good**	120
chr17	45908704	45908824	781615_44389461_MAPT_119	1	+	**good**	120
chr17	45908824	45908944	781615_44389461_MAPT_120	1	+	**good**	120
chr17	45908944	45909064	781615_44389461_MAPT_121	1	+	**good**	120
chr17	45909064	45909184	781615_44389461_MAPT_122	1	+	**good**	120
chr17	45909184	45909304	781615_44389461_MAPT_123	1	+	**good**	120
chr17	45909304	45909424	781615_44389461_MAPT_124	1	+	**good**	120
chr17	45910024	45910144	781615_44389461_MAPT_130	1	+	**good**	120
chr17	45910144	45910264	781615_44389461_MAPT_131	1	+	**good**	120
chr17	45910264	45910384	781615_44389461_MAPT_132	1	+	**good**	120
chr17	45910384	45910504	781615_44389461_MAPT_133	1	+	**good**	120
chr17	45910504	45910624	781615_44389461_MAPT_134	1	+	**good**	120
chr17	45910624	45910744	781615_44389461_MAPT_135	1	+	**good**	120
chr17	45910744	45910864	781615_44389461_MAPT_136	1	+	**good**	120
chr17	45910864	45910984	781615_44389461_MAPT_137	1	+	**good**	120
chr17	45910984	45911104	781615_44389461_MAPT_138	1	+	**good**	120
chr17	45911104	45911224	781615_44389461_MAPT_139	1	+	**good**	120
chr17	45911224	45911344	781615_44389461_MAPT_140	1	+	**good**	120
chr17	45911824	45911944	781615_44389461_MAPT_145	1	+	**good**	120
chr17	45911944	45912064	781615_44389461_MAPT_146	1	+	**good**	120
chr17	45912064	45912184	781615_44389461_MAPT_147	1	+	**good**	120
chr17	45912184	45912304	781615_44389461_MAPT_148	2	+	**good**	120
chr17	45912544	45912664	781615_44389461_MAPT_151	37	+	**good**	120
chr17	45912664	45912784	781615_44389461_MAPT_152	1	+	**good**	120
chr17	45912784	45912904	781615_44389461_MAPT_153	1	+	**good**	120

chr17	45912904	45913024	781615_44389461_MAPT_154	1	+	**good**	120
chr17	45913384	45913504	781615_44389461_MAPT_158	1	+	**good**	120
chr17	45913504	45913624	781615_44389461_MAPT_159	1	+	**good**	120
chr17	45913624	45913744	781615_44389461_MAPT_160	1	+	**good**	120
chr17	45913744	45913864	781615_44389461_MAPT_161	1	+	**good**	120
chr17	45913864	45913984	781615_44389461_MAPT_162	1	+	**good**	120
chr17	45913984	45914104	781615_44389461_MAPT_163	1	+	**good**	120
chr17	45914104	45914224	781615_44389461_MAPT_164	1	+	**good**	120
chr17	45914224	45914344	781615_44389461_MAPT_165	1	+	**good**	120
chr17	45914344	45914464	781615_44389461_MAPT_166	1	+	**good**	120
chr17	45914464	45914584	781615_44389461_MAPT_167	1	+	**good**	120
chr17	45914584	45914704	781615_44389461_MAPT_168	1	+	**good**	120
chr17	45915064	45915184	781615_44389461_MAPT_172	1	+	**good**	120
chr17	45915544	45915664	781615_44389461_MAPT_176	34	+	**good**	120
chr17	45915664	45915784	781615_44389461_MAPT_177	1	+	**good**	120
chr17	45915784	45915904	781615_44389461_MAPT_178	1	+	**good**	120
chr17	45915904	45916024	781615_44389461_MAPT_179	1	+	**good**	120
chr17	45916024	45916144	781615_44389461_MAPT_180	1	+	**good**	120
chr17	45916144	45916264	781615_44389461_MAPT_181	1	+	**good**	120
chr17	45916264	45916384	781615_44389461_MAPT_182	1	+	**good**	120
chr17	45916384	45916504	781615_44389461_MAPT_183	1	+	**good**	120
chr17	45916504	45916624	781615_44389461_MAPT_184	1	+	**good**	120
chr17	45916624	45916744	781615_44389461_MAPT_185	1	+	**good**	120
chr17	45916744	45916864	781615_44389461_MAPT_186	1	+	**good**	120
chr17	45916864	45916984	781615_44389461_MAPT_187	1	+	**good**	120
chr17	45916984	45917104	781615_44389461_MAPT_188	1	+	**good**	120
chr17	45917104	45917224	781615_44389461_MAPT_189	1	+	**good**	120
chr17	45917224	45917344	781615_44389461_MAPT_190	1	+	**good**	120
chr17	45917344	45917464	781615_44389461_MAPT_191	1	+	**good**	120
chr17	45917464	45917584	781615_44389461_MAPT_192	1	+	**good**	120
chr17	45917584	45917704	781615_44389461_MAPT_193	1	+	**good**	120
chr17	45918064	45918184	781615_44389461_MAPT_197	1	+	**good**	120
chr17	45918184	45918304	781615_44389461_MAPT_198	1	+	**good**	120
chr17	45918304	45918424	781615_44389461_MAPT_199	1	+	**good**	120
chr17	45918424	45918544	781615_44389461_MAPT_200	1	+	**good**	120
chr17	45918544	45918664	781615_44389461_MAPT_201	1	+	**good**	120
chr17	45919144	45919264	781615_44389461_MAPT_206	1	+	**good**	120
chr17	45919264	45919384	781615_44389461_MAPT_207	1	+	**good**	120
chr17	45919384	45919504	781615_44389461_MAPT_208	1	+	**good**	120
chr17	45919504	45919624	781615_44389461_MAPT_209	1	+	**good**	120
chr17	45919984	45920104	781615_44389461_MAPT_213	1	+	**good**	120
chr17	45920104	45920224	781615_44389461_MAPT_214	1	+	**good**	120
chr17	45920224	45920344	781615_44389461_MAPT_215	1	+	**good**	120
chr17	45920344	45920464	781615_44389461_MAPT_216	1	+	**good**	120
chr17	45920464	45920584	781615_44389461_MAPT_217	1	+	**good**	120
chr17	45920584	45920704	781615_44389461_MAPT_218	1	+	**good**	120

chr17	45920704	45920824	781615_44389461_MAPT_219	1	+	**good**	120
chr17	45920824	45920944	781615_44389461_MAPT_220	1	+	**good**	120
chr17	45920944	45921064	781615_44389461_MAPT_221	1	+	**good**	120
chr17	45921064	45921184	781615_44389461_MAPT_222	1	+	**good**	120
chr17	45921184	45921304	781615_44389461_MAPT_223	1	+	**good**	120
chr17	45921304	45921424	781615_44389461_MAPT_224	1	+	**good**	120
chr17	45921424	45921544	781615_44389461_MAPT_225	1	+	**good**	120
chr17	45921544	45921664	781615_44389461_MAPT_226	1	+	**good**	120
chr17	45921664	45921784	781615_44389461_MAPT_227	1	+	**good**	120
chr17	45921784	45921904	781615_44389461_MAPT_228	1	+	**good**	120
chr17	45921904	45922024	781615_44389461_MAPT_229	1	+	**good**	120
chr17	45922264	45922384	781615_44389461_MAPT_232	1	+	**good**	120
chr17	45922384	45922504	781615_44389461_MAPT_233	1	+	**good**	120
chr17	45922504	45922624	781615_44389461_MAPT_234	1	+	**good**	120
chr17	45922624	45922744	781615_44389461_MAPT_235	1	+	**good**	120
chr17	45922744	45922864	781615_44389461_MAPT_236	1	+	**good**	120
chr17	45922864	45922984	781615_44389461_MAPT_237	1	+	**good**	120
chr17	45922984	45923104	781615_44389461_MAPT_238	1	+	**good**	120
chr17	45923104	45923224	781615_44389461_MAPT_239	1	+	**good**	120
chr17	45923224	45923344	781615_44389461_MAPT_240	1	+	**good**	120
chr17	45923344	45923464	781615_44389461_MAPT_241	1	+	**good**	120
chr17	45923464	45923584	781615_44389461_MAPT_242	1	+	**good**	120
chr17	45923584	45923704	781615_44389461_MAPT_243	1	+	**good**	120
chr17	45923704	45923824	781615_44389461_MAPT_244	1	+	**good**	120
chr17	45923824	45923944	781615_44389461_MAPT_245	3	+	**good**	120
chr17	45923944	45924064	781615_44389461_MAPT_246	1	+	**good**	120
chr17	45924064	45924184	781615_44389461_MAPT_247	1	+	**good**	120
chr17	45924184	45924304	781615_44389461_MAPT_248	1	+	**good**	120
chr17	45924304	45924424	781615_44389461_MAPT_249	1	+	**good**	120
chr17	45924424	45924544	781615_44389461_MAPT_250	1	+	**good**	120
chr17	45924544	45924664	781615_44389461_MAPT_251	1	+	**good**	120
chr17	45924664	45924784	781615_44389461_MAPT_252	1	+	**good**	120
chr17	45924784	45924904	781615_44389461_MAPT_253	1	+	**good**	120
chr17	45924904	45925024	781615_44389461_MAPT_254	1	+	**good**	120
chr17	45925024	45925144	781615_44389461_MAPT_255	1	+	**good**	120
chr17	45925144	45925264	781615_44389461_MAPT_256	1	+	**good**	120
chr17	45925264	45925384	781615_44389461_MAPT_257	1	+	**good**	120
chr17	45925384	45925504	781615_44389461_MAPT_258	1	+	**good**	120
chr17	45925504	45925624	781615_44389461_MAPT_259	1	+	**good**	120
chr17	45925624	45925744	781615_44389461_MAPT_260	1	+	**good**	120
chr17	45925744	45925864	781615_44389461_MAPT_261	1	+	**good**	120
chr17	45926224	45926344	781615_44389461_MAPT_265	9	+	**good**	120
chr17	45926584	45926704	781615_44389461_MAPT_268	1	+	**good**	120
chr17	45926704	45926824	781615_44389461_MAPT_269	1	+	**good**	120
chr17	45926824	45926944	781615_44389461_MAPT_270	34	+	**good**	120
chr17	45927064	45927184	781615_44389461_MAPT_272	3	+	**good**	120

chr17	45927184	45927304	781615_44389461_MAPT_273	1	+	**good**	120
chr17	45927304	45927424	781615_44389461_MAPT_274	1	+	**good**	120
chr17	45927424	45927544	781615_44389461_MAPT_275	1	+	**good**	120
chr17	45927544	45927664	781615_44389461_MAPT_276	1	+	**good**	120
chr17	45928024	45928144	781615_44389461_MAPT_280	1	+	**good**	120
chr17	45928144	45928264	781615_44389461_MAPT_281	1	+	**good**	120
chr17	45928264	45928384	781615_44389461_MAPT_282	1	+	**good**	120
chr17	45928384	45928504	781615_44389461_MAPT_283	1	+	**good**	120
chr17	45928504	45928624	781615_44389461_MAPT_284	1	+	**good**	120
chr17	45928984	45929104	781615_44389461_MAPT_288	1	+	**good**	120
chr17	45929104	45929224	781615_44389461_MAPT_289	1	+	**good**	120
chr17	45929224	45929344	781615_44389461_MAPT_290	1	+	**good**	120
chr17	45929344	45929464	781615_44389461_MAPT_291	1	+	**good**	120
chr17	45929464	45929584	781615_44389461_MAPT_292	1	+	**good**	120
chr17	45929584	45929704	781615_44389461_MAPT_293	1	+	**good**	120
chr17	45929704	45929824	781615_44389461_MAPT_294	1	+	**good**	120
chr17	45929824	45929944	781615_44389461_MAPT_295	1	+	**good**	120
chr17	45929944	45930064	781615_44389461_MAPT_296	1	+	**good**	120
chr17	45930424	45930544	781615_44389461_MAPT_300	1	+	**good**	120
chr17	45930544	45930664	781615_44389461_MAPT_301	4	+	**good**	120
chr17	45930664	45930784	781615_44389461_MAPT_302	1	+	**good**	120
chr17	45930784	45930904	781615_44389461_MAPT_303	1	+	**good**	120
chr17	45930904	45931024	781615_44389461_MAPT_304	1	+	**good**	120
chr17	45931024	45931144	781615_44389461_MAPT_305	1	+	**good**	120
chr17	45931144	45931264	781615_44389461_MAPT_306	1	+	**good**	120
chr17	45931264	45931384	781615_44389461_MAPT_307	1	+	**good**	120
chr17	45931384	45931504	781615_44389461_MAPT_308	1	+	**good**	120
chr17	45931504	45931624	781615_44389461_MAPT_309	1	+	**good**	120
chr17	45931624	45931744	781615_44389461_MAPT_310	2	+	**good**	120
chr17	45932104	45932224	781615_44389461_MAPT_314	2	+	**good**	120
chr17	45932224	45932344	781615_44389461_MAPT_315	16	+	**good**	120
chr17	45932584	45932704	781615_44389461_MAPT_318	6	+	**good**	120
chr17	45932704	45932824	781615_44389461_MAPT_319	1	+	**good**	120
chr17	45933064	45933184	781615_44389461_MAPT_322	2	+	**good**	120
chr17	45933544	45933664	781615_44389461_MAPT_326	1	+	**good**	120
chr17	45933664	45933784	781615_44389461_MAPT_327	1	+	**good**	120
chr17	45933784	45933904	781615_44389461_MAPT_328	1	+	**good**	120
chr17	45933904	45934024	781615_44389461_MAPT_329	1	+	**good**	120
chr17	45934384	45934504	781615_44389461_MAPT_333	1	+	**good**	120
chr17	45934504	45934624	781615_44389461_MAPT_334	1	+	**good**	120
chr17	45934624	45934744	781615_44389461_MAPT_335	1	+	**good**	120
chr17	45934744	45934864	781615_44389461_MAPT_336	1	+	**good**	120
chr17	45934864	45934984	781615_44389461_MAPT_337	1	+	**good**	120
chr17	45934984	45935104	781615_44389461_MAPT_338	1	+	**good**	120
chr17	45935104	45935224	781615_44389461_MAPT_339	1	+	**good**	120
chr17	45935224	45935344	781615_44389461_MAPT_340	1	+	**good**	120

chr17	45935344	45935464	781615_44389461_MAPT_341	1	+	**good**	120
chr17	45935464	45935584	781615_44389461_MAPT_342	1	+	**good**	120
chr17	45935584	45935704	781615_44389461_MAPT_343	1	+	**good**	120
chr17	45935704	45935824	781615_44389461_MAPT_344	1	+	**good**	120
chr17	45935824	45935944	781615_44389461_MAPT_345	1	+	**good**	120
chr17	45935944	45936064	781615_44389461_MAPT_346	1	+	**good**	120
chr17	45936064	45936184	781615_44389461_MAPT_347	1	+	**good**	120
chr17	45936184	45936304	781615_44389461_MAPT_348	1	+	**good**	120
chr17	45936304	45936424	781615_44389461_MAPT_349	1	+	**good**	120
chr17	45936424	45936544	781615_44389461_MAPT_350	1	+	**good**	120
chr17	45936544	45936664	781615_44389461_MAPT_351	1	+	**good**	120
chr17	45936664	45936784	781615_44389461_MAPT_352	1	+	**good**	120
chr17	45936784	45936904	781615_44389461_MAPT_353	1	+	**good**	120
chr17	45936904	45937024	781615_44389461_MAPT_354	1	+	**good**	120
chr17	45937024	45937144	781615_44389461_MAPT_355	1	+	**good**	120
chr17	45937144	45937264	781615_44389461_MAPT_356	1	+	**good**	120
chr17	45937624	45937744	781615_44389461_MAPT_360	1	+	**good**	120
chr17	45937744	45937864	781615_44389461_MAPT_361	1	+	**good**	120
chr17	45937864	45937984	781615_44389461_MAPT_362	1	+	**good**	120
chr17	45937984	45938104	781615_44389461_MAPT_363	1	+	**good**	120
chr17	45939304	45939424	781615_44389461_MAPT_374	1	+	**good**	120
chr17	45939424	45939544	781615_44389461_MAPT_375	1	+	**good**	120
chr17	45939544	45939664	781615_44389461_MAPT_376	1	+	**good**	120
chr17	45939664	45939784	781615_44389461_MAPT_377	1	+	**good**	120
chr17	45939784	45939904	781615_44389461_MAPT_378	1	+	**good**	120
chr17	45939904	45940024	781615_44389461_MAPT_379	1	+	**good**	120
chr17	45940024	45940144	781615_44389461_MAPT_380	1	+	**good**	120
chr17	45940144	45940264	781615_44389461_MAPT_381	1	+	**good**	120
chr17	45940264	45940384	781615_44389461_MAPT_382	1	+	**good**	120
chr17	45940384	45940504	781615_44389461_MAPT_383	1	+	**good**	120
chr17	45940504	45940624	781615_44389461_MAPT_384	1	+	**good**	120
chr17	45940624	45940744	781615_44389461_MAPT_385	1	+	**good**	120
chr17	45940744	45940864	781615_44389461_MAPT_386	1	+	**good**	120
chr17	45940864	45940984	781615_44389461_MAPT_387	1	+	**good**	120
chr17	45940984	45941104	781615_44389461_MAPT_388	1	+	**good**	120
chr17	45941104	45941224	781615_44389461_MAPT_389	1	+	**good**	120
chr17	45941224	45941344	781615_44389461_MAPT_390	1	+	**good**	120
chr17	45941344	45941464	781615_44389461_MAPT_391	1	+	**good**	120
chr17	45941464	45941584	781615_44389461_MAPT_392	1	+	**good**	120
chr17	45941824	45941944	781615_44389461_MAPT_395	1	+	**good**	120
chr17	45941944	45942064	781615_44389461_MAPT_396	1	+	**good**	120
chr17	45942064	45942184	781615_44389461_MAPT_397	1	+	**good**	120
chr17	45942184	45942304	781615_44389461_MAPT_398	1	+	**good**	120
chr17	45942304	45942424	781615_44389461_MAPT_399	1	+	**good**	120
chr17	45942424	45942544	781615_44389461_MAPT_400	1	+	**good**	120
chr17	45942544	45942664	781615_44389461_MAPT_401	1	+	**good**	120

chr17	45942664	45942784	781615_44389461_MAPT_402	1	+	**good**	120
chr17	45942784	45942904	781615_44389461_MAPT_403	1	+	**good**	120
chr17	45942904	45943024	781615_44389461_MAPT_404	1	+	**good**	120
chr17	45943384	45943504	781615_44389461_MAPT_408	1	+	**good**	120
chr17	45943504	45943624	781615_44389461_MAPT_409	1	+	**good**	120
chr17	45943624	45943744	781615_44389461_MAPT_410	1	+	**good**	120
chr17	45943744	45943864	781615_44389461_MAPT_411	1	+	**good**	120
chr17	45943864	45943984	781615_44389461_MAPT_412	1	+	**good**	120
chr17	45943984	45944104	781615_44389461_MAPT_413	1	+	**good**	120
chr17	45944104	45944224	781615_44389461_MAPT_414	1	+	**good**	120
chr17	45944224	45944344	781615_44389461_MAPT_415	1	+	**good**	120
chr17	45944344	45944464	781615_44389461_MAPT_416	1	+	**good**	120
chr17	45944464	45944584	781615_44389461_MAPT_417	1	+	**good**	120
chr17	45944584	45944704	781615_44389461_MAPT_418	1	+	**good**	120
chr17	45944704	45944824	781615_44389461_MAPT_419	1	+	**good**	120
chr17	45944824	45944944	781615_44389461_MAPT_420	1	+	**good**	120
chr17	45944944	45945064	781615_44389461_MAPT_421	1	+	**good**	120
chr17	45945064	45945184	781615_44389461_MAPT_422	1	+	**good**	120
chr17	45945184	45945304	781615_44389461_MAPT_423	1	+	**good**	120
chr17	45945304	45945424	781615_44389461_MAPT_424	1	+	**good**	120
chr17	45945424	45945544	781615_44389461_MAPT_425	1	+	**good**	120
chr17	45945784	45945904	781615_44389461_MAPT_428	3	+	**good**	120
chr17	45945904	45946024	781615_44389461_MAPT_429	1	+	**good**	120
chr17	45946024	45946144	781615_44389461_MAPT_430	1	+	**good**	120
chr17	45946144	45946264	781615_44389461_MAPT_431	1	+	**good**	120
chr17	45946624	45946744	781615_44389461_MAPT_435	1	+	**good**	120
chr17	45946744	45946864	781615_44389461_MAPT_436	1	+	**good**	120
chr17	45946864	45946984	781615_44389461_MAPT_437	1	+	**good**	120
chr17	45946984	45947104	781615_44389461_MAPT_438	1	+	**good**	120
chr17	45947104	45947224	781615_44389461_MAPT_439	1	+	**good**	120
chr17	45947224	45947344	781615_44389461_MAPT_440	1	+	**good**	120
chr17	45947704	45947824	781615_44389461_MAPT_444	2	+	**good**	120
chr17	45947824	45947944	781615_44389461_MAPT_445	2	+	**good**	120
chr17	45948304	45948424	781615_44389461_MAPT_449	1	+	**good**	120
chr17	45948424	45948544	781615_44389461_MAPT_450	1	+	**good**	120
chr17	45948544	45948664	781615_44389461_MAPT_451	1	+	**good**	120
chr17	45948664	45948784	781615_44389461_MAPT_452	1	+	**good**	120
chr17	45948784	45948904	781615_44389461_MAPT_453	1	+	**good**	120
chr17	45948904	45949024	781615_44389461_MAPT_454	1	+	**good**	120
chr17	45949024	45949144	781615_44389461_MAPT_455	1	+	**good**	120
chr17	45949144	45949264	781615_44389461_MAPT_456	1	+	**good**	120
chr17	45949264	45949384	781615_44389461_MAPT_457	1	+	**good**	120
chr17	45949384	45949504	781615_44389461_MAPT_458	1	+	**good**	120
chr17	45949504	45949624	781615_44389461_MAPT_459	2	+	**good**	120
chr17	45949624	45949744	781615_44389461_MAPT_460	1	+	**good**	120
chr17	45949744	45949864	781615_44389461_MAPT_461	1	+	**good**	120

chr17	45949864	45949984	781615_44389461_MAPT_462	1	+	**good**	120
chr17	45949984	45950104	781615_44389461_MAPT_463	1	+	**good**	120
chr17	45950104	45950224	781615_44389461_MAPT_464	1	+	**good**	120
chr17	45950224	45950344	781615_44389461_MAPT_465	1	+	**good**	120
chr17	45950344	45950464	781615_44389461_MAPT_466	1	+	**good**	120
chr17	45950464	45950584	781615_44389461_MAPT_467	1	+	**good**	120
chr17	45950944	45951064	781615_44389461_MAPT_471	1	+	**good**	120
chr17	45951184	45951304	781615_44389461_MAPT_473	42	+	**good**	120
chr17	45951304	45951424	781615_44389461_MAPT_474	1	+	**good**	120
chr17	45951424	45951544	781615_44389461_MAPT_475	1	+	**good**	120
chr17	45951544	45951664	781615_44389461_MAPT_476	1	+	**good**	120
chr17	45951664	45951784	781615_44389461_MAPT_477	1	+	**good**	120
chr17	45951784	45951904	781615_44389461_MAPT_478	1	+	**good**	120
chr17	45951904	45952024	781615_44389461_MAPT_479	1	+	**good**	120
chr17	45952024	45952144	781615_44389461_MAPT_480	1	+	**good**	120
chr17	45952144	45952264	781615_44389461_MAPT_481	1	+	**good**	120
chr17	45952264	45952384	781615_44389461_MAPT_482	1	+	**good**	120
chr17	45952384	45952504	781615_44389461_MAPT_483	1	+	**good**	120
chr17	45952744	45952864	781615_44389461_MAPT_486	1	+	**good**	120
chr17	45952864	45952984	781615_44389461_MAPT_487	1	+	**good**	120
chr17	45952984	45953104	781615_44389461_MAPT_488	1	+	**good**	120
chr17	45953104	45953224	781615_44389461_MAPT_489	1	+	**good**	120
chr17	45953224	45953344	781615_44389461_MAPT_490	1	+	**good**	120
chr17	45953344	45953464	781615_44389461_MAPT_491	1	+	**good**	120
chr17	45953464	45953584	781615_44389461_MAPT_492	1	+	**good**	120
chr17	45953584	45953704	781615_44389461_MAPT_493	1	+	**good**	120
chr17	45953704	45953824	781615_44389461_MAPT_494	1	+	**good**	120
chr17	45953824	45953944	781615_44389461_MAPT_495	1	+	**good**	120
chr17	45953944	45954064	781615_44389461_MAPT_496	1	+	**good**	120
chr17	45954064	45954184	781615_44389461_MAPT_497	19	+	**good**	120
chr17	45954184	45954304	781615_44389461_MAPT_498	12	+	**good**	120
chr17	45955024	45955144	781615_44389461_MAPT_505	1	+	**good**	120
chr17	45955144	45955264	781615_44389461_MAPT_506	1	+	**good**	120
chr17	45955264	45955384	781615_44389461_MAPT_507	1	+	**good**	120
chr17	45955384	45955504	781615_44389461_MAPT_508	1	+	**good**	120
chr17	45955504	45955624	781615_44389461_MAPT_509	1	+	**good**	120
chr17	45955624	45955744	781615_44389461_MAPT_510	1	+	**good**	120
chr17	45956104	45956224	781615_44389461_MAPT_514	1	+	**good**	120
chr17	45956224	45956344	781615_44389461_MAPT_515	1	+	**good**	120
chr17	45956344	45956464	781615_44389461_MAPT_516	1	+	**good**	120
chr17	45956464	45956584	781615_44389461_MAPT_517	1	+	**good**	120
chr17	45957424	45957544	781615_44389461_MAPT_525	1	+	**good**	120
chr17	45957544	45957664	781615_44389461_MAPT_526	1	+	**good**	120
chr17	45957664	45957784	781615_44389461_MAPT_527	1	+	**good**	120
chr17	45957784	45957904	781615_44389461_MAPT_528	1	+	**good**	120
chr17	45957904	45958024	781615_44389461_MAPT_529	1	+	**good**	120

chr17	45958024	45958144	781615_44389461_MAPT_530	1	+	**good**	120
chr17	45958144	45958264	781615_44389461_MAPT_531	1	+	**good**	120
chr17	45958264	45958384	781615_44389461_MAPT_532	1	+	**good**	120
chr17	45959344	45959464	781615_44389461_MAPT_541	1	+	**good**	120
chr17	45959824	45959944	781615_44389461_MAPT_545	1	+	**good**	120
chr17	45959944	45960064	781615_44389461_MAPT_546	1	+	**good**	120
chr17	45960064	45960184	781615_44389461_MAPT_547	1	+	**good**	120
chr17	45960184	45960304	781615_44389461_MAPT_548	1	+	**good**	120
chr17	45960304	45960424	781615_44389461_MAPT_549	1	+	**good**	120
chr17	45960424	45960544	781615_44389461_MAPT_550	1	+	**good**	120
chr17	45960544	45960664	781615_44389461_MAPT_551	1	+	**good**	120
chr17	45960784	45960904	781615_44389461_MAPT_553	1	+	**good**	120
chr17	45960904	45961024	781615_44389461_MAPT_554	1	+	**good**	120
chr17	45961024	45961144	781615_44389461_MAPT_555	1	+	**good**	120
chr17	45961144	45961264	781615_44389461_MAPT_556	31	+	**good**	120
chr17	45961264	45961384	781615_44389461_MAPT_557	1	+	**good**	120
chr17	45961384	45961504	781615_44389461_MAPT_558	1	+	**good**	120
chr17	45961504	45961624	781615_44389461_MAPT_559	1	+	**good**	120
chr17	45961624	45961744	781615_44389461_MAPT_560	1	+	**good**	120
chr17	45962104	45962224	781615_44389461_MAPT_564	1	+	**good**	120
chr17	45962224	45962344	781615_44389461_MAPT_565	1	+	**good**	120
chr17	45962344	45962464	781615_44389461_MAPT_566	1	+	**good**	120
chr17	45962464	45962584	781615_44389461_MAPT_567	1	+	**good**	120
chr17	45962944	45963064	781615_44389461_MAPT_571	1	+	**good**	120
chr17	45963064	45963184	781615_44389461_MAPT_572	1	+	**good**	120
chr17	45963184	45963304	781615_44389461_MAPT_573	1	+	**good**	120
chr17	45963304	45963424	781615_44389461_MAPT_574	1	+	**good**	120
chr17	45963424	45963544	781615_44389461_MAPT_575	1	+	**good**	120
chr17	45963544	45963664	781615_44389461_MAPT_576	1	+	**good**	120
chr17	45963664	45963784	781615_44389461_MAPT_577	1	+	**good**	120
chr17	45963784	45963904	781615_44389461_MAPT_578	1	+	**good**	120
chr17	45963904	45964024	781615_44389461_MAPT_579	1	+	**good**	120
chr17	45964024	45964144	781615_44389461_MAPT_580	1	+	**good**	120
chr17	45964744	45964864	781615_44389461_MAPT_586	1	+	**good**	120
chr17	45964864	45964984	781615_44389461_MAPT_587	1	+	**good**	120
chr17	45964984	45965104	781615_44389461_MAPT_588	1	+	**good**	120
chr17	45965704	45965824	781615_44389461_MAPT_594	1	+	**good**	120
chr17	45965824	45965944	781615_44389461_MAPT_595	1	+	**good**	120
chr17	45965944	45966064	781615_44389461_MAPT_596	1	+	**good**	120
chr17	45966064	45966184	781615_44389461_MAPT_597	1	+	**good**	120
chr17	45966184	45966304	781615_44389461_MAPT_598	46	+	**good**	120
chr17	45966664	45966784	781615_44389461_MAPT_602	1	+	**good**	120
chr17	45966784	45966904	781615_44389461_MAPT_603	1	+	**good**	120
chr17	45966904	45967024	781615_44389461_MAPT_604	1	+	**good**	120
chr17	45967024	45967144	781615_44389461_MAPT_605	1	+	**good**	120
chr17	45967144	45967264	781615_44389461_MAPT_606	1	+	**good**	120

chr17	45967264	45967384	781615_44389461_MAPT_607	1	+	**good**	120
chr17	45967384	45967504	781615_44389461_MAPT_608	1	+	**good**	120
chr17	45967504	45967624	781615_44389461_MAPT_609	1	+	**good**	120
chr17	45967624	45967744	781615_44389461_MAPT_610	1	+	**good**	120
chr17	45967744	45967864	781615_44389461_MAPT_611	1	+	**good**	120
chr17	45967864	45967984	781615_44389461_MAPT_612	1	+	**good**	120
chr17	45967984	45968104	781615_44389461_MAPT_613	1	+	**good**	120
chr17	45968104	45968224	781615_44389461_MAPT_614	1	+	**good**	120
chr17	45968224	45968344	781615_44389461_MAPT_615	1	+	**good**	120
chr17	45968344	45968464	781615_44389461_MAPT_616	1	+	**good**	120
chr17	45968464	45968584	781615_44389461_MAPT_617	1	+	**good**	120
chr17	45968584	45968704	781615_44389461_MAPT_618	1	+	**good**	120
chr17	45968704	45968824	781615_44389461_MAPT_619	1	+	**good**	120
chr17	45968824	45968944	781615_44389461_MAPT_620	1	+	**good**	120
chr17	45968944	45969064	781615_44389461_MAPT_621	4	+	**good**	120
chr17	45969064	45969184	781615_44389461_MAPT_622	1	+	**good**	120
chr17	45969184	45969304	781615_44389461_MAPT_623	1	+	**good**	120
chr17	45969304	45969424	781615_44389461_MAPT_624	28	+	**good**	120
chr17	45969784	45969904	781615_44389461_MAPT_628	46	+	**good**	120
chr17	45970144	45970264	781615_44389461_MAPT_631	1	+	**good**	120
chr17	45970264	45970384	781615_44389461_MAPT_632	7	+	**good**	120
chr17	45970384	45970504	781615_44389461_MAPT_633	1	+	**good**	120
chr17	45970504	45970624	781615_44389461_MAPT_634	1	+	**good**	120
chr17	45970624	45970744	781615_44389461_MAPT_635	2	+	**good**	120
chr17	45970744	45970864	781615_44389461_MAPT_636	1	+	**good**	120
chr17	45970864	45970984	781615_44389461_MAPT_637	1	+	**good**	120
chr17	45970984	45971104	781615_44389461_MAPT_638	1	+	**good**	120
chr17	45971104	45971224	781615_44389461_MAPT_639	1	+	**good**	120
chr17	45971224	45971344	781615_44389461_MAPT_640	1	+	**good**	120
chr17	45971344	45971464	781615_44389461_MAPT_641	1	+	**good**	120
chr17	45971464	45971584	781615_44389461_MAPT_642	1	+	**good**	120
chr17	45971584	45971704	781615_44389461_MAPT_643	1	+	**good**	120
chr17	45971704	45971824	781615_44389461_MAPT_644	1	+	**good**	120
chr17	45971824	45971944	781615_44389461_MAPT_645	1	+	**good**	120
chr17	45971944	45972064	781615_44389461_MAPT_646	1	+	**good**	120
chr17	45972064	45972184	781615_44389461_MAPT_647	1	+	**good**	120
chr17	45972184	45972304	781615_44389461_MAPT_648	1	+	**good**	120
chr17	45972304	45972424	781615_44389461_MAPT_649	1	+	**good**	120
chr17	45972424	45972544	781615_44389461_MAPT_650	1	+	**good**	120
chr17	45972544	45972664	781615_44389461_MAPT_651	1	+	**good**	120
chr17	45972664	45972784	781615_44389461_MAPT_652	2	+	**good**	120
chr17	45972784	45972904	781615_44389461_MAPT_653	1	+	**good**	120
chr17	45972904	45973024	781615_44389461_MAPT_654	1	+	**good**	120
chr17	45973024	45973144	781615_44389461_MAPT_655	1	+	**good**	120
chr17	45973144	45973264	781615_44389461_MAPT_656	1	+	**good**	120
chr17	45973264	45973384	781615_44389461_MAPT_657	1	+	**good**	120

chr17	45973384	45973504	781615_44389461_MAPT_658	1	+	**good**	120
chr17	45973504	45973624	781615_44389461_MAPT_659	1	+	**good**	120
chr17	45973624	45973744	781615_44389461_MAPT_660	1	+	**good**	120
chr17	45973744	45973864	781615_44389461_MAPT_661	1	+	**good**	120
chr17	45973864	45973984	781615_44389461_MAPT_662	1	+	**good**	120
chr17	45973984	45974104	781615_44389461_MAPT_663	1	+	**good**	120
chr17	45974104	45974224	781615_44389461_MAPT_664	1	+	**good**	120
chr17	45974224	45974344	781615_44389461_MAPT_665	1	+	**good**	120
chr17	45974344	45974464	781615_44389461_MAPT_666	1	+	**good**	120
chr17	45974464	45974584	781615_44389461_MAPT_667	1	+	**good**	120
chr17	45974584	45974704	781615_44389461_MAPT_668	1	+	**good**	120
chr17	45974704	45974824	781615_44389461_MAPT_669	1	+	**good**	120
chr17	45974824	45974944	781615_44389461_MAPT_670	1	+	**good**	120
chr17	45974944	45975064	781615_44389461_MAPT_671	1	+	**good**	120
chr17	45975064	45975184	781615_44389461_MAPT_672	1	+	**good**	120
chr17	45975184	45975304	781615_44389461_MAPT_673	1	+	**good**	120
chr17	45975304	45975424	781615_44389461_MAPT_674	1	+	**good**	120
chr17	45975424	45975544	781615_44389461_MAPT_675	1	+	**good**	120
chr17	45975544	45975664	781615_44389461_MAPT_676	1	+	**good**	120
chr17	45975664	45975784	781615_44389461_MAPT_677	1	+	**good**	120
chr17	45975784	45975904	781615_44389461_MAPT_678	1	+	**good**	120
chr17	45975904	45976024	781615_44389461_MAPT_679	1	+	**good**	120
chr17	45976024	45976144	781615_44389461_MAPT_680	1	+	**good**	120
chr17	45976144	45976264	781615_44389461_MAPT_681	1	+	**good**	120
chr17	45976264	45976384	781615_44389461_MAPT_682	1	+	**good**	120
chr17	45976384	45976504	781615_44389461_MAPT_683	1	+	**good**	120
chr17	45976504	45976624	781615_44389461_MAPT_684	1	+	**good**	120
chr17	45976624	45976744	781615_44389461_MAPT_685	1	+	**good**	120
chr17	45976744	45976864	781615_44389461_MAPT_686	1	+	**good**	120
chr17	45976864	45976984	781615_44389461_MAPT_687	1	+	**good**	120
chr17	45976984	45977104	781615_44389461_MAPT_688	1	+	**good**	120
chr17	45977104	45977224	781615_44389461_MAPT_689	1	+	**good**	120
chr17	45977224	45977344	781615_44389461_MAPT_690	1	+	**good**	120
chr17	45977344	45977464	781615_44389461_MAPT_691	1	+	**good**	120
chr17	45977464	45977584	781615_44389461_MAPT_692	1	+	**good**	120
chr17	45977584	45977704	781615_44389461_MAPT_693	1	+	**good**	120
chr17	45977704	45977824	781615_44389461_MAPT_694	1	+	**good**	120
chr17	45977824	45977944	781615_44389461_MAPT_695	1	+	**good**	120
chr17	45977944	45978064	781615_44389461_MAPT_696	1	+	**good**	120
chr17	45978064	45978184	781615_44389461_MAPT_697	1	+	**good**	120
chr17	45978184	45978304	781615_44389461_MAPT_698	1	+	**good**	120
chr17	45978304	45978424	781615_44389461_MAPT_699	1	+	**good**	120
chr17	45978424	45978544	781615_44389461_MAPT_700	1	+	**good**	120
chr17	45978544	45978664	781615_44389461_MAPT_701	1	+	**good**	120
chr17	45979024	45979144	781615_44389461_MAPT_705	1	+	**good**	120
chr17	45979144	45979264	781615_44389461_MAPT_706	1	+	**good**	120

chr17	45979264	45979384	781615_44389461_MAPT_707	1	+	**good**	120
chr17	45979384	45979504	781615_44389461_MAPT_708	1	+	**good**	120
chr17	45979504	45979624	781615_44389461_MAPT_709	1	+	**good**	120
chr17	45979624	45979744	781615_44389461_MAPT_710	1	+	**good**	120
chr17	45979744	45979864	781615_44389461_MAPT_711	1	+	**good**	120
chr17	45979864	45979984	781615_44389461_MAPT_712	1	+	**good**	120
chr17	45979984	45980104	781615_44389461_MAPT_713	1	+	**good**	120
chr17	45980104	45980224	781615_44389461_MAPT_714	1	+	**good**	120
chr17	45980584	45980704	781615_44389461_MAPT_718	1	+	**good**	120
chr17	45980704	45980824	781615_44389461_MAPT_719	1	+	**good**	120
chr17	45980824	45980944	781615_44389461_MAPT_720	1	+	**good**	120
chr17	45980944	45981064	781615_44389461_MAPT_721	1	+	**good**	120
chr17	45981064	45981184	781615_44389461_MAPT_722	1	+	**good**	120
chr17	45981184	45981304	781615_44389461_MAPT_723	1	+	**good**	120
chr17	45981304	45981424	781615_44389461_MAPT_724	1	+	**good**	120
chr17	45981424	45981544	781615_44389461_MAPT_725	1	+	**good**	120
chr17	45981544	45981664	781615_44389461_MAPT_726	1	+	**good**	120
chr17	45981664	45981784	781615_44389461_MAPT_727	1	+	**good**	120
chr17	45982384	45982504	781615_44389461_MAPT_733	1	+	**good**	120
chr17	45982504	45982624	781615_44389461_MAPT_734	1	+	**good**	120
chr17	45982624	45982744	781615_44389461_MAPT_735	1	+	**good**	120
chr17	45982744	45982864	781615_44389461_MAPT_736	1	+	**good**	120
chr17	45982864	45982984	781615_44389461_MAPT_737	1	+	**good**	120
chr17	45982984	45983104	781615_44389461_MAPT_738	1	+	**good**	120
chr17	45983104	45983224	781615_44389461_MAPT_739	1	+	**good**	120
chr17	45983224	45983344	781615_44389461_MAPT_740	1	+	**good**	120
chr17	45983344	45983464	781615_44389461_MAPT_741	1	+	**good**	120
chr17	45983464	45983584	781615_44389461_MAPT_742	1	+	**good**	120
chr17	45983584	45983704	781615_44389461_MAPT_743	1	+	**good**	120
chr17	45983704	45983824	781615_44389461_MAPT_744	1	+	**good**	120
chr17	45983824	45983944	781615_44389461_MAPT_745	1	+	**good**	120
chr17	45983944	45984064	781615_44389461_MAPT_746	1	+	**good**	120
chr17	45984064	45984184	781615_44389461_MAPT_747	1	+	**good**	120
chr17	45984184	45984304	781615_44389461_MAPT_748	1	+	**good**	120
chr17	45984304	45984424	781615_44389461_MAPT_749	4	+	**good**	120
chr17	45984424	45984544	781615_44389461_MAPT_750	1	+	**good**	120
chr17	45984544	45984664	781615_44389461_MAPT_751	1	+	**good**	120
chr17	45984664	45984784	781615_44389461_MAPT_752	1	+	**good**	120
chr17	45984784	45984904	781615_44389461_MAPT_753	1	+	**good**	120
chr17	45984904	45985024	781615_44389461_MAPT_754	1	+	**good**	120
chr17	45985384	45985504	781615_44389461_MAPT_758	1	+	**good**	120
chr17	45985504	45985624	781615_44389461_MAPT_759	1	+	**good**	120
chr17	45985624	45985744	781615_44389461_MAPT_760	1	+	**good**	120
chr17	45985744	45985864	781615_44389461_MAPT_761	1	+	**good**	120
chr17	45985864	45985984	781615_44389461_MAPT_762	1	+	**good**	120
chr17	45985984	45986104	781615_44389461_MAPT_763	1	+	**good**	120

chr17	45986104	45986224	781615_44389461_MAPT_764	1	+	**good**	120
chr17	45986224	45986344	781615_44389461_MAPT_765	1	+	**good**	120
chr17	45986344	45986464	781615_44389461_MAPT_766	1	+	**good**	120
chr17	45986464	45986584	781615_44389461_MAPT_767	1	+	**good**	120
chr17	45986584	45986704	781615_44389461_MAPT_768	1	+	**good**	120
chr17	45986704	45986824	781615_44389461_MAPT_769	1	+	**good**	120
chr17	45986824	45986944	781615_44389461_MAPT_770	1	+	**good**	120
chr17	45986944	45987064	781615_44389461_MAPT_771	1	+	**good**	120
chr17	45987064	45987184	781615_44389461_MAPT_772	1	+	**good**	120
chr17	45987184	45987304	781615_44389461_MAPT_773	1	+	**good**	120
chr17	45987304	45987424	781615_44389461_MAPT_774	1	+	**good**	120
chr17	45987424	45987544	781615_44389461_MAPT_775	1	+	**good**	120
chr17	45987544	45987664	781615_44389461_MAPT_776	1	+	**good**	120
chr17	45987664	45987784	781615_44389461_MAPT_777	1	+	**good**	120
chr17	45987784	45987904	781615_44389461_MAPT_778	1	+	**good**	120
chr17	45987904	45988024	781615_44389461_MAPT_779	1	+	**good**	120
chr17	45988024	45988144	781615_44389461_MAPT_780	1	+	**good**	120
chr17	45988144	45988264	781615_44389461_MAPT_781	1	+	**good**	120
chr17	45988264	45988384	781615_44389461_MAPT_782	1	+	**good**	120
chr17	45988384	45988504	781615_44389461_MAPT_783	1	+	**good**	120
chr17	45988504	45988624	781615_44389461_MAPT_784	1	+	**good**	120
chr17	45988984	45989104	781615_44389461_MAPT_788	1	+	**good**	120
chr17	45989104	45989224	781615_44389461_MAPT_789	1	+	**good**	120
chr17	45989224	45989344	781615_44389461_MAPT_790	1	+	**good**	120
chr17	45989704	45989824	781615_44389461_MAPT_794	1	+	**good**	120
chr17	45989824	45989944	781615_44389461_MAPT_795	1	+	**good**	120
chr17	45989944	45990064	781615_44389461_MAPT_796	1	+	**good**	120
chr17	45990064	45990184	781615_44389461_MAPT_797	1	+	**good**	120
chr17	45990184	45990304	781615_44389461_MAPT_798	1	+	**good**	120
chr17	45990544	45990664	781615_44389461_MAPT_801	2	+	**good**	120
chr17	45990664	45990784	781615_44389461_MAPT_802	1	+	**good**	120
chr17	45990784	45990904	781615_44389461_MAPT_803	1	+	**good**	120
chr17	45990904	45991024	781615_44389461_MAPT_804	1	+	**good**	120
chr17	45991024	45991144	781615_44389461_MAPT_805	1	+	**good**	120
chr17	45991144	45991264	781615_44389461_MAPT_806	1	+	**good**	120
chr17	45991264	45991384	781615_44389461_MAPT_807	1	+	**good**	120
chr17	45991384	45991504	781615_44389461_MAPT_808	1	+	**good**	120
chr17	45991504	45991624	781615_44389461_MAPT_809	1	+	**good**	120
chr17	45991624	45991744	781615_44389461_MAPT_810	1	+	**good**	120
chr17	45992104	45992224	781615_44389461_MAPT_814	1	+	**good**	120
chr17	45992224	45992344	781615_44389461_MAPT_815	1	+	**good**	120
chr17	45992344	45992464	781615_44389461_MAPT_816	1	+	**good**	120
chr17	45992464	45992584	781615_44389461_MAPT_817	1	+	**good**	120
chr17	45992944	45993064	781615_44389461_MAPT_821	1	+	**good**	120
chr17	45993064	45993184	781615_44389461_MAPT_822	1	+	**good**	120
chr17	45993184	45993304	781615_44389461_MAPT_823	1	+	**good**	120

chr17	45993304	45993424	781615_44389461_MAPT_824	1	+	**good**	120
chr17	45993664	45993784	781615_44389461_MAPT_827	1	+	**good**	120
chr17	45993784	45993904	781615_44389461_MAPT_828	1	+	**good**	120
chr17	45993904	45994024	781615_44389461_MAPT_829	1	+	**good**	120
chr17	45994024	45994144	781615_44389461_MAPT_830	1	+	**good**	120
chr17	45994144	45994264	781615_44389461_MAPT_831	2	+	**good**	120
chr17	45994264	45994384	781615_44389461_MAPT_832	1	+	**good**	120
chr17	45994384	45994504	781615_44389461_MAPT_833	1	+	**good**	120
chr17	45994624	45994744	781615_44389461_MAPT_835	33	+	**good**	120
chr17	45995104	45995224	781615_44389461_MAPT_839	1	+	**good**	120
chr17	45995224	45995344	781615_44389461_MAPT_840	1	+	**good**	120
chr17	45995344	45995464	781615_44389461_MAPT_841	1	+	**good**	120
chr17	45995464	45995584	781615_44389461_MAPT_842	1	+	**good**	120
chr17	45995584	45995704	781615_44389461_MAPT_843	1	+	**good**	120
chr17	45995704	45995824	781615_44389461_MAPT_844	1	+	**good**	120
chr17	45995824	45995944	781615_44389461_MAPT_845	1	+	**good**	120
chr17	45995944	45996064	781615_44389461_MAPT_846	1	+	**good**	120
chr17	45996064	45996184	781615_44389461_MAPT_847	1	+	**good**	120
chr17	45996184	45996304	781615_44389461_MAPT_848	1	+	**good**	120
chr17	45996304	45996424	781615_44389461_MAPT_849	1	+	**good**	120
chr17	45996424	45996544	781615_44389461_MAPT_850	1	+	**good**	120
chr17	45996544	45996664	781615_44389461_MAPT_851	1	+	**good**	120
chr17	45996664	45996784	781615_44389461_MAPT_852	1	+	**good**	120
chr17	45996784	45996904	781615_44389461_MAPT_853	1	+	**good**	120
chr17	45996904	45997024	781615_44389461_MAPT_854	1	+	**good**	120
chr17	45997024	45997144	781615_44389461_MAPT_855	1	+	**good**	120
chr17	45997144	45997264	781615_44389461_MAPT_856	1	+	**good**	120
chr17	45997264	45997384	781615_44389461_MAPT_857	1	+	**good**	120
chr17	45997384	45997504	781615_44389461_MAPT_858	1	+	**good**	120
chr17	45997744	45997864	781615_44389461_MAPT_861	6	+	**good**	120
chr17	45997864	45997984	781615_44389461_MAPT_862	1	+	**good**	120
chr17	45997984	45998104	781615_44389461_MAPT_863	1	+	**good**	120
chr17	45998104	45998224	781615_44389461_MAPT_864	1	+	**good**	120
chr17	45998224	45998344	781615_44389461_MAPT_865	1	+	**good**	120
chr17	45998344	45998464	781615_44389461_MAPT_866	1	+	**good**	120
chr17	45998464	45998584	781615_44389461_MAPT_867	1	+	**good**	120
chr17	45998584	45998704	781615_44389461_MAPT_868	1	+	**good**	120
chr17	45998704	45998824	781615_44389461_MAPT_869	1	+	**good**	120
chr17	45998824	45998944	781615_44389461_MAPT_870	1	+	**good**	120
chr17	45998944	45999064	781615_44389461_MAPT_871	1	+	**good**	120
chr17	45999064	45999184	781615_44389461_MAPT_872	1	+	**good**	120
chr17	45999184	45999304	781615_44389461_MAPT_873	1	+	**good**	120
chr17	45999304	45999424	781615_44389461_MAPT_874	1	+	**good**	120
chr17	45999424	45999544	781615_44389461_MAPT_875	1	+	**good**	120
chr17	45999544	45999664	781615_44389461_MAPT_876	1	+	**good**	120
chr17	45999664	45999784	781615_44389461_MAPT_877	1	+	**good**	120

chr17	45999784	45999904	781615_44389461_MAPT_878	1	+	**good**	120
chr17	45999904	46000024	781615_44389461_MAPT_879	1	+	**good**	120
chr17	46000024	46000144	781615_44389461_MAPT_880	1	+	**good**	120
chr17	46000144	46000264	781615_44389461_MAPT_881	1	+	**good**	120
chr17	46000264	46000384	781615_44389461_MAPT_882	1	+	**good**	120
chr17	46000384	46000504	781615_44389461_MAPT_883	1	+	**good**	120
chr17	46000504	46000624	781615_44389461_MAPT_884	1	+	**good**	120
chr17	46000624	46000744	781615_44389461_MAPT_885	1	+	**good**	120
chr17	46000744	46000864	781615_44389461_MAPT_886	1	+	**good**	120
chr17	46000864	46000984	781615_44389461_MAPT_887	1	+	**good**	120
chr17	46001224	46001344	781615_44389461_MAPT_890	1	+	**good**	120
chr17	46001344	46001464	781615_44389461_MAPT_891	1	+	**good**	120
chr17	46001704	46001824	781615_44389461_MAPT_894	1	+	**good**	120
chr17	46001824	46001944	781615_44389461_MAPT_895	1	+	**good**	120
chr17	46001944	46002064	781615_44389461_MAPT_896	1	+	**good**	120
chr17	46002064	46002184	781615_44389461_MAPT_897	1	+	**good**	120
chr17	46002184	46002304	781615_44389461_MAPT_898	1	+	**good**	120
chr17	46002304	46002424	781615_44389461_MAPT_899	1	+	**good**	120
chr17	46002424	46002544	781615_44389461_MAPT_900	1	+	**good**	120
chr17	46002544	46002664	781615_44389461_MAPT_901	1	+	**good**	120
chr17	46002664	46002784	781615_44389461_MAPT_902	1	+	**good**	120
chr17	46003144	46003264	781615_44389461_MAPT_906	1	+	**good**	120
chr17	46003624	46003744	781615_44389461_MAPT_910	1	+	**good**	120
chr17	46003744	46003864	781615_44389461_MAPT_911	1	+	**good**	120
chr17	46003864	46003984	781615_44389461_MAPT_912	1	+	**good**	120
chr17	46003984	46004104	781615_44389461_MAPT_913	1	+	**good**	120
chr17	46004104	46004224	781615_44389461_MAPT_914	1	+	**good**	120
chr17	46004224	46004344	781615_44389461_MAPT_915	1	+	**good**	120
chr17	46004344	46004464	781615_44389461_MAPT_916	1	+	**good**	120
chr17	46004464	46004584	781615_44389461_MAPT_917	1	+	**good**	120
chr17	46004584	46004704	781615_44389461_MAPT_918	1	+	**good**	120
chr17	46004704	46004824	781615_44389461_MAPT_919	24	+	**good**	120
chr17	46005064	46005184	781615_44389461_MAPT_922	1	+	**good**	120
chr17	46005184	46005304	781615_44389461_MAPT_923	1	+	**good**	120
chr17	46005304	46005424	781615_44389461_MAPT_924	1	+	**good**	120
chr17	46005424	46005544	781615_44389461_MAPT_925	5	+	**good**	120
chr17	46005544	46005664	781615_44389461_MAPT_926	1	+	**good**	120
chr17	46005664	46005784	781615_44389461_MAPT_927	1	+	**good**	120
chr17	46005784	46005904	781615_44389461_MAPT_928	1	+	**good**	120
chr17	46005904	46006024	781615_44389461_MAPT_929	1	+	**good**	120
chr17	46006024	46006144	781615_44389461_MAPT_930	1	+	**good**	120
chr17	46006504	46006624	781615_44389461_MAPT_934	6	+	**good**	120
chr17	46007464	46007584	781615_44389461_MAPT_942	4	+	**good**	120
chr17	46007584	46007704	781615_44389461_MAPT_943	1	+	**good**	120
chr17	46007704	46007824	781615_44389461_MAPT_944	1	+	**good**	120
chr17	46007824	46007944	781615_44389461_MAPT_945	1	+	**good**	120

chr17	46007944	46008064	781615_44389461_MAPT_946	1	+	**good**	120
chr17	46008424	46008544	781615_44389461_MAPT_950	1	+	**good**	120
chr17	46008544	46008664	781615_44389461_MAPT_951	1	+	**good**	120
chr17	46008664	46008784	781615_44389461_MAPT_952	1	+	**good**	120
chr17	46008784	46008904	781615_44389461_MAPT_953	1	+	**good**	120
chr17	46009264	46009384	781615_44389461_MAPT_957	1	+	**good**	120
chr17	46009384	46009504	781615_44389461_MAPT_958	1	+	**good**	120
chr17	46009504	46009624	781615_44389461_MAPT_959	1	+	**good**	120
chr17	46009624	46009744	781615_44389461_MAPT_960	1	+	**good**	120
chr17	46009744	46009864	781615_44389461_MAPT_961	1	+	**good**	120
chr17	46009864	46009984	781615_44389461_MAPT_962	1	+	**good**	120
chr17	46009984	46010104	781615_44389461_MAPT_963	1	+	**good**	120
chr17	46010104	46010224	781615_44389461_MAPT_964	1	+	**good**	120
chr17	46010224	46010344	781615_44389461_MAPT_965	2	+	**good**	120
chr17	46010344	46010464	781615_44389461_MAPT_966	1	+	**good**	120
chr17	46010464	46010584	781615_44389461_MAPT_967	1	+	**good**	120
chr17	46010584	46010704	781615_44389461_MAPT_968	1	+	**good**	120
chr17	46010704	46010824	781615_44389461_MAPT_969	1	+	**good**	120
chr17	46010824	46010944	781615_44389461_MAPT_970	1	+	**good**	120
chr17	46010944	46011064	781615_44389461_MAPT_971	1	+	**good**	120
chr17	46011424	46011544	781615_44389461_MAPT_975	1	+	**good**	120
chr17	46011544	46011664	781615_44389461_MAPT_976	1	+	**good**	120
chr17	46011664	46011784	781615_44389461_MAPT_977	1	+	**good**	120
chr17	46011784	46011904	781615_44389461_MAPT_978	1	+	**good**	120
chr17	46011904	46012024	781615_44389461_MAPT_979	1	+	**good**	120
chr17	46012024	46012144	781615_44389461_MAPT_980	1	+	**good**	120
chr17	46012144	46012264	781615_44389461_MAPT_981	1	+	**good**	120
chr17	46012264	46012384	781615_44389461_MAPT_982	1	+	**good**	120
chr17	46012384	46012504	781615_44389461_MAPT_983	1	+	**good**	120
chr17	46012504	46012624	781615_44389461_MAPT_984	1	+	**good**	120
chr17	46012624	46012744	781615_44389461_MAPT_985	1	+	**good**	120
chr17	46012744	46012864	781615_44389461_MAPT_986	1	+	**good**	120
chr17	46012864	46012984	781615_44389461_MAPT_987	1	+	**good**	120
chr17	46012984	46013104	781615_44389461_MAPT_988	1	+	**good**	120
chr17	46013104	46013224	781615_44389461_MAPT_989	1	+	**good**	120
chr17	46013224	46013344	781615_44389461_MAPT_990	1	+	**good**	120
chr17	46013344	46013464	781615_44389461_MAPT_991	1	+	**good**	120
chr17	46013464	46013584	781615_44389461_MAPT_992	1	+	**good**	120
chr17	46013584	46013704	781615_44389461_MAPT_993	1	+	**good**	120
chr17	46013704	46013824	781615_44389461_MAPT_994	1	+	**good**	120
chr17	46013824	46013944	781615_44389461_MAPT_995	1	+	**good**	120
chr17	46013944	46014064	781615_44389461_MAPT_996	1	+	**good**	120
chr17	46014064	46014184	781615_44389461_MAPT_997	1	+	**good**	120
chr17	46014184	46014304	781615_44389461_MAPT_998	1	+	**good**	120
chr17	46014304	46014424	781615_44389461_MAPT_999	1	+	**good**	120
chr17	46014424	46014544	781615_44389461_MAPT_1000	1	+	**good**	120

chr17	46015864	46015984	781615_44389461_MAPT_1012	1	+	**good**	120
chr17	46015984	46016104	781615_44389461_MAPT_1013	1	+	**good**	120
chr17	46016824	46016944	781615_44389461_MAPT_1020	1	+	**good**	120
chr17	46016944	46017064	781615_44389461_MAPT_1021	1	+	**good**	120
chr17	46017064	46017184	781615_44389461_MAPT_1022	1	+	**good**	120
chr17	46017184	46017304	781615_44389461_MAPT_1023	1	+	**good**	120
chr17	46018144	46018264	781615_44389461_MAPT_1031	1	+	**good**	120
chr17	46018264	46018384	781615_44389461_MAPT_1032	1	+	**good**	120
chr17	46018384	46018504	781615_44389461_MAPT_1033	1	+	**good**	120
chr17	46018504	46018624	781615_44389461_MAPT_1034	1	+	**good**	120
chr17	46018624	46018744	781615_44389461_MAPT_1035	1	+	**good**	120
chr17	46018744	46018864	781615_44389461_MAPT_1036	1	+	**good**	120
chr17	46018864	46018984	781615_44389461_MAPT_1037	1	+	**good**	120
chr17	46020064	46020184	781615_44389461_MAPT_1047	1	+	**good**	120
chr17	46020184	46020304	781615_44389461_MAPT_1048	1	+	**good**	120
chr17	46020304	46020424	781615_44389461_MAPT_1049	1	+	**good**	120
chr17	46020424	46020544	781615_44389461_MAPT_1050	1	+	**good**	120
chr17	46020904	46021024	781615_44389461_MAPT_1054	1	+	**good**	120
chr17	46021024	46021144	781615_44389461_MAPT_1055	1	+	**good**	120
chr17	46021144	46021264	781615_44389461_MAPT_1056	1	+	**good**	120
chr17	46021264	46021384	781615_44389461_MAPT_1057	1	+	**good**	120
chr17	46021384	46021504	781615_44389461_MAPT_1058	1	+	**good**	120
chr17	46021504	46021624	781615_44389461_MAPT_1059	1	+	**good**	120
chr17	46021624	46021744	781615_44389461_MAPT_1060	1	+	**good**	120
chr17	46021744	46021864	781615_44389461_MAPT_1061	1	+	**good**	120
chr17	46021864	46021984	781615_44389461_MAPT_1062	1	+	**good**	120
chr17	46021984	46022104	781615_44389461_MAPT_1063	7	+	**good**	120
chr17	46022344	46022464	781615_44389461_MAPT_1066	5	+	**good**	120
chr17	46022464	46022584	781615_44389461_MAPT_1067	1	+	**good**	120
chr17	46022584	46022704	781615_44389461_MAPT_1068	1	+	**good**	120
chr17	46022704	46022824	781615_44389461_MAPT_1069	1	+	**good**	120
chr17	46022944	46023064	781615_44389461_MAPT_1071	1	+	**good**	120
chr17	46023064	46023184	781615_44389461_MAPT_1072	1	+	**good**	120
chr17	46023184	46023304	781615_44389461_MAPT_1073	1	+	**good**	120
chr17	46023304	46023424	781615_44389461_MAPT_1074	1	+	**good**	120
chr17	46023424	46023544	781615_44389461_MAPT_1075	1	+	**good**	120
chr17	46023904	46024024	781615_44389461_MAPT_1079	1	+	**good**	120
chr17	46024024	46024144	781615_44389461_MAPT_1080	1	+	**good**	120
chr17	46024144	46024264	781615_44389461_MAPT_1081	1	+	**good**	120
chr17	46024264	46024384	781615_44389461_MAPT_1082	1	+	**good**	120
chr17	46024384	46024504	781615_44389461_MAPT_1083	1	+	**good**	120
chr17	46024504	46024624	781615_44389461_MAPT_1084	1	+	**good**	120
chr17	46024624	46024744	781615_44389461_MAPT_1085	1	+	**good**	120
chr17	46024744	46024864	781615_44389461_MAPT_1086	1	+	**good**	120
chr17	46024864	46024984	781615_44389461_MAPT_1087	1	+	**good**	120
chr17	46024984	46025104	781615_44389461_MAPT_1088	1	+	**good**	120

chr17	46025104	46025224	781615_44389461_MAPT_1089	1	+	**good**	120
chr17	46025224	46025344	781615_44389461_MAPT_1090	1	+	**good**	120
chr17	46025344	46025464	781615_44389461_MAPT_1091	1	+	**good**	120
chr17	46025464	46025584	781615_44389461_MAPT_1092	1	+	**good**	120
chr17	46025584	46025704	781615_44389461_MAPT_1093	1	+	**good**	120
chr17	46025704	46025824	781615_44389461_MAPT_1094	1	+	**good**	120
chr17	46025824	46025944	781615_44389461_MAPT_1095	1	+	**good**	120
chr17	46025944	46026064	781615_44389461_MAPT_1096	1	+	**good**	120
chr17	46026064	46026184	781615_44389461_MAPT_1097	1	+	**good**	120
chr17	46026184	46026304	781615_44389461_MAPT_1098	1	+	**good**	120
chr17	46026304	46026424	781615_44389461_MAPT_1099	1	+	**good**	120
chr17	46026424	46026544	781615_44389461_MAPT_1100	1	+	**good**	120
chr17	46026544	46026664	781615_44389461_MAPT_1101	1	+	**good**	120
chr17	46026664	46026784	781615_44389461_MAPT_1102	1	+	**good**	120
chr17	46026784	46026904	781615_44389461_MAPT_1103	1	+	**good**	120
chr17	46026904	46027024	781615_44389461_MAPT_1104	1	+	**good**	120
chr17	46027024	46027144	781615_44389461_MAPT_1105	1	+	**good**	120
chr17	46027144	46027264	781615_44389461_MAPT_1106	1	+	**good**	120
chr17	46027264	46027384	781615_44389461_MAPT_1107	1	+	**good**	120
chr17	46027384	46027504	781615_44389461_MAPT_1108	1	+	**good**	120
chr17	46027504	46027624	781615_44389461_MAPT_1109	1	+	**good**	120
chr17	46027624	46027744	781615_44389461_MAPT_1110	1	+	**good**	120
chr17	46027744	46027864	781615_44389461_MAPT_1111	1	+	**good**	120
chr17	46027864	46027984	781615_44389461_MAPT_1112	1	+	**good**	120
chr17	46027984	46028104	781615_44389461_MAPT_1113	1	+	**good**	120
chr17	46028104	46028224	781615_44389461_MAPT_1114	1	+	**good**	120
chr17	46028224	46028344	781615_44389461_MAPT_1115	1	+	**good**	120

2. The following is the list of "good" probes custom-designed by IDT for the '-' strand from GRCh38 reference genome used in Chapter 7, section 7.2.2.2:

Chromosome	Start	Stop	Name	Score	Strand	Code	Len
chr17	45894484	45894604	781717_44404225_MAPT_1116	1	-	**good**	120
chr17	45894604	45894724	781717_44404225_MAPT_1115	1	-	**good**	120
chr17	45894724	45894844	781717_44404225_MAPT_1114	1	-	**good**	120
chr17	45894844	45894964	781717_44404225_MAPT_1113	1	-	**good**	120
chr17	45894964	45895084	781717_44404225_MAPT_1112	1	-	**good**	120
chr17	45895084	45895204	781717_44404225_MAPT_1111	1	-	**good**	120
chr17	45895204	45895324	781717_44404225_MAPT_1110	1	-	**good**	120
chr17	45895324	45895444	781717_44404225_MAPT_1109	1	-	**good**	120
chr17	45895444	45895564	781717_44404225_MAPT_1108	1	-	**good**	120
chr17	45895564	45895684	781717_44404225_MAPT_1107	1	-	**good**	120
chr17	45895684	45895804	781717_44404225_MAPT_1106	1	-	**good**	120
chr17	45895804	45895924	781717_44404225_MAPT_1105	1	-	**good**	120
chr17	45895924	45896044	781717_44404225_MAPT_1104	1	-	**good**	120

chr17	45896044	45896164	781717_44404225_MAPT_1103	1	-	**good**	120
chr17	45896164	45896284	781717_44404225_MAPT_1102	1	-	**good**	120
chr17	45896284	45896404	781717_44404225_MAPT_1101	1	-	**good**	120
chr17	45896404	45896524	781717_44404225_MAPT_1100	1	-	**good**	120
chr17	45896524	45896644	781717_44404225_MAPT_1099	1	-	**good**	120
chr17	45896644	45896764	781717_44404225_MAPT_1098	1	-	**good**	120
chr17	45896764	45896884	781717_44404225_MAPT_1097	1	-	**good**	120
chr17	45896884	45897004	781717_44404225_MAPT_1096	1	-	**good**	120
chr17	45897004	45897124	781717_44404225_MAPT_1095	1	-	**good**	120
chr17	45897124	45897244	781717_44404225_MAPT_1094	1	-	**good**	120
chr17	45897244	45897364	781717_44404225_MAPT_1093	1	-	**good**	120
chr17	45897364	45897484	781717_44404225_MAPT_1092	1	-	**good**	120
chr17	45897484	45897604	781717_44404225_MAPT_1091	1	-	**good**	120
chr17	45897604	45897724	781717_44404225_MAPT_1090	1	-	**good**	120
chr17	45897724	45897844	781717_44404225_MAPT_1089	1	-	**good**	120
chr17	45897844	45897964	781717_44404225_MAPT_1088	1	-	**good**	120
chr17	45897964	45898084	781717_44404225_MAPT_1087	3	-	**good**	120
chr17	45898084	45898204	781717_44404225_MAPT_1086	1	-	**good**	120
chr17	45898204	45898324	781717_44404225_MAPT_1085	1	-	**good**	120
chr17	45898324	45898444	781717_44404225_MAPT_1084	1	-	**good**	120
chr17	45898444	45898564	781717_44404225_MAPT_1083	1	-	**good**	120
chr17	45898564	45898684	781717_44404225_MAPT_1082	1	-	**good**	120
chr17	45898684	45898804	781717_44404225_MAPT_1081	1	-	**good**	120
chr17	45898804	45898924	781717_44404225_MAPT_1080	1	-	**good**	120
chr17	45898924	45899044	781717_44404225_MAPT_1079	1	-	**good**	120
chr17	45899044	45899164	781717_44404225_MAPT_1078	1	-	**good**	120
chr17	45899284	45899404	781717_44404225_MAPT_1076	1	-	**good**	120
chr17	45899404	45899524	781717_44404225_MAPT_1075	21	-	**good**	120
chr17	45899524	45899644	781717_44404225_MAPT_1074	1	-	**good**	120
chr17	45899644	45899764	781717_44404225_MAPT_1073	1	-	**good**	120
chr17	45899764	45899884	781717_44404225_MAPT_1072	1	-	**good**	120
chr17	45899884	45900004	781717_44404225_MAPT_1071	1	-	**good**	120
chr17	45900004	45900124	781717_44404225_MAPT_1070	1	-	**good**	120
chr17	45900124	45900244	781717_44404225_MAPT_1069	1	-	**good**	120
chr17	45900244	45900364	781717_44404225_MAPT_1068	1	-	**good**	120
chr17	45900364	45900484	781717_44404225_MAPT_1067	1	-	**good**	120
chr17	45900484	45900604	781717_44404225_MAPT_1066	1	-	**good**	120
chr17	45900604	45900724	781717_44404225_MAPT_1065	1	-	**good**	120
chr17	45900724	45900844	781717_44404225_MAPT_1064	1	-	**good**	120
chr17	45900844	45900964	781717_44404225_MAPT_1063	1	-	**good**	120
chr17	45900964	45901084	781717_44404225_MAPT_1062	1	-	**good**	120
chr17	45901084	45901204	781717_44404225_MAPT_1061	1	-	**good**	120
chr17	45901204	45901324	781717_44404225_MAPT_1060	1	-	**good**	120
chr17	45901324	45901444	781717_44404225_MAPT_1059	1	-	**good**	120
chr17	45901444	45901564	781717_44404225_MAPT_1058	1	-	**good**	120
chr17	45901564	45901684	781717_44404225_MAPT_1057	1	-	**good**	120

chr17	45901684	45901804	781717_44404225_MAPT_1056	1	-	**good**	120
chr17	45901804	45901924	781717_44404225_MAPT_1055	1	-	**good**	120
chr17	45901924	45902044	781717_44404225_MAPT_1054	1	-	**good**	120
chr17	45902404	45902524	781717_44404225_MAPT_1050	1	-	**good**	120
chr17	45902524	45902644	781717_44404225_MAPT_1049	1	-	**good**	120
chr17	45902644	45902764	781717_44404225_MAPT_1048	1	-	**good**	120
chr17	45902764	45902884	781717_44404225_MAPT_1047	1	-	**good**	120
chr17	45902884	45903004	781717_44404225_MAPT_1046	1	-	**good**	120
chr17	45903004	45903124	781717_44404225_MAPT_1045	1	-	**good**	120
chr17	45903124	45903244	781717_44404225_MAPT_1044	1	-	**good**	120
chr17	45903244	45903364	781717_44404225_MAPT_1043	1	-	**good**	120
chr17	45904684	45904804	781717_44404225_MAPT_1031	1	-	**good**	120
chr17	45904804	45904924	781717_44404225_MAPT_1030	1	-	**good**	120
chr17	45904924	45905044	781717_44404225_MAPT_1029	1	-	**good**	120
chr17	45905044	45905164	781717_44404225_MAPT_1028	1	-	**good**	120
chr17	45905164	45905284	781717_44404225_MAPT_1027	1	-	**good**	120
chr17	45905284	45905404	781717_44404225_MAPT_1026	1	-	**good**	120
chr17	45905404	45905524	781717_44404225_MAPT_1025	1	-	**good**	120
chr17	45905524	45905644	781717_44404225_MAPT_1024	1	-	**good**	120
chr17	45905644	45905764	781717_44404225_MAPT_1023	1	-	**good**	120
chr17	45905764	45905884	781717_44404225_MAPT_1022	1	-	**good**	120
chr17	45905884	45906004	781717_44404225_MAPT_1021	1	-	**good**	120
chr17	45906004	45906124	781717_44404225_MAPT_1020	1	-	**good**	120
chr17	45906124	45906244	781717_44404225_MAPT_1019	1	-	**good**	120
chr17	45906244	45906364	781717_44404225_MAPT_1018	1	-	**good**	120
chr17	45906364	45906484	781717_44404225_MAPT_1017	1	-	**good**	120
chr17	45906484	45906604	781717_44404225_MAPT_1016	1	-	**good**	120
chr17	45906604	45906724	781717_44404225_MAPT_1015	1	-	**good**	120
chr17	45906724	45906844	781717_44404225_MAPT_1014	1	-	**good**	120
chr17	45906844	45906964	781717_44404225_MAPT_1013	1	-	**good**	120
chr17	45906964	45907084	781717_44404225_MAPT_1012	1	-	**good**	120
chr17	45907084	45907204	781717_44404225_MAPT_1011	1	-	**good**	120
chr17	45907204	45907324	781717_44404225_MAPT_1010	1	-	**good**	120
chr17	45907324	45907444	781717_44404225_MAPT_1009	1	-	**good**	120
chr17	45907444	45907564	781717_44404225_MAPT_1008	1	-	**good**	120
chr17	45907564	45907684	781717_44404225_MAPT_1007	1	-	**good**	120
chr17	45907684	45907804	781717_44404225_MAPT_1006	1	-	**good**	120
chr17	45907804	45907924	781717_44404225_MAPT_1005	1	-	**good**	120
chr17	45907924	45908044	781717_44404225_MAPT_1004	1	-	**good**	120
chr17	45908044	45908164	781717_44404225_MAPT_1003	1	-	**good**	120
chr17	45908164	45908284	781717_44404225_MAPT_1002	1	-	**good**	120
chr17	45908284	45908404	781717_44404225_MAPT_1001	1	-	**good**	120
chr17	45908404	45908524	781717_44404225_MAPT_1000	1	-	**good**	120
chr17	45908524	45908644	781717_44404225_MAPT_999	1	-	**good**	120
chr17	45908644	45908764	781717_44404225_MAPT_998	1	-	**good**	120
chr17	45908764	45908884	781717_44404225_MAPT_997	1	-	**good**	120

chr17	45908884	45909004	781717_44404225_MAPT_996	1	-	**good**	120
chr17	45909004	45909124	781717_44404225_MAPT_995	1	-	**good**	120
chr17	45909124	45909244	781717_44404225_MAPT_994	2	-	**good**	120
chr17	45909244	45909364	781717_44404225_MAPT_993	1	-	**good**	120
chr17	45909364	45909484	781717_44404225_MAPT_992	1	-	**good**	120
chr17	45910084	45910204	781717_44404225_MAPT_986	1	-	**good**	120
chr17	45910204	45910324	781717_44404225_MAPT_985	1	-	**good**	120
chr17	45910324	45910444	781717_44404225_MAPT_984	1	-	**good**	120
chr17	45910444	45910564	781717_44404225_MAPT_983	1	-	**good**	120
chr17	45910564	45910684	781717_44404225_MAPT_982	1	-	**good**	120
chr17	45910684	45910804	781717_44404225_MAPT_981	1	-	**good**	120
chr17	45910804	45910924	781717_44404225_MAPT_980	1	-	**good**	120
chr17	45910924	45911044	781717_44404225_MAPT_979	1	-	**good**	120
chr17	45911044	45911164	781717_44404225_MAPT_978	1	-	**good**	120
chr17	45911164	45911284	781717_44404225_MAPT_977	1	-	**good**	120
chr17	45911284	45911404	781717_44404225_MAPT_976	1	-	**good**	120
chr17	45911764	45911884	781717_44404225_MAPT_972	1	-	**good**	120
chr17	45911884	45912004	781717_44404225_MAPT_971	1	-	**good**	120
chr17	45912004	45912124	781717_44404225_MAPT_970	1	-	**good**	120
chr17	45912124	45912244	781717_44404225_MAPT_969	1	-	**good**	120
chr17	45912364	45912484	781717_44404225_MAPT_967	36	-	**good**	120
chr17	45912724	45912844	781717_44404225_MAPT_964	15	-	**good**	120
chr17	45912844	45912964	781717_44404225_MAPT_963	1	-	**good**	120
chr17	45912964	45913084	781717_44404225_MAPT_962	1	-	**good**	120
chr17	45913444	45913564	781717_44404225_MAPT_958	1	-	**good**	120
chr17	45913564	45913684	781717_44404225_MAPT_957	1	-	**good**	120
chr17	45913684	45913804	781717_44404225_MAPT_956	1	-	**good**	120
chr17	45913804	45913924	781717_44404225_MAPT_955	1	-	**good**	120
chr17	45913924	45914044	781717_44404225_MAPT_954	1	-	**good**	120
chr17	45914044	45914164	781717_44404225_MAPT_953	1	-	**good**	120
chr17	45914164	45914284	781717_44404225_MAPT_952	1	-	**good**	120
chr17	45914284	45914404	781717_44404225_MAPT_951	1	-	**good**	120
chr17	45914404	45914524	781717_44404225_MAPT_950	1	-	**good**	120
chr17	45914524	45914644	781717_44404225_MAPT_949	1	-	**good**	120
chr17	45915004	45915124	781717_44404225_MAPT_945	1	-	**good**	120
chr17	45915124	45915244	781717_44404225_MAPT_944	1	-	**good**	120
chr17	45915604	45915724	781717_44404225_MAPT_940	1	-	**good**	120
chr17	45915724	45915844	781717_44404225_MAPT_939	1	-	**good**	120
chr17	45915844	45915964	781717_44404225_MAPT_938	1	-	**good**	120
chr17	45915964	45916084	781717_44404225_MAPT_937	1	-	**good**	120
chr17	45916084	45916204	781717_44404225_MAPT_936	1	-	**good**	120
chr17	45916204	45916324	781717_44404225_MAPT_935	1	-	**good**	120
chr17	45916324	45916444	781717_44404225_MAPT_934	1	-	**good**	120
chr17	45916444	45916564	781717_44404225_MAPT_933	1	-	**good**	120
chr17	45916564	45916684	781717_44404225_MAPT_932	1	-	**good**	120
chr17	45916684	45916804	781717_44404225_MAPT_931	1	-	**good**	120

chr17	45916804	45916924	781717_44404225_MAPT_930	1	-	**good**	120
chr17	45916924	45917044	781717_44404225_MAPT_929	1	-	**good**	120
chr17	45917044	45917164	781717_44404225_MAPT_928	1	-	**good**	120
chr17	45917164	45917284	781717_44404225_MAPT_927	1	-	**good**	120
chr17	45917284	45917404	781717_44404225_MAPT_926	1	-	**good**	120
chr17	45917404	45917524	781717_44404225_MAPT_925	1	-	**good**	120
chr17	45917524	45917644	781717_44404225_MAPT_924	1	-	**good**	120
chr17	45917644	45917764	781717_44404225_MAPT_923	1	-	**good**	120
chr17	45918124	45918244	781717_44404225_MAPT_919	1	-	**good**	120
chr17	45918244	45918364	781717_44404225_MAPT_918	1	-	**good**	120
chr17	45918364	45918484	781717_44404225_MAPT_917	1	-	**good**	120
chr17	45918484	45918604	781717_44404225_MAPT_916	1	-	**good**	120
chr17	45918604	45918724	781717_44404225_MAPT_915	1	-	**good**	120
chr17	45919084	45919204	781717_44404225_MAPT_911	1	-	**good**	120
chr17	45919204	45919324	781717_44404225_MAPT_910	1	-	**good**	120
chr17	45919324	45919444	781717_44404225_MAPT_909	1	-	**good**	120
chr17	45919444	45919564	781717_44404225_MAPT_908	1	-	**good**	120
chr17	45919924	45920044	781717_44404225_MAPT_904	1	-	**good**	120
chr17	45920044	45920164	781717_44404225_MAPT_903	1	-	**good**	120
chr17	45920164	45920284	781717_44404225_MAPT_902	1	-	**good**	120
chr17	45920284	45920404	781717_44404225_MAPT_901	1	-	**good**	120
chr17	45920404	45920524	781717_44404225_MAPT_900	1	-	**good**	120
chr17	45920524	45920644	781717_44404225_MAPT_899	1	-	**good**	120
chr17	45920644	45920764	781717_44404225_MAPT_898	1	-	**good**	120
chr17	45920764	45920884	781717_44404225_MAPT_897	1	-	**good**	120
chr17	45920884	45921004	781717_44404225_MAPT_896	1	-	**good**	120
chr17	45921004	45921124	781717_44404225_MAPT_895	1	-	**good**	120
chr17	45921124	45921244	781717_44404225_MAPT_894	1	-	**good**	120
chr17	45921244	45921364	781717_44404225_MAPT_893	1	-	**good**	120
chr17	45921364	45921484	781717_44404225_MAPT_892	1	-	**good**	120
chr17	45921484	45921604	781717_44404225_MAPT_891	1	-	**good**	120
chr17	45921604	45921724	781717_44404225_MAPT_890	1	-	**good**	120
chr17	45921724	45921844	781717_44404225_MAPT_889	1	-	**good**	120
chr17	45921844	45921964	781717_44404225_MAPT_888	1	-	**good**	120
chr17	45921964	45922084	781717_44404225_MAPT_887	1	-	**good**	120
chr17	45922204	45922324	781717_44404225_MAPT_885	18	-	**good**	120
chr17	45922324	45922444	781717_44404225_MAPT_884	1	-	**good**	120
chr17	45922444	45922564	781717_44404225_MAPT_883	1	-	**good**	120
chr17	45922564	45922684	781717_44404225_MAPT_882	1	-	**good**	120
chr17	45922684	45922804	781717_44404225_MAPT_881	1	-	**good**	120
chr17	45922804	45922924	781717_44404225_MAPT_880	1	-	**good**	120
chr17	45922924	45923044	781717_44404225_MAPT_879	1	-	**good**	120
chr17	45923044	45923164	781717_44404225_MAPT_878	1	-	**good**	120
chr17	45923164	45923284	781717_44404225_MAPT_877	1	-	**good**	120
chr17	45923284	45923404	781717_44404225_MAPT_876	1	-	**good**	120
chr17	45923404	45923524	781717_44404225_MAPT_875	1	-	**good**	120

chr17	45923524	45923644	781717_44404225_MAPT_874	1	-	**good**	120
chr17	45923644	45923764	781717_44404225_MAPT_873	1	-	**good**	120
chr17	45923764	45923884	781717_44404225_MAPT_872	1	-	**good**	120
chr17	45923884	45924004	781717_44404225_MAPT_871	8	-	**good**	120
chr17	45924004	45924124	781717_44404225_MAPT_870	1	-	**good**	120
chr17	45924124	45924244	781717_44404225_MAPT_869	1	-	**good**	120
chr17	45924244	45924364	781717_44404225_MAPT_868	1	-	**good**	120
chr17	45924364	45924484	781717_44404225_MAPT_867	1	-	**good**	120
chr17	45924484	45924604	781717_44404225_MAPT_866	1	-	**good**	120
chr17	45924604	45924724	781717_44404225_MAPT_865	1	-	**good**	120
chr17	45924724	45924844	781717_44404225_MAPT_864	1	-	**good**	120
chr17	45924844	45924964	781717_44404225_MAPT_863	1	-	**good**	120
chr17	45924964	45925084	781717_44404225_MAPT_862	1	-	**good**	120
chr17	45925084	45925204	781717_44404225_MAPT_861	1	-	**good**	120
chr17	45925204	45925324	781717_44404225_MAPT_860	1	-	**good**	120
chr17	45925324	45925444	781717_44404225_MAPT_859	1	-	**good**	120
chr17	45925444	45925564	781717_44404225_MAPT_858	3	-	**good**	120
chr17	45925564	45925684	781717_44404225_MAPT_857	1	-	**good**	120
chr17	45925684	45925804	781717_44404225_MAPT_856	1	-	**good**	120
chr17	45925804	45925924	781717_44404225_MAPT_855	1	-	**good**	120
chr17	45926644	45926764	781717_44404225_MAPT_848	1	-	**good**	120
chr17	45926764	45926884	781717_44404225_MAPT_847	1	-	**good**	120
chr17	45927004	45927124	781717_44404225_MAPT_845	1	-	**good**	120
chr17	45927124	45927244	781717_44404225_MAPT_844	1	-	**good**	120
chr17	45927244	45927364	781717_44404225_MAPT_843	1	-	**good**	120
chr17	45927364	45927484	781717_44404225_MAPT_842	1	-	**good**	120
chr17	45927484	45927604	781717_44404225_MAPT_841	1	-	**good**	120
chr17	45927604	45927724	781717_44404225_MAPT_840	1	-	**good**	120
chr17	45928084	45928204	781717_44404225_MAPT_836	1	-	**good**	120
chr17	45928204	45928324	781717_44404225_MAPT_835	1	-	**good**	120
chr17	45928324	45928444	781717_44404225_MAPT_834	1	-	**good**	120
chr17	45928444	45928564	781717_44404225_MAPT_833	1	-	**good**	120
chr17	45928564	45928684	781717_44404225_MAPT_832	2	-	**good**	120
chr17	45929044	45929164	781717_44404225_MAPT_828	1	-	**good**	120
chr17	45929164	45929284	781717_44404225_MAPT_827	1	-	**good**	120
chr17	45929284	45929404	781717_44404225_MAPT_826	1	-	**good**	120
chr17	45929404	45929524	781717_44404225_MAPT_825	1	-	**good**	120
chr17	45929524	45929644	781717_44404225_MAPT_824	1	-	**good**	120
chr17	45929644	45929764	781717_44404225_MAPT_823	1	-	**good**	120
chr17	45929764	45929884	781717_44404225_MAPT_822	1	-	**good**	120
chr17	45929884	45930004	781717_44404225_MAPT_821	1	-	**good**	120
chr17	45930004	45930124	781717_44404225_MAPT_820	1	-	**good**	120
chr17	45930484	45930604	781717_44404225_MAPT_816	2	-	**good**	120
chr17	45930604	45930724	781717_44404225_MAPT_815	1	-	**good**	120
chr17	45930724	45930844	781717_44404225_MAPT_814	1	-	**good**	120
chr17	45930844	45930964	781717_44404225_MAPT_813	1	-	**good**	120

chr17	45930964	45931084	781717_44404225_MAPT_812	1	-	**good**	120
chr17	45931084	45931204	781717_44404225_MAPT_811	1	-	**good**	120
chr17	45931204	45931324	781717_44404225_MAPT_810	2	-	**good**	120
chr17	45931324	45931444	781717_44404225_MAPT_809	1	-	**good**	120
chr17	45931444	45931564	781717_44404225_MAPT_808	1	-	**good**	120
chr17	45931564	45931684	781717_44404225_MAPT_807	1	-	**good**	120
chr17	45931684	45931804	781717_44404225_MAPT_806	3	-	**good**	120
chr17	45932164	45932284	781717_44404225_MAPT_802	1	-	**good**	120
chr17	45932644	45932764	781717_44404225_MAPT_798	1	-	**good**	120
chr17	45933124	45933244	781717_44404225_MAPT_794	14	-	**good**	120
chr17	45933604	45933724	781717_44404225_MAPT_790	1	-	**good**	120
chr17	45933724	45933844	781717_44404225_MAPT_789	1	-	**good**	120
chr17	45933844	45933964	781717_44404225_MAPT_788	1	-	**good**	120
chr17	45933964	45934084	781717_44404225_MAPT_787	1	-	**good**	120
chr17	45934324	45934444	781717_44404225_MAPT_784	1	-	**good**	120
chr17	45934444	45934564	781717_44404225_MAPT_783	1	-	**good**	120
chr17	45934564	45934684	781717_44404225_MAPT_782	1	-	**good**	120
chr17	45934684	45934804	781717_44404225_MAPT_781	1	-	**good**	120
chr17	45934804	45934924	781717_44404225_MAPT_780	1	-	**good**	120
chr17	45934924	45935044	781717_44404225_MAPT_779	1	-	**good**	120
chr17	45935044	45935164	781717_44404225_MAPT_778	1	-	**good**	120
chr17	45935164	45935284	781717_44404225_MAPT_777	1	-	**good**	120
chr17	45935284	45935404	781717_44404225_MAPT_776	1	-	**good**	120
chr17	45935404	45935524	781717_44404225_MAPT_775	1	-	**good**	120
chr17	45935524	45935644	781717_44404225_MAPT_774	1	-	**good**	120
chr17	45935644	45935764	781717_44404225_MAPT_773	1	-	**good**	120
chr17	45935764	45935884	781717_44404225_MAPT_772	1	-	**good**	120
chr17	45935884	45936004	781717_44404225_MAPT_771	1	-	**good**	120
chr17	45936004	45936124	781717_44404225_MAPT_770	1	-	**good**	120
chr17	45936124	45936244	781717_44404225_MAPT_769	1	-	**good**	120
chr17	45936244	45936364	781717_44404225_MAPT_768	1	-	**good**	120
chr17	45936364	45936484	781717_44404225_MAPT_767	1	-	**good**	120
chr17	45936484	45936604	781717_44404225_MAPT_766	1	-	**good**	120
chr17	45936604	45936724	781717_44404225_MAPT_765	1	-	**good**	120
chr17	45936724	45936844	781717_44404225_MAPT_764	1	-	**good**	120
chr17	45936844	45936964	781717_44404225_MAPT_763	1	-	**good**	120
chr17	45936964	45937084	781717_44404225_MAPT_762	1	-	**good**	120
chr17	45937084	45937204	781717_44404225_MAPT_761	1	-	**good**	120
chr17	45937204	45937324	781717_44404225_MAPT_760	1	-	**good**	120
chr17	45937684	45937804	781717_44404225_MAPT_756	1	-	**good**	120
chr17	45937804	45937924	781717_44404225_MAPT_755	1	-	**good**	120
chr17	45937924	45938044	781717_44404225_MAPT_754	1	-	**good**	120
chr17	45938044	45938164	781717_44404225_MAPT_753	23	-	**good**	120
chr17	45938284	45938404	781717_44404225_MAPT_751	1	-	**good**	120
chr17	45938404	45938524	781717_44404225_MAPT_750	11	-	**good**	120
chr17	45938524	45938644	781717_44404225_MAPT_749	5	-	**good**	120

chr17	45939244	45939364	781717_44404225_MAPT_743	1	-	**good**	120
chr17	45939364	45939484	781717_44404225_MAPT_742	1	-	**good**	120
chr17	45939484	45939604	781717_44404225_MAPT_741	1	-	**good**	120
chr17	45939604	45939724	781717_44404225_MAPT_740	1	-	**good**	120
chr17	45939724	45939844	781717_44404225_MAPT_739	1	-	**good**	120
chr17	45939844	45939964	781717_44404225_MAPT_738	1	-	**good**	120
chr17	45939964	45940084	781717_44404225_MAPT_737	1	-	**good**	120
chr17	45940084	45940204	781717_44404225_MAPT_736	1	-	**good**	120
chr17	45940204	45940324	781717_44404225_MAPT_735	1	-	**good**	120
chr17	45940324	45940444	781717_44404225_MAPT_734	10	-	**good**	120
chr17	45940444	45940564	781717_44404225_MAPT_733	1	-	**good**	120
chr17	45940564	45940684	781717_44404225_MAPT_732	1	-	**good**	120
chr17	45940684	45940804	781717_44404225_MAPT_731	1	-	**good**	120
chr17	45940804	45940924	781717_44404225_MAPT_730	1	-	**good**	120
chr17	45940924	45941044	781717_44404225_MAPT_729	1	-	**good**	120
chr17	45941044	45941164	781717_44404225_MAPT_728	1	-	**good**	120
chr17	45941164	45941284	781717_44404225_MAPT_727	1	-	**good**	120
chr17	45941284	45941404	781717_44404225_MAPT_726	1	-	**good**	120
chr17	45941404	45941524	781717_44404225_MAPT_725	1	-	**good**	120
chr17	45941764	45941884	781717_44404225_MAPT_722	1	-	**good**	120
chr17	45941884	45942004	781717_44404225_MAPT_721	1	-	**good**	120
chr17	45942004	45942124	781717_44404225_MAPT_720	1	-	**good**	120
chr17	45942124	45942244	781717_44404225_MAPT_719	1	-	**good**	120
chr17	45942244	45942364	781717_44404225_MAPT_718	1	-	**good**	120
chr17	45942364	45942484	781717_44404225_MAPT_717	1	-	**good**	120
chr17	45942484	45942604	781717_44404225_MAPT_716	1	-	**good**	120
chr17	45942604	45942724	781717_44404225_MAPT_715	1	-	**good**	120
chr17	45942724	45942844	781717_44404225_MAPT_714	1	-	**good**	120
chr17	45942844	45942964	781717_44404225_MAPT_713	1	-	**good**	120
chr17	45942964	45943084	781717_44404225_MAPT_712	1	-	**good**	120
chr17	45943324	45943444	781717_44404225_MAPT_709	1	-	**good**	120
chr17	45943444	45943564	781717_44404225_MAPT_708	1	-	**good**	120
chr17	45943564	45943684	781717_44404225_MAPT_707	1	-	**good**	120
chr17	45943684	45943804	781717_44404225_MAPT_706	1	-	**good**	120
chr17	45943804	45943924	781717_44404225_MAPT_705	1	-	**good**	120
chr17	45943924	45944044	781717_44404225_MAPT_704	1	-	**good**	120
chr17	45944044	45944164	781717_44404225_MAPT_703	1	-	**good**	120
chr17	45944164	45944284	781717_44404225_MAPT_702	1	-	**good**	120
chr17	45944284	45944404	781717_44404225_MAPT_701	1	-	**good**	120
chr17	45944404	45944524	781717_44404225_MAPT_700	1	-	**good**	120
chr17	45944524	45944644	781717_44404225_MAPT_699	1	-	**good**	120
chr17	45944644	45944764	781717_44404225_MAPT_698	1	-	**good**	120
chr17	45944764	45944884	781717_44404225_MAPT_697	1	-	**good**	120
chr17	45944884	45945004	781717_44404225_MAPT_696	1	-	**good**	120
chr17	45945004	45945124	781717_44404225_MAPT_695	1	-	**good**	120
chr17	45945124	45945244	781717_44404225_MAPT_694	1	-	**good**	120

chr17	45945244	45945364	781717_44404225_MAPT_693	1	-	**good**	120
chr17	45945364	45945484	781717_44404225_MAPT_692	1	-	**good**	120
chr17	45945844	45945964	781717_44404225_MAPT_688	1	-	**good**	120
chr17	45945964	45946084	781717_44404225_MAPT_687	1	-	**good**	120
chr17	45946084	45946204	781717_44404225_MAPT_686	1	-	**good**	120
chr17	45946204	45946324	781717_44404225_MAPT_685	1	-	**good**	120
chr17	45946684	45946804	781717_44404225_MAPT_681	1	-	**good**	120
chr17	45946804	45946924	781717_44404225_MAPT_680	1	-	**good**	120
chr17	45946924	45947044	781717_44404225_MAPT_679	1	-	**good**	120
chr17	45947044	45947164	781717_44404225_MAPT_678	1	-	**good**	120
chr17	45947164	45947284	781717_44404225_MAPT_677	1	-	**good**	120
chr17	45947284	45947404	781717_44404225_MAPT_676	1	-	**good**	120
chr17	45947764	45947884	781717_44404225_MAPT_672	2	-	**good**	120
chr17	45947884	45948004	781717_44404225_MAPT_671	2	-	**good**	120
chr17	45948364	45948484	781717_44404225_MAPT_667	1	-	**good**	120
chr17	45948484	45948604	781717_44404225_MAPT_666	1	-	**good**	120
chr17	45948604	45948724	781717_44404225_MAPT_665	1	-	**good**	120
chr17	45948724	45948844	781717_44404225_MAPT_664	1	-	**good**	120
chr17	45948844	45948964	781717_44404225_MAPT_663	1	-	**good**	120
chr17	45948964	45949084	781717_44404225_MAPT_662	4	-	**good**	120
chr17	45949084	45949204	781717_44404225_MAPT_661	1	-	**good**	120
chr17	45949204	45949324	781717_44404225_MAPT_660	1	-	**good**	120
chr17	45949324	45949444	781717_44404225_MAPT_659	1	-	**good**	120
chr17	45949444	45949564	781717_44404225_MAPT_658	1	-	**good**	120
chr17	45949564	45949684	781717_44404225_MAPT_657	1	-	**good**	120
chr17	45949684	45949804	781717_44404225_MAPT_656	1	-	**good**	120
chr17	45949804	45949924	781717_44404225_MAPT_655	1	-	**good**	120
chr17	45949924	45950044	781717_44404225_MAPT_654	1	-	**good**	120
chr17	45950044	45950164	781717_44404225_MAPT_653	1	-	**good**	120
chr17	45950164	45950284	781717_44404225_MAPT_652	1	-	**good**	120
chr17	45950284	45950404	781717_44404225_MAPT_651	1	-	**good**	120
chr17	45950404	45950524	781717_44404225_MAPT_650	1	-	**good**	120
chr17	45950524	45950644	781717_44404225_MAPT_649	1	-	**good**	120
chr17	45951004	45951124	781717_44404225_MAPT_645	1	-	**good**	120
chr17	45951124	45951244	781717_44404225_MAPT_644	8	-	**good**	120
chr17	45951364	45951484	781717_44404225_MAPT_642	1	-	**good**	120
chr17	45951484	45951604	781717_44404225_MAPT_641	1	-	**good**	120
chr17	45951604	45951724	781717_44404225_MAPT_640	1	-	**good**	120
chr17	45951724	45951844	781717_44404225_MAPT_639	1	-	**good**	120
chr17	45951964	45952084	781717_44404225_MAPT_637	1	-	**good**	120
chr17	45952084	45952204	781717_44404225_MAPT_636	1	-	**good**	120
chr17	45952204	45952324	781717_44404225_MAPT_635	1	-	**good**	120
chr17	45952324	45952444	781717_44404225_MAPT_634	1	-	**good**	120
chr17	45952684	45952804	781717_44404225_MAPT_631	1	-	**good**	120
chr17	45952804	45952924	781717_44404225_MAPT_630	1	-	**good**	120
chr17	45952924	45953044	781717_44404225_MAPT_629	1	-	**good**	120

chr17	45953044	45953164	781717_44404225_MAPT_628	1	-	**good**	120
chr17	45953164	45953284	781717_44404225_MAPT_627	1	-	**good**	120
chr17	45953284	45953404	781717_44404225_MAPT_626	1	-	**good**	120
chr17	45953404	45953524	781717_44404225_MAPT_625	1	-	**good**	120
chr17	45953524	45953644	781717_44404225_MAPT_624	1	-	**good**	120
chr17	45953644	45953764	781717_44404225_MAPT_623	1	-	**good**	120
chr17	45953764	45953884	781717_44404225_MAPT_622	1	-	**good**	120
chr17	45953884	45954004	781717_44404225_MAPT_621	1	-	**good**	120
chr17	45954004	45954124	781717_44404225_MAPT_620	1	-	**good**	120
chr17	45954124	45954244	781717_44404225_MAPT_619	15	-	**good**	120
chr17	45955084	45955204	781717_44404225_MAPT_611	1	-	**good**	120
chr17	45955204	45955324	781717_44404225_MAPT_610	1	-	**good**	120
chr17	45955324	45955444	781717_44404225_MAPT_609	1	-	**good**	120
chr17	45955444	45955564	781717_44404225_MAPT_608	1	-	**good**	120
chr17	45955564	45955684	781717_44404225_MAPT_607	1	-	**good**	120
chr17	45956044	45956164	781717_44404225_MAPT_603	1	-	**good**	120
chr17	45956164	45956284	781717_44404225_MAPT_602	1	-	**good**	120
chr17	45956284	45956404	781717_44404225_MAPT_601	1	-	**good**	120
chr17	45956404	45956524	781717_44404225_MAPT_600	1	-	**good**	120
chr17	45957364	45957484	781717_44404225_MAPT_592	1	-	**good**	120
chr17	45957484	45957604	781717_44404225_MAPT_591	1	-	**good**	120
chr17	45957604	45957724	781717_44404225_MAPT_590	1	-	**good**	120
chr17	45957724	45957844	781717_44404225_MAPT_589	1	-	**good**	120
chr17	45957844	45957964	781717_44404225_MAPT_588	1	-	**good**	120
chr17	45957964	45958084	781717_44404225_MAPT_587	1	-	**good**	120
chr17	45958084	45958204	781717_44404225_MAPT_586	1	-	**good**	120
chr17	45958204	45958324	781717_44404225_MAPT_585	1	-	**good**	120
chr17	45958324	45958444	781717_44404225_MAPT_584	1	-	**good**	120
chr17	45959044	45959164	781717_44404225_MAPT_578	1	-	**good**	120
chr17	45959284	45959404	781717_44404225_MAPT_576	1	-	**good**	120
chr17	45959404	45959524	781717_44404225_MAPT_575	1	-	**good**	120
chr17	45959764	45959884	781717_44404225_MAPT_572	1	-	**good**	120
chr17	45959884	45960004	781717_44404225_MAPT_571	1	-	**good**	120
chr17	45960004	45960124	781717_44404225_MAPT_570	1	-	**good**	120
chr17	45960124	45960244	781717_44404225_MAPT_569	1	-	**good**	120
chr17	45960244	45960364	781717_44404225_MAPT_568	1	-	**good**	120
chr17	45960364	45960484	781717_44404225_MAPT_567	1	-	**good**	120
chr17	45960484	45960604	781717_44404225_MAPT_566	1	-	**good**	120
chr17	45960844	45960964	781717_44404225_MAPT_563	1	-	**good**	120
chr17	45960964	45961084	781717_44404225_MAPT_562	1	-	**good**	120
chr17	45961204	45961324	781717_44404225_MAPT_560	1	-	**good**	120
chr17	45961324	45961444	781717_44404225_MAPT_559	1	-	**good**	120
chr17	45961444	45961564	781717_44404225_MAPT_558	1	-	**good**	120
chr17	45961564	45961684	781717_44404225_MAPT_557	1	-	**good**	120
chr17	45961684	45961804	781717_44404225_MAPT_556	1	-	**good**	120
chr17	45962044	45962164	781717_44404225_MAPT_553	1	-	**good**	120

chr17	45962164	45962284	781717_44404225_MAPT_552	1	-	**good**	120
chr17	45962284	45962404	781717_44404225_MAPT_551	1	-	**good**	120
chr17	45962404	45962524	781717_44404225_MAPT_550	1	-	**good**	120
chr17	45962524	45962644	781717_44404225_MAPT_549	1	-	**good**	120
chr17	45963004	45963124	781717_44404225_MAPT_545	1	-	**good**	120
chr17	45963124	45963244	781717_44404225_MAPT_544	1	-	**good**	120
chr17	45963244	45963364	781717_44404225_MAPT_543	1	-	**good**	120
chr17	45963364	45963484	781717_44404225_MAPT_542	1	-	**good**	120
chr17	45963484	45963604	781717_44404225_MAPT_541	1	-	**good**	120
chr17	45963604	45963724	781717_44404225_MAPT_540	1	-	**good**	120
chr17	45963724	45963844	781717_44404225_MAPT_539	1	-	**good**	120
chr17	45963844	45963964	781717_44404225_MAPT_538	1	-	**good**	120
chr17	45963964	45964084	781717_44404225_MAPT_537	1	-	**good**	120
chr17	45964084	45964204	781717_44404225_MAPT_536	1	-	**good**	120
chr17	45964684	45964804	781717_44404225_MAPT_531	1	-	**good**	120
chr17	45964804	45964924	781717_44404225_MAPT_530	1	-	**good**	120
chr17	45964924	45965044	781717_44404225_MAPT_529	1	-	**good**	120
chr17	45965044	45965164	781717_44404225_MAPT_528	1	-	**good**	120
chr17	45965764	45965884	781717_44404225_MAPT_522	1	-	**good**	120
chr17	45965884	45966004	781717_44404225_MAPT_521	1	-	**good**	120
chr17	45966004	45966124	781717_44404225_MAPT_520	1	-	**good**	120
chr17	45966124	45966244	781717_44404225_MAPT_519	1	-	**good**	120
chr17	45966604	45966724	781717_44404225_MAPT_515	1	-	**good**	120
chr17	45966724	45966844	781717_44404225_MAPT_514	2	-	**good**	120
chr17	45966844	45966964	781717_44404225_MAPT_513	1	-	**good**	120
chr17	45966964	45967084	781717_44404225_MAPT_512	1	-	**good**	120
chr17	45967084	45967204	781717_44404225_MAPT_511	1	-	**good**	120
chr17	45967204	45967324	781717_44404225_MAPT_510	1	-	**good**	120
chr17	45967324	45967444	781717_44404225_MAPT_509	1	-	**good**	120
chr17	45967444	45967564	781717_44404225_MAPT_508	1	-	**good**	120
chr17	45967564	45967684	781717_44404225_MAPT_507	1	-	**good**	120
chr17	45967684	45967804	781717_44404225_MAPT_506	1	-	**good**	120
chr17	45967804	45967924	781717_44404225_MAPT_505	1	-	**good**	120
chr17	45967924	45968044	781717_44404225_MAPT_504	1	-	**good**	120
chr17	45968044	45968164	781717_44404225_MAPT_503	1	-	**good**	120
chr17	45968164	45968284	781717_44404225_MAPT_502	1	-	**good**	120
chr17	45968284	45968404	781717_44404225_MAPT_501	1	-	**good**	120
chr17	45968404	45968524	781717_44404225_MAPT_500	1	-	**good**	120
chr17	45968524	45968644	781717_44404225_MAPT_499	1	-	**good**	120
chr17	45968644	45968764	781717_44404225_MAPT_498	1	-	**good**	120
chr17	45968764	45968884	781717_44404225_MAPT_497	1	-	**good**	120
chr17	45968884	45969004	781717_44404225_MAPT_496	2	-	**good**	120
chr17	45969004	45969124	781717_44404225_MAPT_495	1	-	**good**	120
chr17	45969124	45969244	781717_44404225_MAPT_494	1	-	**good**	120
chr17	45969244	45969364	781717_44404225_MAPT_493	1	-	**good**	120
chr17	45970204	45970324	781717_44404225_MAPT_485	2	-	**good**	120

chr17	45970444	45970564	781717_44404225_MAPT_483	1	-	**good**	120
chr17	45970564	45970684	781717_44404225_MAPT_482	1	-	**good**	120
chr17	45970684	45970804	781717_44404225_MAPT_481	1	-	**good**	120
chr17	45970804	45970924	781717_44404225_MAPT_480	1	-	**good**	120
chr17	45970924	45971044	781717_44404225_MAPT_479	1	-	**good**	120
chr17	45971044	45971164	781717_44404225_MAPT_478	1	-	**good**	120
chr17	45971164	45971284	781717_44404225_MAPT_477	1	-	**good**	120
chr17	45971284	45971404	781717_44404225_MAPT_476	1	-	**good**	120
chr17	45971404	45971524	781717_44404225_MAPT_475	1	-	**good**	120
chr17	45971524	45971644	781717_44404225_MAPT_474	1	-	**good**	120
chr17	45971644	45971764	781717_44404225_MAPT_473	1	-	**good**	120
chr17	45971764	45971884	781717_44404225_MAPT_472	1	-	**good**	120
chr17	45971884	45972004	781717_44404225_MAPT_471	1	-	**good**	120
chr17	45972004	45972124	781717_44404225_MAPT_470	1	-	**good**	120
chr17	45972124	45972244	781717_44404225_MAPT_469	1	-	**good**	120
chr17	45972244	45972364	781717_44404225_MAPT_468	1	-	**good**	120
chr17	45972364	45972484	781717_44404225_MAPT_467	1	-	**good**	120
chr17	45972484	45972604	781717_44404225_MAPT_466	1	-	**good**	120
chr17	45972604	45972724	781717_44404225_MAPT_465	1	-	**good**	120
chr17	45972724	45972844	781717_44404225_MAPT_464	1	-	**good**	120
chr17	45972844	45972964	781717_44404225_MAPT_463	1	-	**good**	120
chr17	45972964	45973084	781717_44404225_MAPT_462	1	-	**good**	120
chr17	45973084	45973204	781717_44404225_MAPT_461	1	-	**good**	120
chr17	45973204	45973324	781717_44404225_MAPT_460	1	-	**good**	120
chr17	45973324	45973444	781717_44404225_MAPT_459	1	-	**good**	120
chr17	45973444	45973564	781717_44404225_MAPT_458	1	-	**good**	120
chr17	45973564	45973684	781717_44404225_MAPT_457	1	-	**good**	120
chr17	45973684	45973804	781717_44404225_MAPT_456	1	-	**good**	120
chr17	45973804	45973924	781717_44404225_MAPT_455	1	-	**good**	120
chr17	45973924	45974044	781717_44404225_MAPT_454	1	-	**good**	120
chr17	45974044	45974164	781717_44404225_MAPT_453	1	-	**good**	120
chr17	45974164	45974284	781717_44404225_MAPT_452	1	-	**good**	120
chr17	45974284	45974404	781717_44404225_MAPT_451	1	-	**good**	120
chr17	45974404	45974524	781717_44404225_MAPT_450	1	-	**good**	120
chr17	45974524	45974644	781717_44404225_MAPT_449	1	-	**good**	120
chr17	45974644	45974764	781717_44404225_MAPT_448	1	-	**good**	120
chr17	45974764	45974884	781717_44404225_MAPT_447	1	-	**good**	120
chr17	45974884	45975004	781717_44404225_MAPT_446	1	-	**good**	120
chr17	45975004	45975124	781717_44404225_MAPT_445	1	-	**good**	120
chr17	45975124	45975244	781717_44404225_MAPT_444	1	-	**good**	120
chr17	45975244	45975364	781717_44404225_MAPT_443	1	-	**good**	120
chr17	45975364	45975484	781717_44404225_MAPT_442	1	-	**good**	120
chr17	45975484	45975604	781717_44404225_MAPT_441	1	-	**good**	120
chr17	45975604	45975724	781717_44404225_MAPT_440	1	-	**good**	120
chr17	45975724	45975844	781717_44404225_MAPT_439	1	-	**good**	120
chr17	45975844	45975964	781717_44404225_MAPT_438	1	-	**good**	120

chr17	45975964	45976084	781717_44404225_MAPT_437	1	-	**good**	120
chr17	45976084	45976204	781717_44404225_MAPT_436	1	-	**good**	120
chr17	45976204	45976324	781717_44404225_MAPT_435	1	-	**good**	120
chr17	45976324	45976444	781717_44404225_MAPT_434	1	-	**good**	120
chr17	45976444	45976564	781717_44404225_MAPT_433	1	-	**good**	120
chr17	45976564	45976684	781717_44404225_MAPT_432	1	-	**good**	120
chr17	45976684	45976804	781717_44404225_MAPT_431	1	-	**good**	120
chr17	45976804	45976924	781717_44404225_MAPT_430	1	-	**good**	120
chr17	45976924	45977044	781717_44404225_MAPT_429	1	-	**good**	120
chr17	45977044	45977164	781717_44404225_MAPT_428	1	-	**good**	120
chr17	45977164	45977284	781717_44404225_MAPT_427	1	-	**good**	120
chr17	45977284	45977404	781717_44404225_MAPT_426	1	-	**good**	120
chr17	45977404	45977524	781717_44404225_MAPT_425	1	-	**good**	120
chr17	45977524	45977644	781717_44404225_MAPT_424	1	-	**good**	120
chr17	45977644	45977764	781717_44404225_MAPT_423	1	-	**good**	120
chr17	45977764	45977884	781717_44404225_MAPT_422	1	-	**good**	120
chr17	45977884	45978004	781717_44404225_MAPT_421	1	-	**good**	120
chr17	45978004	45978124	781717_44404225_MAPT_420	1	-	**good**	120
chr17	45978124	45978244	781717_44404225_MAPT_419	1	-	**good**	120
chr17	45978244	45978364	781717_44404225_MAPT_418	1	-	**good**	120
chr17	45978364	45978484	781717_44404225_MAPT_417	1	-	**good**	120
chr17	45978484	45978604	781717_44404225_MAPT_416	1	-	**good**	120
chr17	45978604	45978724	781717_44404225_MAPT_415	1	-	**good**	120
chr17	45979084	45979204	781717_44404225_MAPT_411	1	-	**good**	120
chr17	45979204	45979324	781717_44404225_MAPT_410	1	-	**good**	120
chr17	45979324	45979444	781717_44404225_MAPT_409	1	-	**good**	120
chr17	45979444	45979564	781717_44404225_MAPT_408	1	-	**good**	120
chr17	45979564	45979684	781717_44404225_MAPT_407	1	-	**good**	120
chr17	45979684	45979804	781717_44404225_MAPT_406	1	-	**good**	120
chr17	45979804	45979924	781717_44404225_MAPT_405	1	-	**good**	120
chr17	45979924	45980044	781717_44404225_MAPT_404	1	-	**good**	120
chr17	45980044	45980164	781717_44404225_MAPT_403	1	-	**good**	120
chr17	45980524	45980644	781717_44404225_MAPT_399	1	-	**good**	120
chr17	45980644	45980764	781717_44404225_MAPT_398	1	-	**good**	120
chr17	45980764	45980884	781717_44404225_MAPT_397	1	-	**good**	120
chr17	45980884	45981004	781717_44404225_MAPT_396	1	-	**good**	120
chr17	45981004	45981124	781717_44404225_MAPT_395	1	-	**good**	120
chr17	45981124	45981244	781717_44404225_MAPT_394	1	-	**good**	120
chr17	45981244	45981364	781717_44404225_MAPT_393	1	-	**good**	120
chr17	45981364	45981484	781717_44404225_MAPT_392	1	-	**good**	120
chr17	45981484	45981604	781717_44404225_MAPT_391	1	-	**good**	120
chr17	45981604	45981724	781717_44404225_MAPT_390	1	-	**good**	120
chr17	45982324	45982444	781717_44404225_MAPT_384	1	-	**good**	120
chr17	45982444	45982564	781717_44404225_MAPT_383	1	-	**good**	120
chr17	45982564	45982684	781717_44404225_MAPT_382	1	-	**good**	120
chr17	45982684	45982804	781717_44404225_MAPT_381	1	-	**good**	120

chr17	45982804	45982924	781717_44404225_MAPT_380	1	-	**good**	120
chr17	45982924	45983044	781717_44404225_MAPT_379	1	-	**good**	120
chr17	45983044	45983164	781717_44404225_MAPT_378	1	-	**good**	120
chr17	45983164	45983284	781717_44404225_MAPT_377	1	-	**good**	120
chr17	45983284	45983404	781717_44404225_MAPT_376	1	-	**good**	120
chr17	45983404	45983524	781717_44404225_MAPT_375	1	-	**good**	120
chr17	45983524	45983644	781717_44404225_MAPT_374	1	-	**good**	120
chr17	45983644	45983764	781717_44404225_MAPT_373	1	-	**good**	120
chr17	45983764	45983884	781717_44404225_MAPT_372	1	-	**good**	120
chr17	45983884	45984004	781717_44404225_MAPT_371	1	-	**good**	120
chr17	45984004	45984124	781717_44404225_MAPT_370	1	-	**good**	120
chr17	45984124	45984244	781717_44404225_MAPT_369	1	-	**good**	120
chr17	45984244	45984364	781717_44404225_MAPT_368	1	-	**good**	120
chr17	45984364	45984484	781717_44404225_MAPT_367	1	-	**good**	120
chr17	45984484	45984604	781717_44404225_MAPT_366	1	-	**good**	120
chr17	45984604	45984724	781717_44404225_MAPT_365	1	-	**good**	120
chr17	45984724	45984844	781717_44404225_MAPT_364	1	-	**good**	120
chr17	45984844	45984964	781717_44404225_MAPT_363	1	-	**good**	120
chr17	45985324	45985444	781717_44404225_MAPT_359	1	-	**good**	120
chr17	45985444	45985564	781717_44404225_MAPT_358	1	-	**good**	120
chr17	45985564	45985684	781717_44404225_MAPT_357	1	-	**good**	120
chr17	45985684	45985804	781717_44404225_MAPT_356	1	-	**good**	120
chr17	45985804	45985924	781717_44404225_MAPT_355	1	-	**good**	120
chr17	45985924	45986044	781717_44404225_MAPT_354	1	-	**good**	120
chr17	45986044	45986164	781717_44404225_MAPT_353	1	-	**good**	120
chr17	45986164	45986284	781717_44404225_MAPT_352	1	-	**good**	120
chr17	45986284	45986404	781717_44404225_MAPT_351	1	-	**good**	120
chr17	45986404	45986524	781717_44404225_MAPT_350	1	-	**good**	120
chr17	45986524	45986644	781717_44404225_MAPT_349	1	-	**good**	120
chr17	45986644	45986764	781717_44404225_MAPT_348	1	-	**good**	120
chr17	45986764	45986884	781717_44404225_MAPT_347	1	-	**good**	120
chr17	45986884	45987004	781717_44404225_MAPT_346	1	-	**good**	120
chr17	45987004	45987124	781717_44404225_MAPT_345	1	-	**good**	120
chr17	45987124	45987244	781717_44404225_MAPT_344	1	-	**good**	120
chr17	45987244	45987364	781717_44404225_MAPT_343	1	-	**good**	120
chr17	45987364	45987484	781717_44404225_MAPT_342	1	-	**good**	120
chr17	45987484	45987604	781717_44404225_MAPT_341	1	-	**good**	120
chr17	45987604	45987724	781717_44404225_MAPT_340	1	-	**good**	120
chr17	45987724	45987844	781717_44404225_MAPT_339	1	-	**good**	120
chr17	45987844	45987964	781717_44404225_MAPT_338	1	-	**good**	120
chr17	45987964	45988084	781717_44404225_MAPT_337	1	-	**good**	120
chr17	45988084	45988204	781717_44404225_MAPT_336	1	-	**good**	120
chr17	45988204	45988324	781717_44404225_MAPT_335	1	-	**good**	120
chr17	45988324	45988444	781717_44404225_MAPT_334	1	-	**good**	120
chr17	45988444	45988564	781717_44404225_MAPT_333	1	-	**good**	120
chr17	45988924	45989044	781717_44404225_MAPT_329	1	-	**good**	120

chr17	45989044	45989164	781717_44404225_MAPT_328	1	-	**good**	120
chr17	45989164	45989284	781717_44404225_MAPT_327	1	-	**good**	120
chr17	45989284	45989404	781717_44404225_MAPT_326	31	-	**good**	120
chr17	45989644	45989764	781717_44404225_MAPT_323	1	-	**good**	120
chr17	45989764	45989884	781717_44404225_MAPT_322	1	-	**good**	120
chr17	45989884	45990004	781717_44404225_MAPT_321	1	-	**good**	120
chr17	45990004	45990124	781717_44404225_MAPT_320	1	-	**good**	120
chr17	45990124	45990244	781717_44404225_MAPT_319	1	-	**good**	120
chr17	45990244	45990364	781717_44404225_MAPT_318	3	-	**good**	120
chr17	45990604	45990724	781717_44404225_MAPT_315	1	-	**good**	120
chr17	45990724	45990844	781717_44404225_MAPT_314	1	-	**good**	120
chr17	45990844	45990964	781717_44404225_MAPT_313	1	-	**good**	120
chr17	45990964	45991084	781717_44404225_MAPT_312	1	-	**good**	120
chr17	45991084	45991204	781717_44404225_MAPT_311	1	-	**good**	120
chr17	45991204	45991324	781717_44404225_MAPT_310	1	-	**good**	120
chr17	45991324	45991444	781717_44404225_MAPT_309	1	-	**good**	120
chr17	45991444	45991564	781717_44404225_MAPT_308	1	-	**good**	120
chr17	45991564	45991684	781717_44404225_MAPT_307	1	-	**good**	120
chr17	45991684	45991804	781717_44404225_MAPT_306	1	-	**good**	120
chr17	45992164	45992284	781717_44404225_MAPT_302	1	-	**good**	120
chr17	45992284	45992404	781717_44404225_MAPT_301	1	-	**good**	120
chr17	45992404	45992524	781717_44404225_MAPT_300	1	-	**good**	120
chr17	45992524	45992644	781717_44404225_MAPT_299	36	-	**good**	120
chr17	45992884	45993004	781717_44404225_MAPT_296	1	-	**good**	120
chr17	45993004	45993124	781717_44404225_MAPT_295	1	-	**good**	120
chr17	45993124	45993244	781717_44404225_MAPT_294	1	-	**good**	120
chr17	45993244	45993364	781717_44404225_MAPT_293	1	-	**good**	120
chr17	45993724	45993844	781717_44404225_MAPT_289	1	-	**good**	120
chr17	45993844	45993964	781717_44404225_MAPT_288	1	-	**good**	120
chr17	45993964	45994084	781717_44404225_MAPT_287	1	-	**good**	120
chr17	45994084	45994204	781717_44404225_MAPT_286	1	-	**good**	120
chr17	45994204	45994324	781717_44404225_MAPT_285	2	-	**good**	120
chr17	45994324	45994444	781717_44404225_MAPT_284	1	-	**good**	120
chr17	45994444	45994564	781717_44404225_MAPT_283	1	-	**good**	120
chr17	45994804	45994924	781717_44404225_MAPT_280	50	-	**good**	120
chr17	45995044	45995164	781717_44404225_MAPT_278	2	-	**good**	120
chr17	45995164	45995284	781717_44404225_MAPT_277	1	-	**good**	120
chr17	45995284	45995404	781717_44404225_MAPT_276	1	-	**good**	120
chr17	45995404	45995524	781717_44404225_MAPT_275	1	-	**good**	120
chr17	45995524	45995644	781717_44404225_MAPT_274	1	-	**good**	120
chr17	45995644	45995764	781717_44404225_MAPT_273	1	-	**good**	120
chr17	45995764	45995884	781717_44404225_MAPT_272	1	-	**good**	120
chr17	45995884	45996004	781717_44404225_MAPT_271	1	-	**good**	120
chr17	45996004	45996124	781717_44404225_MAPT_270	1	-	**good**	120
chr17	45996124	45996244	781717_44404225_MAPT_269	1	-	**good**	120
chr17	45996244	45996364	781717_44404225_MAPT_268	1	-	**good**	120

chr17	45996364	45996484	781717_44404225_MAPT_267	1	-	**good**	120
chr17	45996484	45996604	781717_44404225_MAPT_266	1	-	**good**	120
chr17	45996604	45996724	781717_44404225_MAPT_265	1	-	**good**	120
chr17	45996724	45996844	781717_44404225_MAPT_264	1	-	**good**	120
chr17	45996844	45996964	781717_44404225_MAPT_263	1	-	**good**	120
chr17	45996964	45997084	781717_44404225_MAPT_262	1	-	**good**	120
chr17	45997084	45997204	781717_44404225_MAPT_261	1	-	**good**	120
chr17	45997204	45997324	781717_44404225_MAPT_260	1	-	**good**	120
chr17	45997324	45997444	781717_44404225_MAPT_259	3	-	**good**	120
chr17	45997924	45998044	781717_44404225_MAPT_254	1	-	**good**	120
chr17	45998044	45998164	781717_44404225_MAPT_253	1	-	**good**	120
chr17	45998164	45998284	781717_44404225_MAPT_252	1	-	**good**	120
chr17	45998284	45998404	781717_44404225_MAPT_251	1	-	**good**	120
chr17	45998404	45998524	781717_44404225_MAPT_250	1	-	**good**	120
chr17	45998524	45998644	781717_44404225_MAPT_249	1	-	**good**	120
chr17	45998644	45998764	781717_44404225_MAPT_248	1	-	**good**	120
chr17	45998764	45998884	781717_44404225_MAPT_247	1	-	**good**	120
chr17	45998884	45999004	781717_44404225_MAPT_246	1	-	**good**	120
chr17	45999004	45999124	781717_44404225_MAPT_245	1	-	**good**	120
chr17	45999124	45999244	781717_44404225_MAPT_244	1	-	**good**	120
chr17	45999244	45999364	781717_44404225_MAPT_243	1	-	**good**	120
chr17	45999364	45999484	781717_44404225_MAPT_242	1	-	**good**	120
chr17	45999484	45999604	781717_44404225_MAPT_241	1	-	**good**	120
chr17	45999604	45999724	781717_44404225_MAPT_240	1	-	**good**	120
chr17	45999724	45999844	781717_44404225_MAPT_239	1	-	**good**	120
chr17	45999844	45999964	781717_44404225_MAPT_238	1	-	**good**	120
chr17	45999964	46000084	781717_44404225_MAPT_237	1	-	**good**	120
chr17	46000084	46000204	781717_44404225_MAPT_236	1	-	**good**	120
chr17	46000204	46000324	781717_44404225_MAPT_235	1	-	**good**	120
chr17	46000324	46000444	781717_44404225_MAPT_234	1	-	**good**	120
chr17	46000444	46000564	781717_44404225_MAPT_233	1	-	**good**	120
chr17	46000564	46000684	781717_44404225_MAPT_232	1	-	**good**	120
chr17	46000684	46000804	781717_44404225_MAPT_231	1	-	**good**	120
chr17	46000804	46000924	781717_44404225_MAPT_230	1	-	**good**	120
chr17	46001164	46001284	781717_44404225_MAPT_227	2	-	**good**	120
chr17	46001284	46001404	781717_44404225_MAPT_226	1	-	**good**	120
chr17	46001404	46001524	781717_44404225_MAPT_225	1	-	**good**	120
chr17	46001644	46001764	781717_44404225_MAPT_223	1	-	**good**	120
chr17	46001764	46001884	781717_44404225_MAPT_222	1	-	**good**	120
chr17	46001884	46002004	781717_44404225_MAPT_221	1	-	**good**	120
chr17	46002004	46002124	781717_44404225_MAPT_220	1	-	**good**	120
chr17	46002124	46002244	781717_44404225_MAPT_219	1	-	**good**	120
chr17	46002244	46002364	781717_44404225_MAPT_218	1	-	**good**	120
chr17	46002364	46002484	781717_44404225_MAPT_217	1	-	**good**	120
chr17	46002484	46002604	781717_44404225_MAPT_216	1	-	**good**	120
chr17	46002604	46002724	781717_44404225_MAPT_215	1	-	**good**	120

chr17	46003084	46003204	781717_44404225_MAPT_211	1	-	**good**	120
chr17	46003564	46003684	781717_44404225_MAPT_207	1	-	**good**	120
chr17	46003684	46003804	781717_44404225_MAPT_206	1	-	**good**	120
chr17	46003804	46003924	781717_44404225_MAPT_205	1	-	**good**	120
chr17	46003924	46004044	781717_44404225_MAPT_204	1	-	**good**	120
chr17	46004044	46004164	781717_44404225_MAPT_203	1	-	**good**	120
chr17	46004164	46004284	781717_44404225_MAPT_202	1	-	**good**	120
chr17	46004284	46004404	781717_44404225_MAPT_201	1	-	**good**	120
chr17	46004404	46004524	781717_44404225_MAPT_200	1	-	**good**	120
chr17	46004524	46004644	781717_44404225_MAPT_199	1	-	**good**	120
chr17	46004644	46004764	781717_44404225_MAPT_198	1	-	**good**	120
chr17	46005124	46005244	781717_44404225_MAPT_194	1	-	**good**	120
chr17	46005244	46005364	781717_44404225_MAPT_193	1	-	**good**	120
chr17	46005364	46005484	781717_44404225_MAPT_192	1	-	**good**	120
chr17	46005604	46005724	781717_44404225_MAPT_190	1	-	**good**	120
chr17	46005724	46005844	781717_44404225_MAPT_189	1	-	**good**	120
chr17	46005844	46005964	781717_44404225_MAPT_188	1	-	**good**	120
chr17	46005964	46006084	781717_44404225_MAPT_187	1	-	**good**	120
chr17	46007524	46007644	781717_44404225_MAPT_174	1	-	**good**	120
chr17	46007644	46007764	781717_44404225_MAPT_173	1	-	**good**	120
chr17	46007764	46007884	781717_44404225_MAPT_172	1	-	**good**	120
chr17	46007884	46008004	781717_44404225_MAPT_171	1	-	**good**	120
chr17	46008004	46008124	781717_44404225_MAPT_170	1	-	**good**	120
chr17	46008364	46008484	781717_44404225_MAPT_167	1	-	**good**	120
chr17	46008484	46008604	781717_44404225_MAPT_166	1	-	**good**	120
chr17	46008604	46008724	781717_44404225_MAPT_165	1	-	**good**	120
chr17	46008724	46008844	781717_44404225_MAPT_164	1	-	**good**	120
chr17	46008844	46008964	781717_44404225_MAPT_163	1	-	**good**	120
chr17	46009204	46009324	781717_44404225_MAPT_160	1	-	**good**	120
chr17	46009324	46009444	781717_44404225_MAPT_159	1	-	**good**	120
chr17	46009444	46009564	781717_44404225_MAPT_158	1	-	**good**	120
chr17	46009564	46009684	781717_44404225_MAPT_157	1	-	**good**	120
chr17	46009684	46009804	781717_44404225_MAPT_156	1	-	**good**	120
chr17	46009804	46009924	781717_44404225_MAPT_155	1	-	**good**	120
chr17	46009924	46010044	781717_44404225_MAPT_154	1	-	**good**	120
chr17	46010044	46010164	781717_44404225_MAPT_153	1	-	**good**	120
chr17	46010164	46010284	781717_44404225_MAPT_152	2	-	**good**	120
chr17	46010284	46010404	781717_44404225_MAPT_151	1	-	**good**	120
chr17	46010404	46010524	781717_44404225_MAPT_150	1	-	**good**	120
chr17	46010524	46010644	781717_44404225_MAPT_149	1	-	**good**	120
chr17	46010644	46010764	781717_44404225_MAPT_148	1	-	**good**	120
chr17	46010764	46010884	781717_44404225_MAPT_147	1	-	**good**	120
chr17	46010884	46011004	781717_44404225_MAPT_146	1	-	**good**	120
chr17	46011004	46011124	781717_44404225_MAPT_145	1	-	**good**	120
chr17	46011364	46011484	781717_44404225_MAPT_142	4	-	**good**	120
chr17	46011484	46011604	781717_44404225_MAPT_141	1	-	**good**	120

chr17	46011604	46011724	781717_44404225_MAPT_140	1	-	**good**	120
chr17	46011724	46011844	781717_44404225_MAPT_139	1	-	**good**	120
chr17	46011844	46011964	781717_44404225_MAPT_138	1	-	**good**	120
chr17	46011964	46012084	781717_44404225_MAPT_137	1	-	**good**	120
chr17	46012084	46012204	781717_44404225_MAPT_136	1	-	**good**	120
chr17	46012204	46012324	781717_44404225_MAPT_135	1	-	**good**	120
chr17	46012324	46012444	781717_44404225_MAPT_134	1	-	**good**	120
chr17	46012444	46012564	781717_44404225_MAPT_133	1	-	**good**	120
chr17	46012564	46012684	781717_44404225_MAPT_132	1	-	**good**	120
chr17	46012684	46012804	781717_44404225_MAPT_131	1	-	**good**	120
chr17	46012804	46012924	781717_44404225_MAPT_130	1	-	**good**	120
chr17	46012924	46013044	781717_44404225_MAPT_129	1	-	**good**	120
chr17	46013044	46013164	781717_44404225_MAPT_128	1	-	**good**	120
chr17	46013164	46013284	781717_44404225_MAPT_127	1	-	**good**	120
chr17	46013284	46013404	781717_44404225_MAPT_126	1	-	**good**	120
chr17	46013404	46013524	781717_44404225_MAPT_125	1	-	**good**	120
chr17	46013524	46013644	781717_44404225_MAPT_124	1	-	**good**	120
chr17	46013644	46013764	781717_44404225_MAPT_123	1	-	**good**	120
chr17	46013764	46013884	781717_44404225_MAPT_122	1	-	**good**	120
chr17	46013884	46014004	781717_44404225_MAPT_121	1	-	**good**	120
chr17	46014004	46014124	781717_44404225_MAPT_120	1	-	**good**	120
chr17	46014124	46014244	781717_44404225_MAPT_119	1	-	**good**	120
chr17	46014244	46014364	781717_44404225_MAPT_118	1	-	**good**	120
chr17	46014364	46014484	781717_44404225_MAPT_117	1	-	**good**	120
chr17	46014484	46014604	781717_44404225_MAPT_116	1	-	**good**	120
chr17	46015924	46016044	781717_44404225_MAPT_104	1	-	**good**	120
chr17	46016044	46016164	781717_44404225_MAPT_103	9	-	**good**	120
chr17	46016404	46016524	781717_44404225_MAPT_100	11	-	**good**	120
chr17	46016764	46016884	781717_44404225_MAPT_97	13	-	**good**	120
chr17	46016884	46017004	781717_44404225_MAPT_96	1	-	**good**	120
chr17	46017004	46017124	781717_44404225_MAPT_95	1	-	**good**	120
chr17	46017124	46017244	781717_44404225_MAPT_94	1	-	**good**	120
chr17	46017724	46017844	781717_44404225_MAPT_89	8	-	**good**	120
chr17	46018204	46018324	781717_44404225_MAPT_85	1	-	**good**	120
chr17	46018324	46018444	781717_44404225_MAPT_84	1	-	**good**	120
chr17	46018444	46018564	781717_44404225_MAPT_83	1	-	**good**	120
chr17	46018564	46018684	781717_44404225_MAPT_82	1	-	**good**	120
chr17	46018684	46018804	781717_44404225_MAPT_81	1	-	**good**	120
chr17	46018804	46018924	781717_44404225_MAPT_80	1	-	**good**	120
chr17	46018924	46019044	781717_44404225_MAPT_79	1	-	**good**	120
chr17	46019644	46019764	781717_44404225_MAPT_73	47	-	**good**	120
chr17	46020004	46020124	781717_44404225_MAPT_70	1	-	**good**	120
chr17	46020124	46020244	781717_44404225_MAPT_69	1	-	**good**	120
chr17	46020244	46020364	781717_44404225_MAPT_68	1	-	**good**	120
chr17	46020364	46020484	781717_44404225_MAPT_67	1	-	**good**	120
chr17	46020844	46020964	781717_44404225_MAPT_63	1	-	**good**	120

chr17	46020964	46021084	781717_44404225_MAPT_62	1	-	**good**	120
chr17	46021084	46021204	781717_44404225_MAPT_61	1	-	**good**	120
chr17	46021204	46021324	781717_44404225_MAPT_60	1	-	**good**	120
chr17	46021324	46021444	781717_44404225_MAPT_59	1	-	**good**	120
chr17	46021444	46021564	781717_44404225_MAPT_58	1	-	**good**	120
chr17	46021564	46021684	781717_44404225_MAPT_57	1	-	**good**	120
chr17	46021684	46021804	781717_44404225_MAPT_56	1	-	**good**	120
chr17	46021804	46021924	781717_44404225_MAPT_55	1	-	**good**	120
chr17	46021924	46022044	781717_44404225_MAPT_54	1	-	**good**	120
chr17	46022404	46022524	781717_44404225_MAPT_50	1	-	**good**	120
chr17	46022524	46022644	781717_44404225_MAPT_49	1	-	**good**	120
chr17	46022644	46022764	781717_44404225_MAPT_48	1	-	**good**	120
chr17	46022884	46023004	781717_44404225_MAPT_46	5	-	**good**	120
chr17	46023004	46023124	781717_44404225_MAPT_45	1	-	**good**	120
chr17	46023124	46023244	781717_44404225_MAPT_44	1	-	**good**	120
chr17	46023244	46023364	781717_44404225_MAPT_43	1	-	**good**	120
chr17	46023364	46023484	781717_44404225_MAPT_42	1	-	**good**	120
chr17	46023844	46023964	781717_44404225_MAPT_38	1	-	**good**	120
chr17	46023964	46024084	781717_44404225_MAPT_37	1	-	**good**	120
chr17	46024084	46024204	781717_44404225_MAPT_36	1	-	**good**	120
chr17	46024204	46024324	781717_44404225_MAPT_35	1	-	**good**	120
chr17	46024324	46024444	781717_44404225_MAPT_34	1	-	**good**	120
chr17	46024444	46024564	781717_44404225_MAPT_33	1	-	**good**	120
chr17	46024564	46024684	781717_44404225_MAPT_32	1	-	**good**	120
chr17	46024684	46024804	781717_44404225_MAPT_31	1	-	**good**	120
chr17	46024804	46024924	781717_44404225_MAPT_30	1	-	**good**	120
chr17	46024924	46025044	781717_44404225_MAPT_29	1	-	**good**	120
chr17	46025044	46025164	781717_44404225_MAPT_28	1	-	**good**	120
chr17	46025164	46025284	781717_44404225_MAPT_27	1	-	**good**	120
chr17	46025284	46025404	781717_44404225_MAPT_26	1	-	**good**	120
chr17	46025404	46025524	781717_44404225_MAPT_25	1	-	**good**	120
chr17	46025524	46025644	781717_44404225_MAPT_24	1	-	**good**	120
chr17	46025644	46025764	781717_44404225_MAPT_23	1	-	**good**	120
chr17	46025764	46025884	781717_44404225_MAPT_22	1	-	**good**	120
chr17	46025884	46026004	781717_44404225_MAPT_21	1	-	**good**	120
chr17	46026004	46026124	781717_44404225_MAPT_20	1	-	**good**	120
chr17	46026124	46026244	781717_44404225_MAPT_19	1	-	**good**	120
chr17	46026244	46026364	781717_44404225_MAPT_18	1	-	**good**	120
chr17	46026364	46026484	781717_44404225_MAPT_17	1	-	**good**	120
chr17	46026484	46026604	781717_44404225_MAPT_16	1	-	**good**	120
chr17	46026604	46026724	781717_44404225_MAPT_15	1	-	**good**	120
chr17	46026724	46026844	781717_44404225_MAPT_14	1	-	**good**	120
chr17	46026844	46026964	781717_44404225_MAPT_13	1	-	**good**	120
chr17	46026964	46027084	781717_44404225_MAPT_12	1	-	**good**	120
chr17	46027084	46027204	781717_44404225_MAPT_11	1	-	**good**	120
chr17	46027204	46027324	781717_44404225_MAPT_10	1	-	**good**	120

chr17	46027324	46027444	781717_44404225_MAPT_9	1	-	**good**	120
chr17	46027444	46027564	781717_44404225_MAPT_8	1	-	**good**	120
chr17	46027564	46027684	781717_44404225_MAPT_7	1	-	**good**	120
chr17	46027684	46027804	781717_44404225_MAPT_6	1	-	**good**	120
chr17	46027804	46027924	781717_44404225_MAPT_5	1	-	**good**	120
chr17	46027924	46028044	781717_44404225_MAPT_4	1	-	**good**	120
chr17	46028044	46028164	781717_44404225_MAPT_3	1	-	**good**	120
chr17	46028164	46028284	781717_44404225_MAPT_2	1	-	**good**	120
chr17	46028284	46028404	781717_44404225_MAPT_1	1	-	**good**	120

Appendix VII

Per-bp mutation rate calculations: (a) per sample and (b) per genomic region per sample

a) The table below describes the calculations for overall per-bp mutation rate for each sample described in Chapter 7, section 7.2.3:

Sample	Template	Depths	Altered nucleotides				No. of mutations detected	Total sum of mutations per sample	Number of each nucleotide (A, T, G, C) sequenced	Total sum of all nucleotides sequenced per sample	Per-bp mutation rate per sample
			T-count	C-count	A-count	G-count					
C_M	A	672028	6406	6654	0	20399	33459		688115		
	C	638747	15782	0	7786	4203	27771		647348		
	G	679135	5404	10668	33040	0	49112		656842		
	T	722137	0	19050	8720	2217	29987	29987	719742	2712047	**1.11e-02**
C_O	A	455264	7708	7063	0	27471	42242		480913		
	C	445923	19609	0	10481	3888	33978		455038		
	G	484707	6049	14492	46199	0	66740		450408		
	T	347264	0	21538	11211	1082	33831	33831	346799	1733158	**1.95e-02**
C_T	A	79859	1043	1347	0	3781	6171		82930		
	C	83650	2210	0	1447	186	3843		85272		
	G	78256	759	1335	6304	0	8398		74054		
	T	85083	0	2783	1491	229	4503	4503	84592	326848	**1.38e-02**
NP_M	A	1059337	29737	37612	0	197867	265216		1027019		
	C	876048	122446	0	26727	37900	187073		942015		
	G	991720	44088	35881	182526	0	262495		1000814		
	T	1007118	0	179547	23645	35822	239014	239014	964375	3934223	**6.08e-02**
NP_O	A	1336685	46204	56657	0	292814	395675		1278015		
	C	1129258	176766	0	38260	56755	271781		1235421		
	G	1430217	64868	51967	265051	0	381886		1448936		
	T	1293915	0	269320	33694	51036	354050	354050	1227703	5190075	**6.82e-02**
NP_T	A	1010114	34428	42254	0	217007	293689		963958		
	C	828293	130392	0	28401	41639	200432		907752		
	G	999372	47295	38301	195063	0	280659		1015533		
	T	922957	0	199336	24069	38174	261579	261579	873493	3760736	**6.96e-02**

Table 35: Overall per-bp mutation rate calculations per sample.
A: Adenine; C576 = Control; C_M = C576 Med; C_O = C576 Occ; C_T = C756 Temp; C: Cytosine, G = Guanine; NP_M = NP16.163 Med; NP_O = NP16.163 Occ; NP_T = NP16.163 Temp; NP16.163 = FTD-Tau disease patient; Med = Medulla, Occ = Occipital lobe; Temp = Temporal lobe; T = Thymine.

The following equations were used to determine the values of:

- Number of mutations detected = T-count + C-count + A-count + G-count (e.g., values of A>T, A>C, A>A, and A>G SNVs can be calculated this way).
- Total sum of mutations per sample = A>T + A>C+ A>A (value = 0) + A>G + T>T (value = 0) + T>C + T>A + T>G + C>T +C>C (value = 0) + C>A + C>G + G>T + G>C + G>A + G>G (value = 0) [highlighted in yellow].
- Number of each nucleotide sequenced:

- Aseq = [Depth of (template = A) – (T-count + C-count + A-count + G-count)] + A-count of (template = C) + A-count of (template = G) + A-count of (template = T).
- Similarly, Cseq, Gseq, and Tseq were determined.

- Total sum of all nucleotides sequenced per sample = Aseq + Cseq + Gseq + Tseq [highlighted in blue].
- Per-bp mutation rate per sample = $\frac{Total\ sum\ of\ mutations\ per\ sample}{Total\ sum\ of\ all\ nucleotides\ sequenced\ per\ sample}$ [highlighted in **green font**].

b) The table below describes the calculations for overall per-bp mutation rate per genomic region for each sample described in Chapter 7, section 7.2.3:

Sample	Genomic Region of MAPT gene	Template	Depths	Altered nucleotides				No. of mutations detected per genomic region	No. of nucleotides sequenced per genomic region	Total no. of nucleotides sequenced per genomic region	Per-bp mutation rate per genomic region per sample
				T-count	C-count	A-count	G-count				
C_M	UTR3	A	31243	493	1	0	2091	2585	28901		
		C	26183	111	0	89	105	305	30960		
		G	32594	1	169	121	0	291	34505		
		T	22100	0	4912	33	6	4951	17754		
		Sum of mutations detected in UTR3 region:						**8132**	**Sum:**	**112120**	**3.00e-03**
	Exonic	A	14637	0	0	0	131	131	14643		
		C	17431	99	0	88	22	209	19758		
		G	4925	0	238	0	0	238	5018		
		T	28905	0	2298	49	178	2525	26479		
		Sum of mutations detected in exonic region:						**3103**	**Sum:**	**65898**	**1.14e-03**
	intronic	A	608996	5846	6599	0	18116	30561	627153		
		C	538024	15415	0	7461	3793	26669	539569		
		G	593361	5284	10004	32733	0	48021	568929		
		T	603092	0	11611	8524	1680	21815	607822		
		Sum of mutations detected in intronic region:						**127066**	**Sum:**	**2343473**	**4.69e-02**
	ncRNA exonic	A	8728	67	54	0	1	122	8842		
		C	25571	31	0	44	210	285	25695		
		G	37315	81	257	116	0	454	37221		
		T	35063	0	98	76	149	323	34919		
		Sum of mutations detected in ncRNA exonic region:						**1184**	**Sum:**	**106677**	**4.37e-04**
	ncRNA intronic	A	8424	0	0	0	60	60	8576		
		C	31538	126	0	104	73	303	31366		
		G	10940	38	0	70	0	108	11112		
		T	28675	0	131	38	147	316	28523		
		Sum of mutations detected in ncRNA intronic region:						**787**	**Sum:**	**79577**	**2.90e-04**
	Splicing	T	4302	0	0	0	57	57	4245	**4245**	**2.10e-05**
		Sum of mutations detected in splicing region:						**57**	**Total sum of all nucleotides sequenced per sample:**	**2711990**	
C_O	UTR3	A	31630	549	1	0	3176	3726	28001		
		C	14343	36	0	39	32	107	21205		
		G	22512	5	88	58	0	151	25569		
		T	6881	0	6880	0	0	6880	591		
		Sum of mutations detected in UTR3 region:						**10864**	**Sum:**	**75366**	**6.27e-03**
	exonic	A	11948	0	0	0	63	63	11989		
		C	7412	0	0	104	0	104	11295		
		G	7650	0	281	0	0	281	7584		
		T	21140	0	3706	0	152	3858	17282		
		Sum of mutations detected in exonic region:						**4306**	**Sum:**	**48150**	**2.48e-03**
	intronic	A	399949	7098	7015	0	24230	38343	428992		
		C	391004	19551	0	10293	3637	33481	389290		
		G	423867	5956	13947	46002	0	65905	386568		
		T	264060	0	10805	11091	739	22635	274030		
		Sum of mutations detected in intronic region:						**160364**	**Sum:**	**1478880**	**9.25e-02**

	ncRNA exonic	A	11737	61	47	0	2	110	11870		
		C	27285	2	0	27	219	248	27325		
		G	29093	83	176	122	0	381	29021		
		T	31072	0	65	94	88	247	30971		
			Sum of mutations detected in ncRNA exonic region:					**986**	**Sum:**	**99187**	**5.69e-04**
	ncRNA intronic	C	5879	20	0	18	0	38	5846		
		G	1585	5	0	17	0	22	1601		
		T	17628	0	82	26	38	146	17507		
			Sum of mutations detected in ncRNA intronic region:					**206**	**Sum:**	**24954**	**1.19e-04**
	Splicing	T	6483	0	0	0	65	65	6418	**6418**	**3.75e-05**
			Sum of mutations detected in splicing region:					**65**	**Total sum of all nucleotides sequenced per sample:**	**1732955**	
C_T	UTR3	A	3921	26	0	0	401	427	3539		
		C	3526	8	0	20	0	28	3524		
		G	5198	0	17	25	0	42	5568		
		T	4907	0	1141	0	11	1152	3789		
			Sum of mutations detected in UTR3 region:					**1649**	**Sum:**	**16420**	**5.06e-03**
	exonic	A	3022	0	0	0	17	17	3040		
		C	2133	6	0	25	0	31	2658		
		G	1116	0	37	0	0	37	1111		
		T	5302	0	519	10	15	544	4764		
			Sum of mutations detected in exonic region:					**629**	**Sum:**	**11573**	**1.93e-03**
	intronic	A	71252	1017	1332	0	3358	5707	74689		
		C	73715	2173	0	1402	172	3747	73680		
		G	69442	751	1273	6272	0	8296	64832		
		T	66306	0	1107	1470	156	2733	67514		
			Sum of mutations detected in intronic region:					**20483**	**Sum:**	**280715**	**6.29e-02**
	ncRNA exonic	A	793	0	15	0	0	15	796		
		C	2432	5	0	0	14	19	2445		
		G	2500	8	8	7	0	23	2517		
		T	4769	0	9	11	26	46	4736		
			Sum of mutations detected in ncRNA exonic region:					**103**	**Sum:**	**10494**	**3.16e-04**
	ncRNA intronic	A	871	0	0	0	5	5	866		
		C	1844	18	0	0	0	18	1833		
		T	2900	0	7	0	13	20	2898		
			Sum of mutations detected in ncRNA intronic region:					**43**	**Sum:**	**5597**	**1.32e-04**
	Splicing	T	899	0	0	0	8	8	891	**891**	**2.46e-05**
			Sum of mutations detected in splicing region:					**8**	**Total sum of all nucleotides sequenced per sample:**	**325690**	
NP_M	UTR3	A	35306	491	4266	0	6948	11705	27660		
		C	28093	2450	0	74	69	2593	48846		
		G	28911	0	2101	3985	0	6086	29891		
		T	54745	0	16979	0	49	17028	40658		
			Sum of mutations detected in UTR3 region:					**37412**	**Sum:**	147055	**9.52e-03**
	UTR5	A	4570	0	0	0	2298	2298	2272		
			Sum of mutations detected in UTR5 region:					**2298**	**Sum:**	2272	**5.84e-04**
	exonic	A	16846	0	0	0	4136	4136	16970		
	exonic	C	23803	4369	0	71	16	4456	27863		
	exonic	G	12992	0	165	4155	0	4320	12916		
	exonic	T	30620	0	8351	34	92	8477	26512		
			Sum of mutations detected in exonic region:					**21389**	**Sum:**	84261	**5.44e-03**
	intronic	A	995022	29246	33283	0	184460	246989	972384		
		C	791290	115534	0	26512	37720	179766	832448		
		G	922990	44027	33515	174302	0	251844	928797		
		T	876657	0	154126	23537	35471	213134	852330		
			Sum of mutations detected in intronic region:					**891733**	**Sum:**	3585959	**2.27e-01**
	ncRNA exonic	A	3410	0	63	0	0	63	3420		
		C	13980	52	0	21	36	109	14061		
		G	21418	51	100	28	0	179	21342		
		T	13463	0	27	24	67	118	13448		
			Sum of mutations detected in ncRNA exonic region:					**469**	**Sum:**	52271	**1.19e-04**
	ncRNA intronic	A	4183	0	0	0	25	25	4313		
		C	18882	41	0	49	59	149	18797		
		G	5409	10	0	56	0	66	5524		
		T	27449	0	64	50	97	211	27289		
			Sum of mutations detected in ncRNA intronic region:					**451**	**Sum:**	55923	**1.15e-04**
	Splicing	T	4184	0	0	0	46	46	4138	4138	**1.17e-05**
			Sum of mutations detected in splicing region:					**46**	**Total sum of all nucleotides sequenced per sample:**	**3931879**	
NP_O	UTR3	A	53876	812	6210	0	11196	18218	41397	196359	**1.07e-02**

		C	29491	3245	0	121	62	3428	60293		
		G	51145	6	3243	5618	0	8867	53582		
		T	61847	0	24777	0	46	24823	41087		
							Sum of mutations detected in UTR3 region:	**55336**	**Sum:**		
	UTR5	A	6974	0	0	0	3456	3456	3518		
							Sum of mutations detected in UTR5 region:	**3456**	**Sum:**	3518	**6.66e-04**
	exonic	A	32905	0	1	0	6210	6211	33338		
		C	31394	6701	0	87	37	6825	37679		
		G	20423	0	246	6505	0	6751	20031		
		T	47355	0	12863	52	112	13027	41029		
							Sum of mutations detected in exonic region:	**32814**	**Sum:**	132077	**6.33e-03**
	intronic	A	1230364	45355	50322	0	271952	367629	1187002		
		C	1018275	166678	0	37982	56377	261037	1087246		
		G	1308702	64718	48225	252728	0	365671	1321937		
		T	1119112	0	231461	33557	50577	315595	1080268		
							Sum of mutations detected in intronic region:	**1309932**	**Sum:**	4676453	**2.53e-01**
	ncRNA exonic	A	12566	37	124	0	0	161	12527		
		C	24229	1	0	43	140	184	24529		
		G	39617	132	253	79	0	464	39411		
		T	24568	0	107	0	118	225	24513		
							Sum of mutations detected in ncRNA exonic region:	**1034**	**Sum:**	100980	**1.99e-04**
	ncRNA intronic	C	25869	141	0	27	139	307	25673		
		G	10330	12	0	121	0	133	10450		
		T	34730	0	111	85	114	310	34573		
							Sum of mutations detected in ncRNA intronic region:	**750**	**Sum:**	70696	**1.45e-04**
	Splicing	T	6303	0	1	0	69	70	6233	6233	**1.35e-05**
							Sum of mutations detected in splicing region:	**70**	**Total sum of all nucleotides sequenced per sample:**	**5186316**	
NP_T	UTR3	A	40025	506	4661	0	8269	13436	30674		
		C	21464	2258	0	50	50	2358	44706		
		G	32874	2	2512	4035	0	6549	34646		
		T	41187	0	18427	0	2	18429	25524		
							Sum of mutations detected in UTR3 region:	**40772**	**Sum:**	135550	**1.08e-02**
	UTR5	A	5086	0	0	0	2463	2463	2623		
							Sum of mutations detected in UTR5 region:	**2463**	**Sum:**	2623	**6.55e-04**
	exonic	A	23842	0	0	0	4653	4653	24069		
		C	28020	4917	0	64	27	5008	32546		
		G	14943	0	194	4784	0	4978	14738		
		T	34475	0	9340	32	93	9465	29927		
							Sum of mutations detected in exonic region:	**24104**	**Sum:**	101280	**6.41e-03**
	intronic	A	932341	33922	37535	0	201598	273055	897621		
		C	741658	123078	0	28238	41425	192741	793331		
		G	914132	47177	35455	186106	0	268738	926256		
		T	797152	0	171424	23991	37839	233254	768075		
							Sum of mutations detected in intronic region:	**967788**	**Sum:**	3385283	**2.58e-01**
	ncRNA exonic	A	4906	0	58	0	1	59	4966		
		C	14396	31	0	48	35	114	14531		
		G	25153	77	139	71	0	287	24968		
		T	14123	0	52	0	66	118	14113		
							Sum of mutations detected in ncRNA exonic region:	**578**	**Sum:**	58578	**1.54e-04**
	ncRNA intronic	A	3914	0	0	0	23	23	4005		
		C	22755	108	0	1	102	211	22638		
		G	12270	39	1	67	0	107	12408		
		T	31149	0	93	46	120	259	31037		
							Sum of mutations detected in ncRNA intronic region:	**600**	**Sum:**	70088	**1.60e-04**
	Splicing	T	4871	0	0	0	54	54	4817	4817	**1.44e-05**
							Sum of mutations detected in splicing region:	**54**	**Total sum of all nucleotides sequenced per sample:**	**3758219**	

Table 36: Overall per-bp mutation rate per genomic region per sample calculations.

A = Adenine; C576 = Age-matched control; C_M = C576 Med; C_O = C576 Occ; C_T = C756 Temp; C = Cytosine; G = Guanine; Med = Medulla; ncRNA: Non-coding RNA; NP_M = NP16.163 Med; NP_O = NP16.163 Occ; NP_T = NP16.163 Temp; NP16.163 = FTD-Tau disease patient; Occ = Occipital lobe; RNA: Ribonucleic acid; Temp = Temporal lobe; T = Thymine; UTR: Untranslated regions.

The following equations were used to determine the values of:

- Number of mutations detected per genomic region per sample = T-count + C-count + A-count + G-count (e.g., values of A>T, A>C, A>A, and A>G SNVs can be calculated this way).
- Total sum of mutations per sample per genomic region per sample = A>T + A>C+ A>A (value = 0) + A>G + T>T (value = 0) + T>C + T>A + T>G + C>T +C>C (value = 0) + C>A + C>G + G>T + G>C + G>A + G>G (value = 0) [highlighted in yellow].
- Number of each nucleotide sequenced per genomic region per sample:
    - Aseq = [Depth of (template = A) – (T-count + C-count + A-count + G-count)] + A-count of (template = C) + A-count of (template = G) + A-count of (template = T).
    - Similarly, Cseq, Gseq, and Tseq were determined.
- Total sum of all nucleotides sequenced genomic region per sample = Aseq + Cseq + Gseq + Tseq.
- Total sum of all nucleotides sequenced per sample = sum of all mutations detected in (UTR3 + UTR5 + exonic + intronic + ncRNA exonic + ncRNA intronic + splicing) regions per sample [highlighted in blue].
- Per-bp mutation rate per sample = $\frac{Total\ sum\ of\ mutations\ per\ sample}{Total\ sum\ of\ all\ nucleotides\ sequenced\ per\ sample}$ [highlighted in **green font**].

****