

**ROLE OF NON-CODING RNAS IN AMYLOID BETA NEUROPATHOLOGY AND
ALZHEIMER'S DISEASE**

LUKE SAVILLE

B.Sc. Biochemistry, University of Lethbridge, 2020

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge, Lethbridge, AB

© Luke Saville, 2023

ROLE OF NON-CODING RNAS IN AMYLOID BETA NEUROPATHOLOGY AND
ALZHEIMER'S DISEASE

LUKE SAVILLE

Date of defense: 2023 04 27

Dr. Athanasios Zovoilis Thesis Supervisor	Associate Professor	M.D. Ph.D.
Dr. Majid Mohajerani Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Steve Wiseman Thesis Examination Committee Member	Associate professor	Ph.D.
Dr. Gerlinde Metz Internal External Examiner Department of Neurosciences Faculty of Arts and Science	Professor	Ph.D.
Dr. Michelle Scott External External Examiner Université de Sherbrooke	Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

Alzheimer's disease has a complex etiology and pathology. At the forefront of the disease is the amyloid beta neuropathology where amyloid plaques accumulate as the disease progresses, amidst a state of molecular dysfunction. A relatively understudied component of the molecular pathogenesis underlying Alzheimer's disease is the potential role of non-coding RNAs as well as their RNA modifications. Here, an integrative molecular and computational approach is used to study the role of non-coding RNAs in the APP NL-G-F Alzheimer's disease mouse model and human Alzheimer's disease patients, unveiling that non-coding RNAs transcribed from Short Interspersed Nuclear Elements (SINEs) in mice and humans are aberrantly processed in connection with alterations in gene expression. As well, adenosine to inosine RNA editing is shown to be dysregulated in SINE RNAs and non-coding RNAs involved in protein expression are differentially modified in the hippocampus of APP NL-G-F mice.

Preface – Author Contributions

The works collected in this thesis represent highly collaborate projects with incredible colleagues.

The work presented in chapter 2 was a collaboration between myself and my colleagues Dr. Yubo Cheng, Mr. Chris Isaac and Dr. Babita Gollen. Yubo performed the bioinformatics analysis, data curation and writing portions of the manuscript, Chris helped test the methodology, performed cell culture and molecular biology experiments, Babita performed much of the RNA sequencing and helped test the methodology, and I performed cell culture experiments, molecular biology experiments, helped with the methodology, wrote portions of the manuscript and assembled figures. Additionally, my colleagues Mr. Abel Belay performed some qPCR experiments, Dr. Jogender Mehla prepared the mouse tissue for analysis and Mr. Kush Patel performed a Western blot analysis. Dr. Nehal Thakor provided support, materials and reagents for our Western blot analysis, Dr. Majid Mohajerani provided support, our animal model, some data analysis, materials and reagents. Dr. Athanasios Zovoilis provided support, materials and reagents, guided the study with conceptualization and investigation, formal analysis, supervision, funding, data interpretation, validation, methodology, figure generation and manuscript writing. The data I directly contributed that is visualized in chapter 2 includes: Fig 2D, Fig2 supplement 2B, Fig 4B, Fig 5, Fig6 supplement, Fig 6, Fig 7, Fig 8, Fig 8 supplement1 and supplement 2. I also contributed data that is organized in supplementary table 1.

The work presented in Chapter 3 was a collaboration between myself and my colleague Dr. Yubo Cheng. Yubo performed the bioinformatics analysis, data curation and writing portions of the manuscript. I performed the tissue processing and RNA sequencing, the cell culture experiments,

qPCR data, data interpretation, writing of the manuscript and figure generation. Additionally, my colleague Dr. Babita Gollen helped test methodology. Our collaborator Ana Alvarez Veronesi prepared the Calgary Brain Bank tissue. Dr. Majid Mohajerani helped with data interpretation, Dr. Jeffrey T. Joseph provided pathology analysis, provision of material, study design, data interpretation, and Dr. Athanasios Zovoilis provided conception and design, establishment and testing of data generation and analysis pipelines, data interpretation, data visualization and writing of the manuscript, overall supervision.

The data I directly contributed that is visualized in chapter 3 includes: Fig 1, Fig 2, Fig 7, Fig 8, Fig S9, Fig S10.

The work presented in Chapter 4 was a collaboration between myself and my colleagues Mr. Liam Mitchell, Dr. Yubo Cheng and Dr. Babita Gollen. Liam and Yubo both contributed bioinformatic analysis, Babita contributed methodology development, I conducted the sequencing and contributed to writing the manuscript. As well, my colleagues, Mr. Travis Haight contributed testing of the next-generation-sequencing pipeline and Mr. Matthew Stuart-Edwards contributed testing of the base calling pipeline. Dr. Majid Mohajerani contributed data interpretation and Dr. Athanasios Zovoilis contributed conception and design, establishment and testing of data generation and analysis pipelines, bioinformatics analysis, data interpretation, data visualization and writing of the manuscript, and overall supervision.

The data I directly contributed that is visualized in chapter 4 is included in all of the figures in the manuscript.

Lastly, the work presented in Chapter 5 includes some work by colleague Mr. Liam Mitchell, particularly Figures 1 and 2 which help contextualize the study, as well as Liam's work developing the XPore analysis workflow which made producing supplementary table 3 and Fig 6 and 8 possible. Dr. Babita Gollen also performed the sequencing for the APP+ mice while I sequenced the wild type mice and Dr. Zovoilis contributed conception and design, establishment and testing of data generation and analysis pipelines, bioinformatics analysis, data interpretation, data visualization and overall supervision.

The data I directly contributed that is visualized in chapter 5 is included in Fig 3,4,5,6,7,8,9 and both supplementary tables.

Acknowledgements

I would like to start off by acknowledging my supervisor and mentor, Dr. Athanasios Zovoilis. Since starting in your lab as an undergraduate student, my life has changed a lot and I have matured a great deal as a person, in large part because of your guidance and encouragement. I immensely value our work together and I look forward to our continued collaborations and achievements. The works collected into this thesis would not have been possible without your vision or mentorship.

To my wife, Melanie. Thank you for your patience with the late nights in the lab and me always taking on way too many extracurriculars. I would not have been able to get through the trials of graduate school without your support. I love you and look forward to moving onto the next steps with you.

To my Dad, thank you for saving up money so I could go to school and for making sure I got good grades. Your support for me has been an immense privilege that I will cherish forever. You might have been on your own raising us for a while, and that was probably a tiring experience, but I think you knocked it out of the park. To Karen, thank you for coming into our lives and giving Dad a hand. Learning to cook with you was surprisingly fun and apparently was very applicable to my work in the wet lab. We didn't use pipettes to cook but learning different techniques and how to manipulate ingredients applies to experiments too.

To my brothers, Josh, Zach and Jason, thanks for being great friends on top of being great brothers. Josh, thanks for always looking like me and confirming that I always have the correct opinion. Zach, you always taught me to not take things too seriously and to always try and have some fun.

Jason, thanks for always feeding me beer, fancy coffee and smoked meat. You're a man who knows how to live.

To my colleagues in the lab: Chris, it was always a good time playing smash bros on my Nintendo Switch while we waited for gels to stain. It was also fun coming up with whacky project ideas with you. Babita, thank you for teaching me how to do my first sequencing libraries. You were very patient with me calling you randomly with questions about where an enzyme was. Travis, thank you for always making me feel smarter than I am by always asking me "biology questions." Going for drinks at the Zoo when we were done experiments for the day helped the pill go down easier when the experiments didn't work. Yubo, thank you for trying to teach me bioinformatics. I understand some of it now. You were really great to collaborate with and I hold our work together in very high regard. Liam, thank you for also trying to teach me bioinformatics, I think I understand some of it now but certainly less than what Yubo tried to teach me. I always enjoyed our weird philosophical discussions and ramblings. I'm still not going to watch your 2-hour YouTube video recommendations though. Riya, thanks for always laughing at my jokes. Especially when they weren't funny. I still have the pickled mango in my fridge, but I swear one day I will finish it. Matthew, you were always there for me when I needed a software downloaded or when I accidentally copied 100 gigabytes of files into my work folder that I didn't have read or write permissions to. Your technology support and your linux tricks and tips really made my life less miserable. Alexey, thanks for always answering my questions about life in your hometown. I look forward to working on the pig data with you in the future. Elly, thank you for analyzing all of the data I made into nice graph frames and talking to me about Elden Ring. One day, I promise I will get out of the castle.

To my committee, I really appreciate the feedback I've received from you over the years. Majid, thank you for always making sure I know a lot about the models we work with and Steve, thank you for making sure I never use the word "prove" in academic writing again. The advice you gave me throughout my academic programming has been invaluable and I cannot thank you enough.

I would also like to take a moment to acknowledge the folks on the graduate student's labour union. Rachel, Jackson, Jenn and John, thanks for letting me be a part of the bargaining team with you all. It was a lot of fun and was one of the most rewarding experiences of my graduate studies. I also really valued knowing I could go to the union for help if I had any issues.

To my dogs, Bondi and Chester (not that you will ever read this). You were not very helpful actually, but you made my life a lot more enjoyable, and I think that is pretty cool.

Table of Contents

Thesis Examination Committee	Error! Bookmark not defined.
Abstract	iii
Preface – Author Contributions	iv
Acknowledgements	vii
Table of Contents	Error! Bookmark not defined.
List of Figures	xiii
List of Tables	xvi
List of Abbreviations.....	xvii
Chapter 1 - Introduction	1
1.1 ncRNAs	2
1.1.2 ncRNAs fold into complex structures	4
1.1.3 ncRNAs are active contributors to disease when improperly regulated	6
1.2.1 Alu and B2 SINE RNAs.....	8
1.2.2 B2 and Alu SINE RNAs regulate RNA polymerase II mediated transcription	10
1.3 Alzheimer’s disease	11
1.3.1 Alzheimer’s disease etiology	11
1.3.2 Amyloid hypothesis of AD pathogenesis	13
1.3.3 APP NL-G-F mouse model.....	14
1.3.4 ncRNAs and AD	14
1.4 Technological innovations in RNA sequencing.....	15
1.5 Rationale Objectives and Thesis Overview	18
1.6 References.....	19
Chapter 2 - Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology.....	28
2.1 Foreword	29
2.2 Abstract.....	30
2.3 Introduction	30
2.4 Materials and Methods.....	34
2.4.1 Amyloid beta peptide preparations	34
2.4.2 Animals and behavioral measurement.....	34
2.4.3 Cell culture and transfections.....	35
2.4.4 Reverse transcription and quantitative PCR.....	36
2.4.5 RNA in vitro transcription and RNA-protein incubations	37
2.4.6 Short-RNA-seq and long-RNA-seq	40
2.4.7 Western blotting	41
2.4.8 Bioinformatics analysis.....	41
2.5 Results.....	44
2.5.1 B2 RNA regulated SRGs (B2-SRGs) are enriched in neural functions	44

2.5.3 A number of B2-SRGs get hyper-activated during the neurodegeneration phase of amyloid pathology	48
2.5.4 B2 RNA processing ratio increases during the neurodegeneration phase of in vivo amyloid pathology	52
2.5.5 Hsf1 accelerates B2 RNA processing	56
2.5.6 Hsf1 mediates increased B2 RNA processing in response to amyloid beta toxicity	62
2.6 Discussion	69
2.7 References	77
<i>Chapter 3 - Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes</i>	82
3.1 Foreword	83
3.2 Abstract	84
3.3 Introduction	84
3.4 Materials and Methods	88
3.4.1 Post-mortem hippocampal tissues	88
3.4.2 RNA in vitro transcription and RNA-protein incubations	89
3.4.3 Short RNA-seq	91
3.4.4 SH-SY5Y cell culture	92
3.4.5 LNA transfections	92
3.4.6 RT-qPCR	93
3.4.7 Bioinformatics analysis	93
3.4.8 Ethics approvals	98
3.5 Results	98
3.5.1 Short RNA sequencing identifies processing at the right arm of Alu RNAs in human hippocampus	98
3.5.2 Alu RNA processing is accelerated in hippocampi of AD patients	102
3.5.3 Standard RNA sequencing is able to detect the processing area in the right arm of Alu RNAs	105
3.5.4 Alu RNA processing is accelerated in the cortex of AD patients	109
3.5.5 Changes in Alu RNA processing ratio are associated with changes in P53 levels	114
3.5.6 Changes in Alu RNA expression and processing are associated with changes in gene expression	115
3.5.7 HSF1 can accelerate Alu RNA processing in vitro	122
3.3 Discussion	126
3.7 References	134
<i>Chapter 4 - NERD-seq: A novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs</i>	139
4.1 Foreword	140
4.2 Abstract	141
4.3 Introduction	141
4.4 Materials and methods	144
4.4.1 Post mortem hippocampal tissues	144
4.4.2 Direct RNA sequencing using Nanopore	144
4.4.3 Bioinformatics analysis	146
4.4.4 Data access	147
4.5 Results	147
4.5.1 The modified library construction protocol used in NERD-seq bypasses the limitations posed by the standard Nanopore direct RNA-seq approach	147

4.5.2 NERD-seq enables the generation of reads with higher coverage for the non-coding genome, while still detecting mRNAs and poly(A) ncRNAs	151
4.5.3 NERD-seq can efficiently detect various classes of short non-poly(A) ncRNAs in contrast to the standard approach	154
4.5.4 NERD-seq allows for the detection of LINE 1 produced ncRNAs	157
4.6 Discussion.....	160
4.7 References.....	163
<i>Chapter 5 - ncRNA modifications: an unmined source of information in amyloid beta neuropathology.....</i>	<i>171</i>
5.1 Foreword	172
5.2 Abstract.....	172
5.3 Introduction	173
5.4 Material and methods	175
5.4.1 NERD-seq of hippocampus tissue.....	175
5.4.2 Nanopore sequencing of Human cerebral cortex tissue	176
5.4.3 7SK RNA sequencing	177
5.4.4 Cell culture	177
5.4.5 Endonuclease V assay.....	178
5.4.6 qPCR	178
5.4.7 Bioinformatics analysis and SINE A-to-I editing method rational.....	178
5.4.8 Differential expression analysis	179
5.4.9 Gene expression analysis of 5S rRNA	179
5.4.10 Differential modification analysis	179
5.5 Results.....	180
5.5.1 B2 RNAs' A-to-I editing stoichiometry is reduced in APP NL-G-F mice.....	180
5.5.2 Interferon stimulated gene expression is elevated in 12 month APP NL-G-F mice.....	185
5.5.3 NERD-Seq can enable RNA modification detection.....	190
5.5.4 APP+ mice have differentially modified RNAs	192
5.6 Discussion.....	194
5.7 References.....	197
<i>Chapter 6 - Conclusion</i>	<i>203</i>
6.1 Conclusions and future directions	204
6.2 References.....	209
<i>Appendix 1 – Supplementary for chapter 2: Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology.....</i>	<i>212</i>
<i>Appendix 2 – Supplementary for chapter 3: Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes</i>	<i>225</i>
<i>Appendix 3 – Supplementary for chapter 4: NERD-seq: A novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs.....</i>	<i>237</i>

List of Figures

Chapter 1

Figure 1. RNA structure entries into the NDB database – Page 4

Figure 2. Life cycle of Alu RNA – Page 8

Chapter 2

Figure 1. B2-SRGs are enriched in neural functions – Page 45

Figure 2. A number of B2-SRGs are hyper-activated in amyloid pathology – Page 49

Figure 3. B2 RNA processing ratio is increased in 6-month-old APP mice – Page 54

Figure 4. Hsf1 is upregulated in 6-month-old APP mice – Page 58

Figure 5. Hsf1 accelerates B2 RNA processing – Page 60

Figure 6. A hippocampal cell culture assay for tracking effects of amyloid beta toxicity on B2 RNA stability – Page 63

Figure 7. B2 RNA destabilization leads to increase in expression of B2-SRGs – Page 66

Figure 8. Hsf1 mediates B2 RNA processing in amyloid toxicity – Page 68

Figure 9. Representation of the role of B2 RNA processing in amyloid pathology – Page 71

Chapter 3

Figure 1. Processing areas of Alu RNAs in the human hippocampus as revealed by short RNA-seq – Page 99

Figure 2. Alu RNA processing ratio is increased in the hippocampi of AD patients in CBB cohort – Page 103

Figure 3. Processing areas of Alu RNAs in the human cortex as revealed by RNA-seq – Page 106

Figure 4. Alu RNA processing ratio is increased in the cortex of AD patients in the MAP cohort – 110

Figure 5. Alu RNA processing ratio is associated with clinical, pathology, genetic, and molecular markers of AD in MAP patients – Page 113

Figure 6. Alu RNA expression and processing levels are associated with transcriptome changes in the cortex of MAP patients – Page 116

Figure 7. Alu RNA destabilization leads to increase in expression of Alu RNA processing correlated genes – Page 120

Figure 8. HSF1 accelerates Alu RNA processing in vitro – Page 124

Figure 9. Representation of the observed changes in Alu RNA processing in AD – Page 128

Chapter 4

Figure 1. Description of the NERD-seq approach – Page 148

Figure 2. NERD-seq enables the generation of reads with higher coverage for the non-coding genome – Page 150

Figure 3. NERD-seq can detect mRNAs and poly(A) ncRNAs – Page 153

Figure 4. NERD-seq can detect snoRNAs, snRNAs, scRNAs and srpRNAs – Page 156

Figure 5. NERD-seq can detect tRNAs and ribosomal RNA fragments (rRFs) – Page 156

Figure 6. NERD-seq detects LINE L1 associated ncRNAs – Page 159

Chapter 5

Figure 1. Workflow describing the method for estimating relative A-to-I editing rate across SINE B2 RNAs – Page 181

Figure 2. A-to-I editing is increased at specific B2 RNA positions and time points in a mouse model of amyloid beta pathology – Page 182

Figure 3. When A-to-I editing is reduced, full length B2 RNA levels are reduced – Page 184

Figure 4. Interferon stimulated genes emerge in 12mo APP+ mice – Page 187

Figure 5. Adar and Adarb1 expression is elevated before interferon stimulated genes – Page 189

Figure 6. NERD-seq detects RNA modifications in Rn7SK – Page 191

Figure 7. APP+ modified RNAs are primarily involved in translation – Page 192

Figure 8. BC1 is dysregulated in APP+ mice – Page 193

Figure 9. 5S rRNA levels are reduced in APP+ mice Page 194

Chapter 6

Figure 1. Model of ncRNA derived dysregulation in amyloid beta neuropathology – Page 206

Appendix 1 – Supplementary for chapter 2:

Figure 2—figure supplement 1. Expression of known hippocampal markers in our RNA-seq data of mice hippocampi – Page 212

Figure 2—figure supplement 2. Validation of RNA-seq data in WT and APP mice by RT-qPCR – Page 213

Figure 2—figure supplement 3. Expression levels of all B2-SRGs in amyloid pathology – Page 214

Figure 3—figure supplement 1 – Page 216

Figure 3—figure supplement 2 – Page 217

Figure 4—figure supplement 1. Expression of Hsf1 in neural tissues – Page 218

Figure 5—figure supplement 1. Position of B2 RNA fragments generated in vitro and downregulation of Hsf1 protein levels – Page 219

Figure 6—figure supplement 1 – Page 220

Figure 8—figure supplement 1. B2 RNA levels in HT22 cells and relationship with Hsf1 levels – Page 221

Figure 8—figure supplement 2. Expression levels of all B2-SRGs in amyloid beta toxicity – Page 223

Appendix 2 – Supplementary for chapter 3

Appendix Figure S1 – Page 225

Appendix Figure S2 – Page 226

Appendix Figure S3 – Page 227

Appendix Figure S4 – Page 228

Appendix Figure S5. H3K9ac ChIP-seq for MAP patients – Page 229

Appendix Figure S6. Pathways impaired in AD up-regulated genes that are strongly correlated with Alu RNA processing ratio – Page 230

Appendix Figure S7. Additional pathways impaired in AD up-regulated genes that are strongly correlated with Alu RNA processing ratio – Page 232

Appendix Figure S8. Expression of known hippocampal markers and cellular deconvolution in the MAP RNA-seq data – Page 233

Appendix Figure S9 – Page 234

Appendix Figure S10 – Page 235

Appendix 3 – Supplementary for chapter 4

Supplementary Figure 1 – Page 237

Supplementary Figure 2 – Page 238

Supplementary Figure 3 – Page 238

List of Tables

The following tables are submitted as external tables due to their size:

supplementary_table1.xlsx referred to as “supplementary table 1.”

table_EV1.xlsx referred to as “table EV1.”

Dataset_EV1.xlsx referred to as “dataset EV1.”

Dataset_EV2.xlsx referred to as “dataset EV2.”

Dataset_EV3.xlsx referred to as “dataset EV3.”

Dataset_EV4.xlsx referred to as “dataset EV4.”

Dataset_EV5.xlsx referred to as “dataset EV5.”

Dataset_EV6.xlsx referred to as “dataset EV6.”

Dataset_EV7.xlsx referred to as “dataset EV7.”

Dataset_EV8.xlsx referred to as “dataset EV8.”

supplementary_table2.xlsx referred to as “supplementary table 2.”

supplementary_table3.csv referred to as “supplementary table 3.”

List of Abbreviations

42	Amyloid Peptide (1-42 amino acid)
8-Aza	8-aza adenosine
A-to-I	Adenosine to inosine
AD	Alzheimer's Disease
Adar	Adenosine Deaminase Acting on RNA
Adarb1	Adenosine Deaminase RNA Specific B1
Anti-adar	Adar and Adarb1 siRNA Treated
APP mice/ APP+ mice	APP NL-G-F mouse model
B2-SRGs	B2 SINE RNA regulated stress response genes
CBB	Calgary Brain Bank
cDNA	Complementary DNA
Cm	Cytosine methylation
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dsRNA	double-stranded RNA
E4	APOE e4 allele
EOFAD	Early Onset Familial Alzheimer's Disease
FCS	Fetal Calf Serum
GB	Gigabases
GO terms	Gene Ontology terms
Hsf1	Heat Shock Factor 1
IFN	Interferon
IHC	Immunohistochemistry
ISG	Interferon Stimulated Gene
Kmer	Nucleotide string of "mer" length (generally 5nt long)
KS test	Kolmogorov-Smirnov test
LINE	Long Interspersed Nuclear Element
LNA	Locked Nucleic Acid
lncRNA	Long non-coding RNA
LOAD	Late Onset Alzheimer's Disease
M	Million reads
m1G	N1-methylguanosine
m2,2,7G	N2,2,7-trimethylguanosine
m6a	N6-methyladenosine
m7G	N7-methylguanosine
MAP	Memory and Aging Project
MCI	Mild Cognitive Impairment
miRNA	MicroRNA
mo	Months Old
mRNA	Messenger RNA
NCI	No Cognitive Impairment
ncRNA	Non-coding RNA
nt	Nucleotide

ONT	Oxford Nanopore Technologies
PCR	Polymerase Chain Reaction
piRNA	PIWI interacting RNA
poly(A)/ poly-A	Polyadenylated
poly(T)	Polythymidine
R	Reverse peptide (42-1 amino acid)
RNA	Ribonucleic acid
RNA Pol I	RNA Polymerase I
RNA Pol II	RNA Polymerase II
RNA Pol III	RNA Polymerase III
RNA-seq	RNA sequencing
ROS	Reactive Oxygen Species
rRNA	Ribosomal RNA
RT-qPCR	Reverse Transcription quantitative PCR
scRNA	Small Cytoplasmic RNA
SINE	Short Interspersed Nuclear Element
siRNA	Small Interfering RNA
snoRNA	Small Nucleolar RNA
snRNA	Small Nuclear RNA
srpRNA	Signal Recognition Particle RNA
tRFs	tRNA Fragments
tRNA	Transfer RNA
TSS	Transcription Start Site
Um	Uracil methylation
UTR	Untranslated Region
WT	Wild type mice (C57BL/6J)

Chapter 1 - Introduction

1.1 ncRNAs

1.1.1 A short history of ncRNAs

Non-coding RNAs (ncRNAs) are a diverse biomolecule class with functions distinct from messenger RNAs' (mRNAs) template role in protein synthesis. The first ncRNA described was the yeast alanine tRNA (Ala-tRNA) in 1965 (Holley *et al*, 1965; Hube & Francastel, 2018) demonstrating that RNAs are a functional biomolecule with structure and unique functions, not just as templates for translation. Before the discovery of Ala-tRNAs and mRNAs, the prevailing theory of RNA as a biomolecule class was that ribosomes, which are particles comprising RNA and protein subunits, contained intermediate information from the genome, resulting in the one gene-one ribosome-one protein theory (Brenner *et al*, 1961). It was not until the discovery of mRNAs in 1961 that RNAs as a biomolecule class could be divided into “coding” and “non-coding” classes, whereby mRNAs were synthesized as intermediate message carriers to ribosomes, for the information directed synthesis of proteins (Brenner *et al.*, 1961; Cobb, 2015; Gros *et al*, 1961). The model postulated by Brenner *et al*, 1961 and Gros *et al*, 1961 soon became the central dogma of molecular biology whereby, DNA encodes RNA, and RNA encodes proteins. Thus began a decades long endeavor: to uncover the diversity and importance of ncRNAs as essential biomolecules in all cells.

Since their initial discovery, ncRNAs are now thought to be the predominant products of the transcription process whereby, some RNAs like rRNAs (ribosomal RNAs, which form the core of the ribosome) are so highly represented in the transcriptome (approximately 80% of all RNAs in normal cells (O'Neil *et al*, 2013)), eukaryotic cells dedicate the entirety of RNA polymerase I (RNA pol I) catalyzed transcription to synthesizing rRNAs (except 5S rRNA: a rRNA which does

not form part of the ribosome core and is transcribed by RNA polymerase III (RNA pol III)) (Turowski & Boguta, 2021). Of course, rRNAs are critical for cell function and arguably, eukaryotic cells dedicate the majority of an entire organelle: the nucleolus to producing ribosomes by chemically modifying and processing rRNAs into catalytic structures with the aid of snoRNAs (small nucleolar RNAs) (Melese & Xue, 1995; Shishova *et al*, 2015). Beyond, tRNAs, rRNAs and snoRNAs, snRNAs (small nuclear RNAs) have diverse functions including participation in pre-mRNA splicing and regulation of transcription processes (Briese & Sendtner, 2021; Goh *et al*, 2020), SINE RNAs (short interspersed nuclear elements) inhibit RNA polymerase II (RNA pol-II) catalyzed transcription (Espinoza *et al*, 2004; Ponicsan *et al*, 2013; Ponicsan *et al*, 2015), selectively enhance translation (Ivanova *et al*, 2015), and promote genome instability (Ade *et al*, 2013), and miRNAs (micro RNAs) post-transcriptionally inhibit mRNA transcript maturation by recruiting the RISC complex (Bartel, 2018; Gregory *et al*, 2005) – to name a few.

Indeed, ncRNAs are essential components to cell function, dysregulation of which can lead to disease in complex eukaryotes – such as mammals. For instance, decreased adenosine to inosine (A-to-I) modifications in the wobble position of tRNAs due to autosomal mutations in ADAT3 (V144M) may cause intellectual disability (Ramos *et al*, 2019). Additionally, miRNAs are often dysregulated in cancer (Iorio *et al*, 2005; Reddy, 2015) and are often exploited as predictive markers for different kinds of cancers such as miR-15a, miR-103, miR-148a, miR-320a, miR-451 and miR-596 in the prediction of colon cancer recurrence (McGuire *et al*, 2015; Shivapurkar *et al*, 2014). Since many ncRNAs hold important housekeeping roles in all cells, it is of little surprise that perturbations in their regulatory pathways can be highly damaging.

1.1.2 ncRNAs fold into complex structures

It is often stated, ‘structure determines function,’ and while it is usually attributed to proteins, whose structure is determined by its amino acid content and its function determined by its structure, the same is often true for RNAs and particularly, ncRNAs, by their nucleotide content (Miladi *et al*, 2020). Additionally, because of the maturity of bioinformatics and interrogative techniques to study protein structure, the field has developed insofar as being capable of computationally predicting many structures, simply from the amino acid content (Jumper *et al*, 2021). The same cannot be said for RNAs, despite ribonucleotides being much less diverse than amino acids. However, RNA structural biology has massively expanded as a field and efforts are being made to catalogue and computationally predict RNA structures. RNA 3D structure cataloguing has thus massively expanded in the past decade (Figure 1) whereby, as of March 1st, 2023, 12 666 3D RNA structures have been deposited in the NDB database (<http://ndbserver.rutgers.edu/> accessed on 2023/03/05).

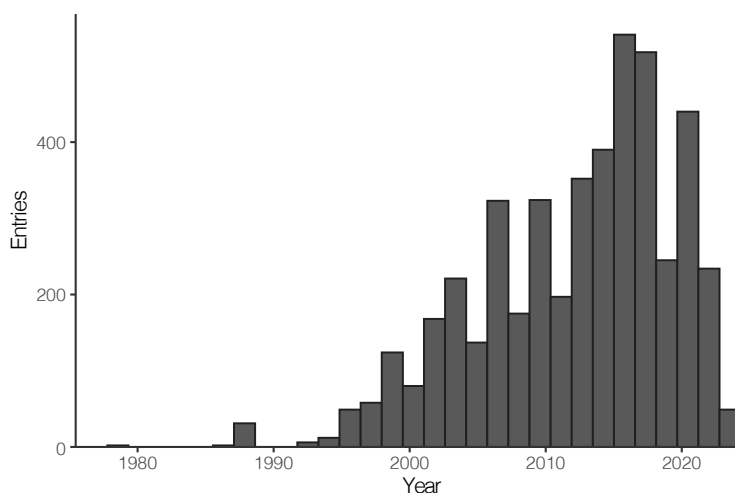


Figure 1. RNA structure entries into the NDB database.

Histogram depicting the year over year RNA 3D structure entries into the NDB database (<http://ndbserver.rutgers.edu/> accessed on 2023/03/05).

While the protein database (PDB) hosts 201 979 3D protein structures and 1,068,577 computationally predicted structures (<https://www.rcsb.org/> accessed on 2023/03/05), the improved focus on 3D RNA structures has developed the overall understanding of how ncRNAs and even mRNAs function. In fact, the structural composition of mRNAs may be even more important than previously thought. Leppek *et al* (2022) for instance, showed that by specifically engineering mRNA structures while retaining codon redundancy, they could optimize the rate of translation within an interplay of ribosome load and mRNA half-life (Leppek *et al*, 2022). Conversely, ncRNAs have diverse functions which are demonstrably impacted by their structures. Some ribozymes for example require magnesium ions (Mg^{2+}) to catalyze their reaction (Beebe *et al*, 1996). The positive ion ligands interact with the negatively charged ribonucleotide backbone, stabilizing its structure, enhancing the catalytic function (Yamagami *et al*, 2021).

RNA structures usually fold into single and/ or double stranded formations with some inter hairpin interactions and quadruplex formations. Depending on their localization, this has different connotations. dsRNAs in the cytoplasm for instance can be harmful to eukaryotic cells unless they are modified with A-to-I edits. In fact, the way A-to-I editing was discovered was by an experiment where antisense RNAs were transfected into *Xenopus* oocytes and embryos but instead of gene silencing as the authors expected, the dsRNA formation was transient (Bass & Weintraub, 1987; Rebagliati & Melton, 1987). It was then determined many adenosines were covalently converted into inosines (Bass & Weintraub, 1988). Later, A-to-I editing by adenosine deamination acting on RNA (Adar) was discovered to inhibit MDA5-MAVS signaling which activates the type I interferon response and interferon stimulated genes (Pestal *et al*, 2015). Interestingly, complete Adar knockouts in mice are embryonic lethal (Pestal *et al*, 2015), unless the Mavs or Mda5 gene

is also knocked out (Bajad *et al*, 2020). Additionally, the Adar enzyme is interferon stimulated to isoform switch into its p150 form (Sun *et al*, 2021). Adar p150 hyperedits dsRNAs in the cytoplasm to produce bulge structure motifs and minimize their dsRNA motifs (Vogel *et al*, 2020). If dsRNA accumulates, the MDA5-MAVS pathway is triggered and the cell will commit to an apoptotic state (Huang *et al*, 2022).

1.1.3 ncRNAs are active contributors to disease when improperly regulated

Since ncRNAs are often involved in housekeeping roles and can participate in cell response to stress pathways, dysregulation of their sequence and structure, their signaling partners, and their ligands can challenge cells. Furthermore, many ncRNAs are transcribed from LINE and SINE elements which can act as semi-autonomous and autonomous retrotransposons (Kramerov & Vassetzky, 2011; Ponicsan *et al*, 2010). If a retrotransposon insertion impacts an essential gene, it can cause disease. As such, mammalian germ cells highly express a repetitive class of RNAs called piRNAs (PIWI interacting RNAs). piRNAs, amongst other functions, inhibit retrotransposons' insertions into genomic space by complementary base pairing and inhibition of their interactions with genomes. piRNAs reside in the genome as pattern recognition sequences and if they encounter a novel transposon, can themselves be inserted into genomic space as regulators of transposon insertion in future generations (Ozata *et al*, 2019). Doing so, they limit the genome instability posed by transposons produced in germ cells; whereas, insertions in somatic cells are less deleterious and harmful to an organism, germ cell insertions can infect an entire offspring. Hancks and Kazazian (2012) reviewed some of the disorders thought to be caused by transposons such as the human SINE Alu RNA and L1 LINE RNA. Adrenoleukodystrophy, hemophilia, and glycerol kinase deficiency can have Alu derived etiologies while duchennes muscular dystrophy, colon

cancer and choroideremia can have L1 derived etiologies (Hancks & Kazazian, 2012) – to name a few.

ncRNAs also play diverse and important roles in neurological disorders. tRNAs for instance, can be processed into shorter RNAs (tRFs) by the Ang protein, for which about 40 mutations in the ANG gene have been associated with neurological disorders, like stroke, Alzheimer’s disease, epilepsy and Parkinson’s (Mathew *et al*, 2023). In the case of stroke, tRFs have been shown to feature differential expression in post-stroke recovery (Winek *et al*, 2020).

Mounting evidence suggests an important role of miRNAs in Parkinson’s disease. LRRK for example, is a very important gene in Parkinson’s disease and has been implicated in the negative regulation of let-t and mir-148, allowing the overexpression of E2F1 and DP, which limits cell division and promotes cell death (Gehrke *et al*, 2010; Salta & De Strooper, 2012). As well, the lncRNA Neat1 has been shown in multiple reports to sequester by a “sponge” mechanism ncRNAs mir124-3p and mir-374-5p, upregulating PDE4B, an active contributor to Parkinson’s disease (Chen *et al*, 2021; Dong *et al*, 2021; Kuo *et al*, 2021).

Overall, ncRNAs are massively influential to cell function and are experiencing growing attention and interest. Expanding the study of ncRNAs to relatively understudied species will further uncover important contributors to cell function and disorder. SINE RNAs, such as B2 and Alu RNAs in mouse and human cells, have received some attention for their roles in promoting genome instability but as of recently, are also receiving increasing interest in their roles regulating their respective RNA Pol-II derived transcription.

1.2 SINE RNAs

1.2.1 Alu and B2 SINE RNAs

SINE (short interspersed nuclear elements) repeats are genomic motifs that appear ubiquitously throughout eukaryotic genomes. Through evolutionary time, they have inserted themselves into their respective genomes as transposable elements, correlating insertion events with increasing genomic size and complexity (Symmons *et al*, 2014). Additionally, SINE elements often insert themselves between exons, increasing exonization and thus the complexity of genetic isoforms (Chen & Yang, 2017). SINE RNAs themselves have complex and multi-functional roles in the cell, contributing to translational and transcriptional control during stress (Figure 1).

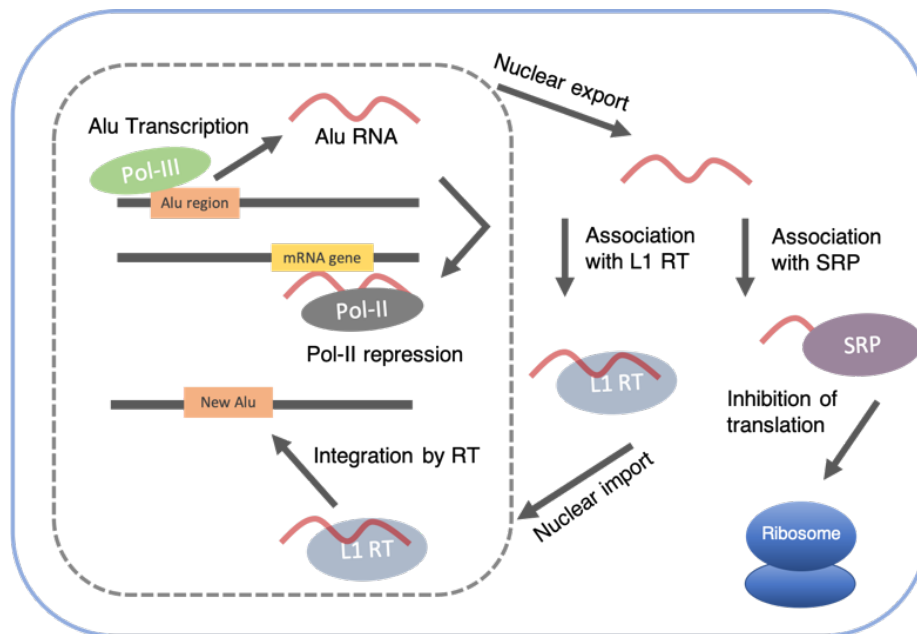


Fig 2. Life cycle of Alu RNA

Alu RNA is a SINE ribozyme retrotransposon with multifunctional roles in the cell. Alu RNAs can inhibit mRNA transcription, reproduce in the genome with new insertions, and selectively inhibit translation.

SINE elements are actively transcribed by RNA Polymerase III (RNA Pol-III) (Agarwal *et al*, 2016). SINE B2 and Alu RNAs are the most transcribed SINEs in the murine and primate genomes respectively and are capable of inhibiting site-specific, RNA polymerase II derived transcription by directly binding to Pol-II (Yakovchuk *et al*, 2009). Additionally, inverted SINE transcripts have been shown to post-transcriptionally repress pre-mRNA maturation (LeGendre *et al*, 2013). Multiple works have shown that B2 and Alu SINE RNAs control transcription in stress related gene regions, preventing the stress response during cellular homeostasis (Allen *et al*, 2004; Espinoza *et al*, 2004). This transcriptional control is regulated by transcription factors like the EZH2 protein by destabilizing the RNA (Zovoilis *et al*, 2016).

The human Alu element is evolutionarily derived from the snRNA gene, 7SL (Deininger, 2011; Deininger *et al*, 2003), while the murine B2 element is evolutionarily derived from tRNA genes (Deininger *et al*, 2003). Interesting enough, the murine genome also has a 7SL derived SINE called B1 (Deininger *et al*, 2003) and while also transcribed, B2 and Alu are apparently more similar on a functional level and are often thought of as orthologous. Obvious differences exist, such as their lengths, whereby B2 RNA is 200nt long and Alu RNA is 300nt long. Structurally, they are quite dissimilar. Alu RNAs have two, primarily double-stranded heterodimeric arms linked by an internal poly(A) segment, while B2 RNA is more monomeric with multiple hairpins, linked by a dynamic single-stranded ring (Dunker *et al*, 2017). Both RNAs have a poly(A) 3' end which is used by LINE RNAs to mediate their insertions in new genomic sites (Deininger, 2011; Dunker *et al*, 2017).

1.2.2 B2 and Alu SINE RNAs regulate RNA polymerase II mediated transcription

An important characteristic of B2 and Alu RNA is their affinity for RNA polymerase II whereby, they can selectively inhibit the transcription of many RNA Pol-II transcribed genes (Allen *et al.*, 2004; Espinoza *et al.*, 2004; Mariner *et al.*, 2008). In particular, B2 and Alu RNAs bind to the DNA cleft of RNA Pol-II and inhibit the phosphorylation of serine 5 in the carboxy terminal domain. Doing so, the general transcription factor, TFIID is prevented from allowing the dislocation of RNA Pol-II from the promoter (Ponicsan *et al.*, 2013; Yakovchuk *et al.*, 2009). Inhibition of RNA Pol-II exists within an interplay of stress and pre-stress conditions whereby, in pre-stress B2 and Alu RNAs inhibit the transcription of so called stress response genes, while during stress conditions they inhibit housekeeping genes such as Actin b and Hexokinase II (Allen *et al.*, 2004; Hernandez *et al.*, 2020; Zovoilis *et al.*, 2016). As many as 1684 genes may be controlled by B2 RNA, according to a CHART-seq study performed previously (Zovoilis *et al.*, 2016). Of course, B2 elements are much less numerous than Alu elements (350 000 B2 elements in the murine genome compared to approximately 3 million Alu elements (Ponicsan *et al.*, 2010)). Further study is needed to ascertain how many genes are controlled by Alu RNAs.

An interesting development in the story of B2 and Alu RNA's control over RNA Pol-II transcription emerged when it was discovered the polycomb repressor protein: Ezh2 binds directly to B2 RNA (Zovoilis *et al.*, 2016). Furthermore, interactions *in vitro* showed that Ezh2 had a direct and specific effect on B2 RNAs, rapidly processing them into fragments. In a heat shock model, a higher number of B2 RNA fragments were produced and the CHART-seq peaks corresponding to sites of B2 RNA binding, correlated positively with the differential transcript analysis pre vs post-heatshock. This all indicated a model by which B2 RNA pre-heatshock, binds to RNA Pol II,

inhibiting target mRNA transcription. Heat shock occurs and the Ezh2 protein binds to B2 RNA, processing it into fragments. Thus, RNA Pol-II can proceed with transcription and mount a stress response (Zovoilis *et al.*, 2016).

A follow up study to Zovoilis *et al.*, 2016 elucidated the mechanism by which B2 RNA process and replicated the findings in a human model with Alu RNA (Hernandez *et al.*, 2020). In both cases it was confirmed, Ezh2 processes B2 and Alu RNA. Furthermore, the authors postulated that the mechanism by which B2 and Alu RNAs were being processed is not because Ezh2 is a nuclease, but that B2 and Alu RNAs have a ribozyme activity which is accelerated by contact with Ezh2 (Hernandez *et al.*, 2020).

1.3 Alzheimer's disease

1.3.1 Alzheimer's disease etiology

Alzheimer's disease (AD) has a highly complex probabilistic etiology (Barnard *et al.*, 2014; Huang *et al.*, 2019). From genetic to environmental causes, it is rarely simple to determine the exact causes of any single diagnosis. Initially characterized by its telltale amyloid beta plaque formation and tau neurofibrillary entanglement, AD is the most commonly diagnosed dementia, yet no treatment to date can successfully slow its progression (Cornutiu, 2015). The genetic causes are complex. GWAS studies have elucidated only a few reasonably predictive gene types, but largely for familial AD which usually occurs in early onset. Mutations in APP, PSEN1 and PSEN2 genes are thought to have high disease penetration in early onset familial AD (EOFAD) (Bertram *et al.*, 2010). In the case of late onset AD (LOAD), the ApoE allele types for example are considered the most predictive, whereby the ApoE4 allele can increase risk of LOAD by approximately 4 fold (Bertram

et al., 2010; Strittmatter *et al.*, 1993). Beyond genetic associations, many environmental and physiological theories have emerged.

Accumulation of reactive oxygen species (ROS) for example, is another feature of AD which has received a large amount of attention (Llanos-Gonzalez *et al.*, 2019; Moulton *et al.*, 2021). ROS is a normal part of cell function, tightly controlled by multiple systems, including antioxidant production and the Nrf2-Keap1 gene pathway (Schieber & Chandel, 2014). Once ROS accumulates to the extent by which it permanently oxidizes important biomolecules, thus inhibiting their function, cells begin to experience ROS stress (Schieber & Chandel, 2014). Because mitochondrial dysfunction and apoptosis are more common in AD afflicted neurons, commensurate to low antioxidant levels, it has been suggested oxidative stress is responsible for much of the neurodegeneration that occurs during AD (Manoharan *et al.*, 2016; Xie *et al.*, 2002) particularly because of its role in triggering neuroinflammation. Neuroinflammation results from increasing immune activity in brain regions, particularly from microglial activation. The current prevailing theory is that amyloid plaques and tau neurofibrillary tangles trigger microglial activation by triggering synaptic lesions and accumulating ROS, causing a release of cytokines such as interleukins (IL) IL-1 β , IL-8 and tumor necrosis factor (TNF) (Leng & Edison, 2021). The potent inflammatory signals may cause an increased apoptotic state by recruiting TNFR1 and Casp8 (Leng & Edison, 2021; Mischeau & Tschopp, 2003).

AD etiology is however, highly complex and is likely heterogenous on a case per case basis. Elucidating the molecular mechanisms that result in cell perturbation is a key focus of emerging research and no etiological factor is more well studied than the accumulation of amyloid plaques.

1.3.2 Amyloid hypothesis of AD pathogenesis

The amyloid pathology of Alzheimer's disease (AD) is among the most studied hallmarks of the disease (Gotz *et al*, 2004; He *et al*, 2013; Seino *et al*, 2010), yet the causes of amyloid aggregation and its resultant contributions to neurodegeneration remain poorly understood (Grimm *et al*, 2016; Rasgon & Kenna, 2005; Tiiman *et al*, 2013). Given the rising age of global populations and thus the incidence of AD, increasing focus is being applied to treat and prevent AD (Hane *et al*, 2017). Alas, treatment and prevention strategies are typically focused on reducing amyloid plaque formation, yet such strategies remain unsuccessful (Cummings *et al*, 2020; Maia & Sousa, 2019).

Amyloid is produced as a processed product of the APP protein where it can exist in a monomeric, oligomeric and fibrillary form. It has been suggested that oligomeric amyloid is the most cytotoxic and as the oligomeric order increases, the cytotoxicity decreases, until it nucleates into a fibrillary form (Ono *et al*, 2009). In fact, some studies have suggested the monomeric form is neuroprotective and that maximizing the monomeric to oligomeric ratio could be a treatment strategy (Giuffrida *et al*, 2009).

An important consideration to keep in mind when studying AD through the lens of the amyloid hypothesis is that while amyloid seems to be an important contributor to the pathophysiology of AD, it may not be the convenient explanation it was initially thought to be. Multiple clinical trials have assessed drugs targeting amyloid accumulation but no successful treatment has been produced (Asher & Priefer, 2022; Huang *et al.*, 2019).

1.3.3 APP NL-G-F mouse model

To enable the study of AD, mouse models expressing some of the important molecular pathologies of AD have been developed. Most commonly, mice are genetically altered to produce humanized forms of genes like tau and APP. As well, mutations mirroring genes with predictive power for AD may also be added. The APP NL-G-F model (APP⁺) for instance is an APP knock-in model in mice which contains the Swedish, Iberian and Arctic mutations to stabilize, extend the half-life of and promote the productions of amyloid plaques (Nilsson *et al*, 2014; Saito *et al*, 2014).

The APP⁺ model mimics many of the hallmark pathophysiological features of AD, developing plaques, followed by microgliosis, synaptic loss and cognitive impairment (Saito *et al*, 2014). An important consideration is that while tau is increasingly phosphorylated, tangles are not produced, and neuronal loss does not occur like it does in AD patients (Saito *et al*, 2014). Despite this, the model is useful means to study the amyloid hypothesis on its own and particularly, how it induces neuroinflammation and ROS accumulation (Izumi *et al*, 2020; Uruno *et al*, 2020).

1.3.4 ncRNAs and AD

ncRNAs hold many important roles in cells and in varying disorders. ncRNAs in AD is a relatively new field of study and is gaining traction. BC200 and BC1, which are human and mouse orthologues (respectively), are thought to aide the activity of a protein called FMRP, which selectively inhibits translation for proteins such as those encoded by the Arc/Arg3.1 and MAP1B, and the kinase α -CaMKII genes at synapses (Zalfa *et al*, 2003). Depending on the modification state of BC1 however, the localization of the ribonucleoprotein complex can vary. If the 5'UTR of BC1 is 2'-O-methylated, the complex will inhibit translation outside of synapses, otherwise both

FMRP and BC1 will localize to the synapse and inhibit mRNA transcription in the region (Lacoux *et al*, 2012). In Tg2576-APP^{swe} mice, it was reported that BC1 expression increased after 6 months, into old age mice and correlated with APP translation. In turn, amyloid deposition increased. When BC1 was inhibited using a locked nucleic acid (LNA) antisense oligonucleotide (ASO), amyloid deposition decreased (Zhang *et al*, 2018). Furthermore, it has been suggested ROS accumulation can trigger ROS stress by oxidizing RNA molecules, often producing an RNA modification called 8-oxo-guanine (Song *et al*, 2011). Oxidation of RNA can affect their functionality. For example, when rRNA species are more readily oxidized, the rate of protein-synthesis is reduced (Willi *et al*, 2018). rRNA dysfunction is a common feature of AD and it is possible that ROS may play a role in that dysfunction – in addition to lowered levels of rRNA expression overall (Ding *et al*, 2005; Nunomura & Perry, 2020).

1.4 Technological innovations in RNA sequencing

The study of RNA and gene expression has been massively improved by techniques like RNA sequencing (RNA-seq). Compared to techniques like microarray gene expression and RT-qPCR, RNA-seq increases throughput and genetic context (Rao *et al*, 2018). And, since RNA-seq's inception many different techniques have emerged from it. Gene expression analysis is largely conducted from the mRNA-seq technique whereby RNA is captured by hybridizing probes to the polyadenylation signature of mRNAs and other polyadenylated transcripts (Zhao *et al*, 2014). However, other techniques are employed to sequence total RNA by rRNA depleting RNA samples and hybridizing known adaptor sequences on the ends of all RNAs (Zhao *et al*, 2014; Cheng *et al*, 2020). As well, a non-fragmenting approach to study small RNAs (sRNAs) has been developed which enables the enrichment of small RNAs such as miRNAs and snRNAs. sRNA sequencing

relies on size fractionation and a random adaptor ligation chemistry for sequencing (Cheng *et al*, 2020).

All standard RNA-seq approaches rely on similar preparative approaches for sequencing. Because Illumina sequencing technologies – the foremost used DNA sequencing technology in molecular biology – requires a PCR reaction during sequencing to reincorporate fluorescent nucleotides in DNA molecules, the sequencing of RNA polynucleotides requires the prior conversion of RNA to cDNA via the reverse transcription reaction (Ozsolak and Milos, 2011). While reverse transcription is generally unavoidable with the Illumina based RNA-seq technique, it can often introduce bias, whereby due to certain structures and sequence complexities, some reverse transcriptase enzymes such as the Moloney Murine Leukemia Virus (M-MuLV) enzyme, cannot efficiently reverse transcribe some RNA classes like snoRNAs (Minshall and Git, 2020). As such, methodological variations have developed to specifically improve the sequencing of broad RNA classes. TGIRT-seq for instance uses the TGIRT enzyme which has a strand switching capability, coupling 3' end adaptors without the use of ligation factors (Boiving *et al*, 2018; Boivin *et al*, 2020). TGIRT-seq demonstrably improves structure resolution and enables less biased reconstruction of the transcriptome.

Beyond standard sequencing approaches, other sequencing approaches have been developed to enrich for RNA transcripts bound to different proteins such as RNA immunoprecipitation sequencing (RIP-Seq) (Zhao *et al*, 2010) and RNA transcripts bound to genomic locations such as capture hybridization analysis of RNA targets (CHART-Seq) (Simon *et al*, 2011). As well, many sequencing innovations have endeavoured to characterize the epitranscriptome: a dynamic system

of endogenously produced RNA nucleotide modifications which change in differing contexts and hold important signalling and structural roles in RNA biology (Begik, Mattick and Novoa, 2022). Some sequencing approaches to character RNA modifications include: endoVIPER-Seq (Knutson and Heemstra, 2020) uses endonuclease V's reactivity towards inosines to characterize A-to-I editing; HydraPsiSeq (Marchand *et al*, 2020) uses aniline/hydrazine cleavage reaction of uridine nucleobases to differentiate pseudouridylation sites; GLORI (Liu *et al*, 2023) uses a deamination reaction to produce inosine nucleobases from unmodified adenosines whereby N6-Methyladenosine is resistant to deamination and can be characterized in sequencing reads as A to A, whereas unmethylated adenosine is read as A to G. All of the aforementioned methods are excellent approaches at discriminating single RNA modifications, and while they are limited to the Illumina sequencing platform, there is a highly anticipated development in third generation sequencing by Oxford Nanopore Technologies (ONT)'s native RNA sequencing platforms, which has the potential of resolving any number of RNA modifications present in any RNA sample (Garalde *et al*, 2018). Although, the development of RNA modification characterization with ONT's native RNA sequencing approaches has been slow, the technology is now capable of characterizing N6-Methyladenosine, pseudouridine, inosine and more (Liu *et al*, 2019; Begik *et al*, 2021; Nguyen *et al*, 2022).

ONT's native RNA sequencing technique primarily utilizes ligation chemistry to capture polyadenylated reads and further ligate motor protein adaptors to enable the titular nanopore based sequencing of nucleotides, which as they travel through the nanopore, differentially disturb ion gradients, producing a detectable current (Garalde *et al*, 2018). ONT's native RNA sequencing

also permits cDNA, PCR and fragmenting free sequencing, reducing sequencing bias and potentially enabling the sequencing of new differentially spliced transcripts (Garalde *et al*, 2018).

Since its inception, RNA sequencing has been at the forefront of transcriptomics and RNA biology and even with recent advancements, is benefitting from immense improvements. Throughout this thesis, much of the work is conducted by RNA-seq and variations thereof. The study of Alzheimer's disease and ncRNAs have benefitted greatly by RNA-seq (Mills *et al*, 2013; Jiang *et al*, 2020; Di Bella *et al*, 2020) and the work presented in this thesis continues in that spirit.

1.5 Rationale Objectives and Thesis Overview

Previous works studying the function of Alu and B2 SINE RNAs have uncovered a mode of regulation whereby Alu and B2 SINE RNAs are processed to elicit a stress response gene cascade. How this mode of regulation contributes to pathological processes is not understood and particularly for pathologies such as AD, has never been studied. Additionally, many ncRNAs have been studied for their RNA modifications and to an extent, this has been studied in AD as well as how some Alu RNAs are modified in AD. However, the resultant changes to the RNAs' properties are not well known. The objective of this thesis is to expand our understanding of the contributions of such ncRNAs to AD. To that end, in the following chapters, I address a number of objectives: I study how B2 controlled stress response genes (B2-SRGs) are regulated in the APP NL-G-F model (Chapter 2) and explore further how many of the same modes of regulation are dysregulated in AD patients (Chapter 3), whereby in their respective cases B2 and Alu RNAs are aberrantly processed, correlated with signatures of a stress response (Chapter 2 and 3). In Chapter 4 I develop a methodology to enable the sequencing of native ncRNAs for RNA modification analysis that we

use to analyze the differential RNA modification patterns between APP NL-G-F and wild type mice in Chapter 5. As well, I explore the regulation of A-to-I editing in the APP NL-G-F mice and how that effects some of the other molecular pathologies present in amyloid pathology.

1.6 References

Ade C, Roy-Engel AM, Deininger PL (2013) Alu elements: an intrinsic source of human genome instability. *Curr Opin Virol* 3: 639-645

Agarwal P, Enroth S, Teichmann M, Jernberg Wiklund H, Smit A, Westermarck B, Singh U (2016) Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. *Cell Cycle* 15: 1558-1571

Allen TA, Von Kaenel S, Goodrich JA, Kugel JF (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 11: 816-821

Asher S, Priefer R (2022) Alzheimer's disease failed clinical trials. *Life Sci* 306: 120861

Bajad P, Ebner F, Amman F, Szabo B, Kapoor U, Manjali G, Hildebrandt A, Janisiw MP, Jantsch MF (2020) An internal deletion of ADAR rescued by MAVS deficiency leads to a minute phenotype. *Nucleic Acids Res* 48: 3286-3303

Barnard ND, Bush AI, Ceccarelli A, Cooper J, de Jager CA, Erickson KI, Fraser G, Kesler S, Levin SM, Lucey B *et al* (2014) Dietary and lifestyle guidelines for the prevention of Alzheimer's disease. *Neurobiol Aging* 35 Suppl 2: S74-78

Bartel DP (2018) Metazoan MicroRNAs. *Cell* 173: 20-51

Bass BL, Weintraub H (1987) A developmentally regulated activity that unwinds RNA duplexes. *Cell* 48: 607-613

Bass BL, Weintraub H (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55: 1089-1098

Beebe JA, Kurz JC, Fierke CA (1996) Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA^{Asp}. *Biochemistry* 35: 10493-10505

Begik O, Lucas MC, Pryszcz LP, Ramirez JM, Medina R, Milenkovic I, Cruciani S, Liu H, Vieira HGS, Sas-Chen A *et al* (2021) Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol* 39: 1278-1291

Begik O, Mattick JS, Novoa EM (2022) Exploring the epitranscriptome by native RNA sequencing. *RNA*

Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. *Neuron* 68: 270-281

Boivin V, Deschamps-Francoeur G, Couture S, Nottingham RM, Bouchard-Bourelle P, Lambowitz AM, Scott MS, Abou-Elela S (2018) Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA* 24: 950-965

Boivin V, Reulet G, Boisvert O, Couture S, Elela SA, Scott MS (2020) Reducing the structure bias of RNA-Seq reveals a large number of non-annotated non-coding RNA. *Nucleic Acids Res* 48: 2271-2286

Brenner S, Jacob F, Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190: 576-581

Briese M, Sendtner M (2021) Keeping the balance: The noncoding RNA 7SK as a master regulator for neuron development and function. *Bioessays* 43: e2100092

Chen LL, Yang L (2017) ALU alternative Regulation for Gene Expression. *Trends Cell Biol* 27: 480-490

Chen MY, Fan K, Zhao LJ, Wei JM, Gao JX, Li ZF (2021) Long non-coding RNA nuclear enriched abundant transcript 1 (NEAT1) sponges microRNA-124-3p to up-regulate phosphodiesterase 4B (PDE4B) to accelerate the progression of Parkinson's disease. *Bioengineered* 12: 708-719

Cobb M (2015) Who discovered messenger RNA? *Curr Biol* 25: R526-532

Cornutiu G (2015) The Epidemiological Scale of Alzheimer's Disease. *J Clin Med Res* 7: 657-666

Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K (2020) Alzheimer's disease drug development pipeline: 2020. *Alzheimers Dement (N Y)* 6: e12050

Deininger P (2011) Alu elements: know the SINEs. *Genome Biology* 12: 236

Deininger PL, Moran JV, Batzer MA, Kazazian HH, Jr. (2003) Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 13: 651-658

Di Bella S, La Ferlita A, Carapezza G, Alaimo S, Isacchi A, Ferro A, Pulvirenti A, Bosotti R (2020) A benchmarking of pipelines for detecting ncRNAs from RNA-Seq data. *Brief Bioinform* 21: 1987-1998

Ding Q, Markesbery WR, Chen Q, Li F, Keller JN (2005) Ribosome dysfunction is an early event in Alzheimer's disease. *J Neurosci* 25: 9171-9175

- Dong LI, Zheng Y, Gao L, Luo X (2021) lncRNA NEAT1 prompts autophagy and apoptosis in MPTP-induced Parkinson's disease by impairing miR-374c-5p. *Acta Biochim Biophys Sin (Shanghai)* 53: 870-882
- Dunker W, Zhao Y, Song Y, Karijolic J (2017) Recognizing the SINEs of Infection: Regulation of Retrotransposon Expression and Modulation of Host Cell Processes. *Viruses* 9
- Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* 11: 822-829
- Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15: 201-206
- Gehrke S, Imai Y, Sokol N, Lu B (2010) Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. *Nature* 466: 637-641
- Giuffrida ML, Caraci F, Pignataro B, Cataldo S, De Bona P, Bruno V, Molinaro G, Pappalardo G, Messina A, Palmigiano A et al (2009) Beta-amyloid monomers are neuroprotective. *J Neurosci* 29: 10582-10587
- Goh YT, Koh CWQ, Sim DY, Roca X, Goh WSS (2020) METTL4 catalyzes m6Am methylation in U2 snRNA to regulate pre-mRNA splicing. *Nucleic Acids Res* 48: 9250-9261
- Gotz J, Schild A, Hoernndli F, Pennanen L (2004) Amyloid-induced neurofibrillary tangle formation in Alzheimer's disease: insight from transgenic mouse and tissue-culture models. *Int J Dev Neurosci* 22: 453-465
- Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123: 631-640
- Grimm A, Friedland K, Eckert A (2016) Mitochondrial dysfunction: the missing link between aging and sporadic Alzheimer's disease. *Biogerontology* 17: 281-296
- Gros F, Hiatt H, Gilbert W, Kurland CG, Risebrough RW, Watson JD (1961) Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. *Nature* 190: 581-585
- Hancks DC, Kazazian HH, Jr. (2012) Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 22: 191-203
- Hane FT, Robinson M, Lee BY, Bai O, Leonenko Z, Albert MS (2017) Recent Progress in Alzheimer's Disease Research, Part 3: Diagnosis and Treatment. *J Alzheimers Dis* 57: 645-665
- He N, Jin WL, Lok KH, Wang Y, Yin M, Wang ZJ (2013) Amyloid-beta(1-42) oligomer accelerates senescence in adult hippocampal neural stem/progenitor cells via formylpeptide receptor 2. *Cell Death Dis* 4: e924

- Hernandez AJ, Zovoilis A, Cifuentes-Rojas C, Han L, Bujisic B, Lee JT (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc Natl Acad Sci U S A* 117: 415-425
- Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A (1965) Structure of a Ribonucleic Acid. *Science* 147: 1462-1465
- Huang N, Groover D, Damania B, Moody C (2022) Apoptotic caspases suppress an MDA5-driven IFN response during productive replication of human papillomavirus type 31. *Proc Natl Acad Sci U S A* 119: e2200206119
- Huang YM, Shen J, Zhao HL (2019) Major Clinical Trials Failed the Amyloid Hypothesis of Alzheimer's Disease. *J Am Geriatr Soc* 67: 841-844
- Hube F, Francastel C (2018) Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Front Genet* 9: 140
- Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M *et al* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 65: 7065-7070
- Ivanova E, Berger A, Scherrer A, Alkalaeva E, Strub K (2015) Alu RNA regulates the cellular pool of active ribosomes by targeted delivery of SRP9/14 to 40S subunits. *Nucleic Acids Res* 43: 2874-2887
- Izumi H, Sato K, Kojima K, Saito T, Saido TC, Fukunaga K (2020) Oral glutathione administration inhibits the oxidative stress and the inflammatory responses in App(NL-G-F/NL-G-F) knock-in mice. *Neuropharmacology* 168: 108026
- Jiang J, Wang C, Qi R, Fu H, Ma Q (2020) scREAD: A Single-Cell RNA-Seq Database for Alzheimer's Disease. *iScience* 23: 101769
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583-589
- Knutson SD, Heemstra JM (2020) EndoVIPER-seq for Improved Detection of A-to-I Editing Sites in Cellular RNA. *Curr Protoc Chem Biol* 12: e82
- Kramerov DA, Vassetzky NS (2011) SINEs. *Wiley Interdiscip Rev RNA* 2: 772-786
- Kuo MC, Liu SC, Hsu YF, Wu RM (2021) The role of noncoding RNAs in Parkinson's disease: biomarkers and associations with pathogenic pathways. *J Biomed Sci* 28: 78
- Lacoux C, Di Marino D, Boyl PP, Zalfa F, Yan B, Ciotti MT, Falconi M, Urlaub H, Achsel T, Mougin A *et al* (2012) BC1-FMRP interaction is modulated by 2'-O-methylation: RNA-binding activity of the tudor domain and translational regulation at synapses. *Nucleic Acids Res* 40: 4086-4096

LeGendre JB, Campbell ZT, Kroll-Conner P, Anderson P, Kimble J, Wickens M (2013) RNA targets and specificity of Staufien, a double-stranded RNA-binding protein in *Caenorhabditis elegans*. *J Biol Chem* 288: 2532-2545

Leng F, Edison P (2021) Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat Rev Neurol* 17: 157-172

Leppek K, Byeon GW, Kladwang W, Wayment-Steele HK, Kerr CH, Xu AF, Kim DS, Topkar VV, Choe C, Rothschild D *et al* (2022) Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat Commun* 13: 1536

Liu C, Sun H, Yi Y, Shen W, Li K, Xiao Y, Li F, Li Y, Hou Y, Lu B *et al* (2023) Absolute quantification of single-base m(6)A methylation in the mammalian transcriptome using GLORI. *Nat Biotechnol* 41: 355-366

Liu H, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, Schwartz S, Mattick JS, Smith MA, Novoa EM (2019) Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat Commun* 10: 4079

Llanos-Gonzalez E, Henares-Chavarino AA, Pedrero-Prieto CM, Garcia-Carpintero S, Frontinan-Rubio J, Sancho-Bielsa FJ, Alcain FJ, Peinado JR, Rabanal-Ruiz Y, Duran-Prado M (2019) Interplay Between Mitochondrial Oxidative Disorders and Proteostasis in Alzheimer's Disease. *Front Neurosci* 13: 1444

Maia MA, Sousa E (2019) BACE-1 and gamma-Secretase as Therapeutic Targets for Alzheimer's Disease. *Pharmaceuticals (Basel)* 12

Manoharan S, Guillemin GJ, Abiramasundari RS, Essa MM, Akbar M, Akbar MD (2016) The Role of Reactive Oxygen Species in the Pathogenesis of Alzheimer's Disease, Parkinson's Disease, and Huntington's Disease: A Mini Review. *Oxid Med Cell Longev* 2016: 8590578

Marchand V, Pichot F, Neybecker P, Ayadi L, Bourguignon-Igel V, Wacheul L, Lafontaine DLJ, Pinzano A, Helm M, Motorin Y (2020) HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res* 48: e110

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 29: 499-509

Mathew BA, Katta M, Ludhiadch A, Singh P, Munshi A (2023) Role of tRNA-Derived Fragments in Neurological Disorders: a Review. *Mol Neurobiol* 60: 655-671

McGuire A, Brown JA, Kerin MJ (2015) Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring. *Cancer Metastasis Rev* 34: 145-155

Melese T, Xue Z (1995) The nucleolus: an organelle formed by the act of building a ribosome. *Curr Opin Cell Biol* 7: 319-324

- Micheau O, Tschopp J (2003) Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* 114: 181-190
- Miladi M, Raden M, Diederichs S, Backofen R (2020) MutaRNA: analysis and visualization of mutation-induced changes in RNA structure. *Nucleic Acids Res* 48: W287-W291
- Mills JD, Nalpathamkalam T, Jacobs HI, Janitz C, Merico D, Hu P, Janitz M (2013) RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neurosci Lett* 536: 90-95
- Minshall N, Git A (2020) Enzyme- and gene-specific biases in reverse transcription of RNA raise concerns for evaluating gene expression. *Sci Rep* 10: 8151
- Moulton MJ, Barish S, Ralhan I, Chang J, Goodman LD, Harland JG, Marcogliese PC, Johansson JO, Ioannou MS, Bellen HJ (2021) Neuronal ROS-induced glial lipid droplet formation is altered by loss of Alzheimer's disease-associated genes. *Proc Natl Acad Sci U S A* 118
- Nilsson P, Saito T, Saido TC (2014) New mouse model of Alzheimer's. *ACS Chem Neurosci* 5: 499-502
- Nguyen TA, Heng JWW, Kaewsapsak P, Kok EPL, Stanojevic D, Liu H, Cardilla A, Praditya A, Yi Z, Lin M et al (2022) Direct identification of A-to-I editing sites with nanopore native RNA sequencing. *Nat Methods*
- Nunomura A, Perry G (2020) RNA and Oxidative Stress in Alzheimer's Disease: Focus on microRNAs. *Oxid Med Cell Longev* 2020: 2638130
- O'Neil D, Glowatz H, Schlumpberger M (2013) Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol* Chapter 4: Unit 4 19
- Ono K, Condron MM, Teplow DB (2009) Structure-neurotoxicity relationships of amyloid beta-protein oligomers. *Proc Natl Acad Sci U S A* 106: 14745-14750
- Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* 20: 89-108
- Pestal K, Funk CC, Snyder JM, Price ND, Treuting PM, Stetson DB (2015) Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development. *Immunity* 43: 933-944
- Ponicsan SL, Houel S, Old WM, Ahn NG, Goodrich JA, Kugel JF (2013) The non-coding B2 RNA binds to the DNA cleft and active-site region of RNA polymerase II. *J Mol Biol* 425: 3625-3638
- Ponicsan SL, Kugel JF, Goodrich JA (2010) Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* 20: 149-155

Ponicsan SL, Kugel JF, Goodrich JA (2015) Repression of RNA Polymerase II Transcription by B2 RNA Depends on a Specific Pattern of Structural Regions in the RNA. *Noncoding RNA* 1: 4-16

Ramos J, Han L, Li Y, Hagelskamp F, Kellner SM, Alkuraya FS, Phizicky EM, Fu D (2019) Formation of tRNA Wobble Inosine in Humans Is Disrupted by a Millennia-Old Mutation Causing Intellectual Disability. *Mol Cell Biol* 39

Rao MS, Van Vleet TR, Ciurlionis R, Buck WR, Mittelstadt SW, Blomme EAG, Liguori MJ (2018) Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front Genet* 9: 636

Rasgon NL, Kenna HA (2005) Insulin resistance in depressive disorders and Alzheimer's disease: revisiting the missing link hypothesis. *Neurobiol Aging* 26 Suppl 1: 103-107

Rebagliati MR, Melton DA (1987) Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell* 48: 599-605

Reddy KB (2015) MicroRNA (miRNA) in cancer. *Cancer Cell Int* 15: 38

Saito T, Matsuba Y, Mihira N, Takano J, Nilsson P, Itohara S, Iwata N, Saido TC (2014) Single App knock-in mouse models of Alzheimer's disease. *Nat Neurosci* 17: 661-663

Salta E, De Strooper B (2012) Non-coding RNAs with essential roles in neurodegenerative disorders. *Lancet Neurol* 11: 189-200

Schieber M, Chandel NS (2014) ROS function in redox signaling and oxidative stress. *Curr Biol* 24: R453-462

Seino Y, Kawarabayashi T, Wakasaya Y, Watanabe M, Takamura A, Yamamoto-Watanabe Y, Kurata T, Abe K, Ikeda M, Westaway D *et al* (2010) Amyloid beta accelerates phosphorylation of tau and neurofibrillary tangle formation in an amyloid precursor protein and tau double-transgenic mouse model. *J Neurosci Res* 88: 3547-3554

Shishova KV, Khodarovich YM, Lavrentyeva EA, Zatsepina OV (2015) High-resolution microscopy of active ribosomal genes and key members of the rRNA processing machinery inside nucleolus-like bodies of fully-grown mouse oocytes. *Exp Cell Res* 337: 208-218

Shivapurkar N, Weiner LM, Marshall JL, Madhavan S, Deslattes Mays A, Juhl H, Wellstein A (2014) Recurrence of early stage colon cancer predicted by expression pattern of circulating microRNAs. *PLoS One* 9: e84686

Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE (2011) The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* 108: 20497-20502

Song XN, Zhang LQ, Liu DG, Lin J, Zheng JD, Dai DP, Hei AL, Hayakawa H, Sekiguchi M, Cai JP (2011) Oxidative damage to RNA and expression patterns of MTH1 in the hippocampi of

senescence-accelerated SAMP8 mice and Alzheimer's disease patients. *Neurochem Res* 36: 1558-1565

Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A* 90: 1977-1981

Sun T, Yu Y, Wu X, Acevedo A, Luo JD, Wang J, Schneider WM, Hurwitz B, Rosenberg BR, Chung H *et al* (2021) Decoupling expression and editing preferences of ADAR1 p150 and p110 isoforms. *Proc Natl Acad Sci U S A* 118

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24: 390-400

Tiiman A, Palumaa P, Tougu V (2013) The missing link in the amyloid cascade of Alzheimer's disease - metal ions. *Neurochem Int* 62: 367-378

Turowski TW, Boguta M (2021) Specific Features of RNA Polymerases I and III: Structure and Assembly. *Front Mol Biosci* 8: 680090

Urano A, Matsumaru D, Ryoke R, Saito R, Kadoguchi S, Saigusa D, Saito T, Saido TC, Kawashima R, Yamamoto M (2020) Nrf2 Suppresses Oxidative Stress and Inflammation in App Knock-In Alzheimer's Disease Model Mice. *Mol Cell Biol* 40

Vogel OA, Han J, Liang CY, Manicassamy S, Perez JT, Manicassamy B (2020) The p150 Isoform of ADAR1 Blocks Sustained RLR signaling and Apoptosis during Influenza Virus Infection. *PLoS Pathog* 16: e1008842

Willi J, Kupfer P, Evequoz D, Fernandez G, Katz A, Leumann C, Polacek N (2018) Oxidative stress damages rRNA inside the ribosome and differentially affects the catalytic center. *Nucleic Acids Res* 46: 1945-1957

Winek K, Lobentanzer S, Nadorp B, Dubnov S, Dames C, Jagdmann S, Moshitzky G, Hotter B, Meisel C, Greenberg DS *et al* (2020) Transfer RNA fragments replace microRNA regulators of the cholinergic poststroke immune blockade. *Proc Natl Acad Sci U S A* 117: 32606-32616

Xie Z, Wei M, Morgan TE, Fabrizio P, Han D, Finch CE, Longo VD (2002) Peroxynitrite mediates neurotoxicity of amyloid beta-peptide1-42- and lipopolysaccharide-activated microglia. *J Neurosci* 22: 3484-3492

Yakovchuk P, Goodrich JA, Kugel JF (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci U S A* 106: 5569-5574

Yamagami R, Sieg JP, Bevilacqua PC (2021) Functional Roles of Chelated Magnesium Ions in RNA Folding and Function. *Biochemistry* 60: 2374-2386

Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, Oostra B, Bagni C (2003) The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112: 317-327

Zhang T, Pang P, Fang Z, Guo Y, Li H, Li X, Tian T, Yang X, Chen W, Shu S *et al* (2018) Expression of BC1 Impairs Spatial Learning and Memory in Alzheimer's Disease Via APP Translation. *Mol Neurobiol* 55: 6007-6020

Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* 15: 419

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40: 939-953

Zovoilis A, Cifuentes-Rojas C, Chu HP, Hernandez AJ, Lee JT (2016) Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* 167: 1788-1802 e1713

Chapter 2 - Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology

2.1 Foreword

This chapter is presented in the form of the manuscript published in eLife (Cheng *et al*, 2020). As a co-first author on the manuscript, alongside Yubo Cheng, Babita Gollen and Chris Isaac, I contributed “Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing” (Cheng *et al*, 2020). Yubo Cheng contributed “Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing” (Cheng *et al*, 2020). Babita Gollen contributed “Investigation, Methodology, Project administration” (Cheng *et al*, 2020). Chris Isaac contributed “Data curation, Formal analysis, Investigation, Methodology” (Cheng *et al*, 2020). Additional work was performed by Abel Belay (“Investigation, qPCR assays”), Jogender Mehla (“Resources, Methodology”), Kush Patel (“WB Assays”), Nehal Thakor (“Resources, WB assays”), Majid Mohajerani (“Resources, Supervision, Project administration”), and Athanasios Zovoilis (“Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing”) (Cheng *et al*, 2020).

The research presented in this chapter investigates the phenomenon of aberrant B2 RNA processing, correlating with constitutive expression of B2 RNA regulated genes identified in Zovoilis *et al*, 2016, presented in mouse hippocampus tissue of the NL-G-F APP⁺ model. This model is commonly used for its presentation of amyloidogenesis after 6 months of age and presents symptoms consistent with the course of amyloidogenesis in Alzheimer’s disease patients.

In addition to *in vivo* models, numerous *in vitro* models were employed to establish the mechanism by which B2 RNA processing was accelerated. Briefly, the master heat shock regulator protein, Hsf1 was identified as a possible candidate due to its high expression in neural tissue and its Chip-Seq feature overlap with B2 regulated genes. When incubated with Hsf1, B2 RNA is more rapidly processed.

2.2 Abstract

The functional importance of many non-coding RNAs (ncRNAs) generated by repetitive elements and their connection with pathological processes remains elusive. B2 RNAs, a class of ncRNAs of the B2 family of SINE repeats, mediate through their processing the transcriptional activation of various genes in response to stress. Here, we show that this response is dysfunctional during amyloid beta toxicity and pathology in the mouse hippocampus due to increased levels of B2 RNA processing, leading to constitutively elevated B2 RNA target gene expression and high Trp53 levels. Evidence indicates that Hsf1, a master regulator of stress response, mediates B2 RNA processing in hippocampal cells and is activated during amyloid toxicity, accelerating the processing of SINE RNAs and gene hyper-activation. Our study reveals that in mouse, SINE RNAs constitute a novel pathway deregulated in amyloid beta pathology, with potential implications for similar cases in the human brain, such as Alzheimer's disease (AD).

2.3 Introduction

The number of patients with Alzheimer's disease (AD) is expected to skyrocket in the upcoming years (Cornutiu, 2015). The exact molecular mechanisms underlying AD are not fully understood, a fact that is underlined by the inauspicious results of recent clinical trials for potential therapeutic

agents (Hane et al., 2017; Cummings et al., 2019). Amyloid pathology, and particularly, amyloid beta peptides and their aggregated forms have been connected with AD pathogenesis (Bloom, 2014) as well as with neurotoxicity in mouse models of amyloid pathology (Ittner et al., 2010). Nevertheless, the transcriptome changes involved in cell stress response to amyloid toxicity in brains with extensive amyloid beta pathology are still not entirely clear. The hippocampus is a primary target of amyloid pathology in humans. In healthy hippocampi, among the genes that have been implicated in transcriptome-environment interactions are stress response genes (SRGs) (Gallo et al., 2018). These genes have been initially described in other biological contexts, such as thermal and oxidative stress, as pro-survival genes activated early after the application of a stress stimulus, such as heat shock, that help the cell overcome the stress condition (Mahat et al., 2016). However, in addition to the cellular response to stress, many SRGs were shown to have a central role in the function of the mouse hippocampus, by mediating cell signaling and genome-environment interactions. In particular, we and others have shown that in the healthy hippocampus during neural response to environmental stimuli, SRGs, such as those of the MAPK pathway, are transiently activated during various hippocampal processes including learning and response to cellular stress (Peleg et al., 2010; Yutsudo et al., 2013; Sananbenesi et al., 2003). Activation of SRGs is followed by a transient upregulation of the pro-apoptotic factor Trp53 (Peleg et al., 2010) and, subsequently, of a pro-apoptotic miRNA, Mir34c, which are transiently induced by and as a response to the activation of pro-survival SRGs (Zovoilis et al., 2011; Yamakuchi and Lowenstein, 2009). Trp53 activates the expression of genes engaged in promoting cell death in response to multiple forms of cellular stress including Mir34c (Yamakuchi and Lowenstein, 2009). This miRNA acts transiently as a guard and fine tuner of the expression of many SRGs by targeting them, thus, creating a negative feedback regulatory loop that keeps SRG expression in healthy

cells under strict control. This facilitates the return to the pro-stimulation state in approximately 3 hr after the application of the stimulus (Zovoilis et al., 2011). In contrast, hippocampi of mouse models of amyloid pathology and postmortem brains of human patients of AD are characterized by abnormally high Mir34c levels that subsequently can lead to prolonged high Trp53 levels and neural death (Zovoilis et al., 2011; Yamakuchi and Lowenstein, 2009). Given that many SRGs are upstream regulators of Trp53-Mir34c activation (Gao et al., 2010), high Trp53 and Mir34c levels in amyloid pathology implied a possible transcriptome deregulation of the pathways that involve SRGs but whether such a deregulation exists, and which is the mechanism underlying this, it remained unknown. Interestingly, in a recent publication, we showed that expression of a number of SRGs is regulated by a class of non-protein coding (non-coding) RNAs called B2 SINE RNAs (Zovoilis et al., 2016), raising the possibility that these non-coding RNAs may be a missing link in the pathways connecting amyloid beta toxicity with transcriptome changes in mouse hippocampus during amyloid pathology.

SINE non-coding RNAs are transcribed by repetitive small interspersed nuclear elements (SINEs), with the subclass B2-repetitive elements being one of the most frequent in mouse (Kramarov and Vassetzky, 2011). SINE B2 elements, which are retrotransposons present in hundreds of thousands of copies, are part of the non-protein coding genome, and over the long haul, they have been regarded as genomic parasites and ‘junk DNA’ with no function (Karijolich et al., 2017). However, SINE B2 elements can be transcribed by RNA Polymerase III into SINE B2 RNAs (Kramarov and Vassetzky, 2011) and a number of recent studies have revealed a key role for SINE B2 RNAs in cellular response to stress. In particular, studies from the J. Kugel and J. Goodrich labs have shown that during response to cellular stress, levels of SINE B2 RNAs increase and suppress the

transcription of a number of housekeeping genes through binding of RNA Polymerase II (RNA Pol II), potentially facilitating the redirection of cell resources to pro-survival pathways (Yakovchuk et al., 2009). In addition, we have recently shown that SINE B2 RNAs mediate cellular response to stress through the regulation of pro-survival stress response genes by acting as transcriptional switches. In particular, in the pro-cellular stress state SINE B2 RNAs bind RNA Pol II at several SRGs and suppress their transcription. In this way, stalled or delayed RNA Pol II remains poised for a fast activation and ramp-up of transcription when needed. Upon application of a stress stimulus, SINE B2 RNAs, which have a self-cleavage activity that is accelerated by their interaction with a protein (Ezh2) (Hernandez et al., 2020), are processed into unstable fragments that lack the ability to bind and suppress RNA Pol II. This event releases the delayed or stalled RNA Pol II and enables fast transcriptional activation of stress response genes (Zovoilis et al., 2016).

The above findings have revealed a novel role in cellular function for processing of SINE B2 RNAs (hereafter referred simply as B2 RNAs) through the activation of a number of SRGs regulated by them (hereafter called B2-SRGs). However, an association of the processing and destabilization of B2 RNAs with any pathological cellular process remains unknown. Here, given the importance of SRGs in hippocampal neuronal function, we examine whether this newly described B2 RNA-SRG regulatory mechanism is linked with pathological processes, focusing on transcriptome response to amyloid beta toxicity and pathology. To this end, we investigate whether B2-SRGs are indeed deregulated during amyloid pathology, which would imply a role for B2 RNAs in this condition and we examine whether amyloid toxicity is connected with changes in B2 RNA processing. Subsequently, we investigate the further upstream molecular mechanisms

underlying any potential deregulation of B2 RNA processing in response to amyloid toxicity in hippocampal neural cells.

2.4 Materials and Methods

2.4.1 Amyloid beta peptide preparations

The amyloid beta 1–42 peptides and the respective control peptides (having the reverse aa sequence compared to 1–42 peptides) were synthesized by Sigma Aldrich’s custom synthesis service using the following sequences: DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA (peptide 1-42) and AIVVGGVMLGIIAGKNSGVDEAFFVLKQHHVEYGS DHRFEAD (Reverse). Upon receiving, the peptides were dissolved in 10% NH₄OH at a concentration of 2.1 mg/mL, sonicated for 5 min, aliquoted, dried, and stored at –80°C for further use (see below).

2.4.2 Animals and behavioral measurement

Immunohistochemistry was done as described in our previous study (Mehla et al., 2019) for the same mice cohorts for which behavioral studies were performed in that study (Mehla et al., 2019). In brief, mouse pairs of APP-KI mice carrying Arctic, Swedish, and Beyreuther/Iberian mutations (APPNL-G-F/NL-G-F) were gifted by RIKEN Center for Brain Science, Japan, and the colony of these mice was maintained at Canadian Center for Behavioral Neuroscience vivarium. C57BL/6J mice were used as a WT control and all animals were housed in groups of four mice in each cage in a controlled environment (22°C–25°C, 50% humidity and a 12 hr light:dark cycle). All experimental procedures were approved by the institutional animal care committee and performed in accordance with the standards set out by the Canadian Council for Animal Care. In total, we

extracted RNA from hippocampi of three different mice per age group per condition. During the hippocampal RNA extractions, the RNA of one of the three 3-month-old wild-type mice had very low RIN scores, which could be a confounding factor for the short-RNA-seq. As this happened some months after the hippocampal extractions, we did not have available other 3-month mice of the same cohort used in our previous study of these mice for the behavioral and IHC studies. Thus, we decided to include only two replicates in this condition. Since the results presented in the current study involve mainly 6-month-old mice we expect the impact to be minimal.

2.4.3 Cell culture and transfections

HT22 cells, from an immortalized mouse hippocampal cell line (Davis and Maher, 1994) (Millipore Cat# SCC129, RRID:CVCL_0321). Cells were provided by the vendor tested negative for infectious diseases by a Mouse Essential CLEAR panel by Charles River Animal Diagnostic Services and verified to be of mouse origin and negative for inter-species contamination from rat, chinese hamster, Golden Syrian hamster, human, and non-human primate (NHP) (as assessed by a Contamination Clear panel by Charles River Animal Diagnostic Services) as well as negative for mycoplasma contamination. Cells were cultured in DMEM (Sigma) and 1% Penicillin/Streptomycin (Gibco). Cells were thawed and passaged at least 3 days before the transfection date in order to allow sufficient time for cells to recover from the stress of cryopreservation and not interfere with the assessment of cellular response to stress in subsequent experiments. For knocking down of Hsf1 mRNA levels, we used an LNA long RNA GapmeR against Hsf1 (Exiqon/Qiagen) with the following sequence: 5'-GAAGGATGGAGTCAA-3' and an LNA long RNA GapmeR with the following scramble sequence: 5'-CCTCAATTTTATCAC-3'. For knocking down B2 RNA transcripts, we used an anti-B2 LNA pool consisting of five LNAs,

synthesized using the K and E DNA and RNA synthesizer: 1. 5' - G*T*T*A*CGGATGGTT*G*T*G* - 3'; 2. 5'- A*G*A*T*CTCATTACA*G*A*T*- 3'; 3. 5' - A*G*A*T*CTCATTACG*G*A*T* - 3''; 4. 5' - A*G*A*T*CCCATTACA*G*A*T* - 3'; 5. 5' - A*G*A*T*CCCATTACG*G*A*T* - 3'; 6. 5'-T*G*TAGCTGTCTTCA*G*-3'. Nucleotides were ordered from Sigma Aldrich as DMT phosphoramidites. The day of LNA transfections, following 5-min incubation with TrypLE Express Enzyme (Gibco) (1x), cells were passaged and transferred to a six-well plate at a 100,000 cells/well density and LNA transfections were performed simultaneously, using the HiPerfect reagent (Qiagen). Transfection was performed as follows: Firstly, LNAs were reconstituted in nuclease-free water to 50 μ M. Subsequently, 3 μ L 50 μ M LNA were mixed and incubated with 4 μ L HiPerfect reagent and 30 μ L nuclease-free water, at room temperature for 20 min, and then added drop-wise to cells that had just been plated in 1 ml of medium/well (still not attached) to a final LNA concentration of 150 nM. For anti-Hsf1 LNAs transfections, incubation with amyloid beta (peptide 1-42) and control peptides (Reverse) was performed 24 hr after transfection with LNAs. Peptides were initially dissolved in DMSO for incubation at 37°C for 1 hr and then added to cells to a final concentration of 30 μ M for 6 hr before treating with 0.5 mL TrypLE (Gibco), pelleting cells at 1000 rpm for 5 min and resuspending the pellet in 1 mL Trizol reagent (Thermofisher) for RNA extraction based on Manufacturer's instructions. For anti-B2 LNA, transfection and incubation lasted 6 hr before RNA extraction.

2.4.4 Reverse transcription and quantitative PCR

RNA was extracted using Trizol reagent as described and reverse transcribed using Superscript III (18080093, Invitrogen) by the following method: 50 ng total RNA was mixed with 100 ng random primers (C1181, Promega) and 1 μ L of 10 mM dNTP mix. The mixture was incubated for 5 min

at 65°C and placed immediately on ice. The mixture was then incubated with 4 µL of 5x First strand buffer, 1 µL of 0.1M DTT and 0.4 µL Superscript III (18080093, Invitrogen) for 5 min at 25°C, 60 min at 55°C and 15 min at 70°C. cDNA was analyzed by qPCR using 2 µL of 1:20 diluted cDNA, 0.5 µL of 10 µM of each gene-specific primer, 2 µL H₂O and 5 µL of Luna Universal qPCR Master Mix (M3003E, NEB). Thermocycler conditions are as follows: 3 min at 95°C (15 s at 95°C, 30 s at 54°C, 30 s at 66°C) × 40 cycles. Fluorometer readings were taken during extension and qPCR was performed using the Bio-Rad CFX384 Real-time detection system. Standard curves were prepared for relative expression and the analysis of PCR efficiency by pooling 2 µL of each cDNA sample and standard diluting SD1: 1:5, SD2: 1:10, SD3: 1:20, SD4: 1:40. Samples were either analyzed by standard curve relative expression or 2^{-ΔΔCT} fold change analysis. Student's T-test were used to study significance as described in the respective figure legends. Primers were ordered from IDT as custom oligos and are listed in Appendix 1—table 1.

2.4.5 RNA in vitro transcription and RNA-protein incubations

B2 template for in vitro RNA transcription was ordered as IDT g-block (lower case denotes the T7 promoter sequence): 5'-taatacgactcactata GGGGCTGGTGAGATGGCTCAGTGGGTAAGAGCACCCGACTGCTCTTCCGAAGGTCC GGAGTTCAAATCCCAGCAACCACATGGTGGCTCACAACCATCCGTAACGAGATCTG ACTCCCTCTTCTGGAGTGTCTGAAGACAGCTACAGTGTACTTACATATAATAAATAA ATAAATCTTTAAAAAAA - 3'. The template was amplified by PCR using a T7 promoter sequence as the forward primer: 5'-TAATACGACTCACTATAG and the following sequence as reverse primer: 5'-TTTTTTTTTAAAGATTTATTTATTTATTATATGTAAGTACA. B2mut4b

was transcribed from the following template (same primers):
 taatacgaactcactataGGGCTGGTGAGATGGCTCAGTGGGTAAGAGCACCCGACTGCTCTTCC
 GAAGGTCCGGAGTTCAAATCCCAGCAACCACATGGTGGCTCACAACCATCCGTAAC
 GAGATCTGACTCCCTCTTCTTCTGAAGACAGCTACAGTGTACTTACATATAATAAAT
 AAATAAATCTTTAAAAA. Primers were diluted to 10 mM and PCR was performed
 using the NEB Q5 polymerase, Q5 reaction buffer (10x), Q5 high GC enhancer (10x). The reaction
 proceeded at hot start 98°C – 30 s (98°C – 5 s, 58°C – 10 s, 72°C – 10 s) X35 cycles, 72°C – 10
 min. The samples were then analyzed by agarose gel electrophoresis (Bio-Rad, 1613100EDU) and
 the bands were gel extracted at the desired size and purified using the BioBasic EZ-10 gel
 extraction kit (BS353). A subsequent PCR was then repeated and 1 µg of the amplified g-block
 was then in vitro transcribed by T7 RNA polymerase (NEB, M0251) for 2 hr at 37°C. The reaction
 was buffered using the T7 RNA polymerase buffer in addition to 10 mM NTPs (ATP: P1132, CTP:
 P1142, GTP: P1152, UTP: P1162) in a final 20 µL reaction. RNA was purified using the Zymo
 Research RNA Clean and Concentrator - 25 kit. Sequences of RNA controls used were as follows:

Control #1(G-44U):
 5'GCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAACAACCCCAUCCGUCAUGG
 AGACAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUG
 CUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC 3', control #2:(5' Zika UTR
 scramble):5'
 UACAAUCACGAAAGUCAAUUAUAGUUUCGAGUCGUAACGAGAACAUUUCGCG
 GACCAAUUUAAGGAGUAACUAAAGUGUGAAAUGAUUCCGGAAUACUGUUGAAAU
 UGCGGAUCGAGCUUGCAGCCGUUAAAUAACCGGACGUUAGUGAAGUGCAGAUAU
 G 3'.

B2 RNA re-folding and B2 RNA – Hsf1 incubations were performed as described previously (14). In brief, in vitro-transcribed B2 RNA was folded with 300 mM NaCl through incubation for 1 min at 50°C and cooling at a rate of 1°C/10 s until 4°C. Subsequently, B2 RNAs were incubated at a final concentration of 0.4 μ M unless otherwise stated with the addition of Hsf1 diluted in TAP buffer (final reaction concentrations: 5 nM Tris pH 7.9, 0.5 mM MgCl₂, 0.02 mM EDTA, 0.01% NP-40, 1% glycerol, 0.2 mM DTT). Hsf1 protein incubations were performed with phosphorylated, recombinant, His-tagged Hsf1 (Enzo Life Sciences: ADI-SPP-902-F; ~60KDa). Hsf1 working concentrations were 250 nM unless otherwise specified, diluted in TAP buffer. PNK (NEB, M0201; ~35KDa) was used as a control protein because it is an RNA-binding protein that is used in the construction of our short RNA libraries and we wanted show that short RNA seq data are free of such confounding factors that could potentially generate artificial fragments. PolyA (NEB, M0276; ~56KDa) was used as control protein because it is an RNA-binding protein that should not process B2 RNAs. Denatured Hsf1 was prepared by diluting in TAP buffer as above and heating at 95°C for 3 min; denatured Hsf1 was used as a control to ensure buffer components of Hsf1 themselves were not responsible for B2 RNA processing.

Fragmentation of B2 RNA was analyzed on 8M urea 10% PAGE gels stained by SYBR II (Invitrogen, S7564). Gel analysis occurred on Amersham Typhoon instruments. Band absorbance was analyzed using ImageJ area under the curve software and normalized by the ratio of experimental over initial as described previously (Zovoilis et al., 2016).

2.4.6 Short-RNA-seq and long-RNA-seq

Using the miRvana miRNA size selection kit (Thermo Fisher) as described before, 1.5 µg total RNA was size separated into short and long fractions (Zovoilis et al., 2016). In brief, following addition of the lysis/binding buffer and the homogenate additive solution to the RNA, 1/3 of the volume 100% EtOH was added and the mix was passed through the column for binding long RNAs. 100% EtOH at 2/3 of the flow through volume was subsequently added to the flow through and passed through a second column for binding short RNAs. Eluted RNAs were tested for size and quality using the Agilent Bioanalyzer RNA pico-kit. For long-RNA-seq, the long RNA fractions were cleaned and concentrated using the RNeasy Minelute kit (Qiagen) and ribodepleted using the rRNA depletion kit (NEB). The library was then prepared using the NEBNext Ultra II direction Library preparation kit (NEB, E7760), and sample cleanups were performed using the Omega NGS Total Pure Mag Beads (Omega, SKU: M1378-01) 0.5X and 1.2X before library amplification and 0.9X following amplification. nine cycles were used during amplification. For short-RNA-seq, the short RNA fractions were subjected to 3'-phosphoryl removal for 1 hr at 37°C, treated with the T4 PNK enzyme (NEB, M0201), using exclusively the 10X PNK buffer (NEB). The short fractions were cleaned and concentrated using the RNeasy Minelute kit and the library was prepared using the NEBNext small RNA library prep set (E7330) as described before (Zovoilis et al., 2016). Sample cleanups were performed using again the Omega NGS Total Pure Mag Beads (Omega) 1.2X following library amplification. Library amplification used 15 cycles. Quantification of libraries was done by qPCR using the NEBNext library quant kit for Illumina (NEB, E7630) and library sizes were analyzed using the Agilent bioanalyzer 2100 HS DNA kit. Equimolar amounts were prepared for sequencing. Libraries were sequenced on an Illumina HiSeq platform using 150nt read lengths. For B2 RNA in vitro sequencing, a 10X reaction as described

of B2 RNA in solution with Hsf1 was incubated for 60 min at 37°C. The RNA was prepared as described above at twice the concentration of the recommended kit components. RNA was cleaned and concentrated using the modified RNeasy minElute kit.

2.4.7 Western blotting

HT22 cells were seeded at 100,000 cells per well and transfected with LNA in a six-well plate as described. Cells were harvested and lysed using RIPA lysis buffer with protease and phosphatase inhibitors added. Next, the cell suspension was centrifuged at 13,000 rpm for 20 min at 4°C, supernatant was aspirated and pellet discarded. Protein concentration was determined using a Bradford assay (bio-rad; 500–0205). Equal amounts of soluble protein were loaded (25 µg) for resolving with SDS-PAGE. Proteins were transferred onto nitrocellulose membrane (GE healthcare), blocked for 1 hr with 1% milk in 0.02% PBST at room temperature. Individual proteins were detected using polyclonal rabbit HSF1 antibody (1:1000; Enzo Life Sciences; ADI-SPA-901; RRID:AB_10616511). After incubating primary antibody, blots were washed using 0.1% PBST and reprobed with anti-rabbit-HRP conjugate secondary antibody (1:5000; abcam; ab97051). Proteins were visualized using Pierce enhanced chemiluminescence detection system (Thermo Fisher Scientific; 32106). Blots were imaged in an AI600 imager (GE Healthcare) and densitometry performed using ImageJ.

2.4.8 Bioinformatics analysis

For the tissue enrichment and GO term analysis of B2-SRGs (Supplementary file 1), we used the DAVID function annotation platform (DAVID 6.8, February 2020) with default parameters (EASE score 0.1, max. 1000 entries), and a reporting EASE score threshold of 0.05 and p-adjusted

values calculated based on the Benjamini method (Huang et al., 2009a; Huang et al., 2009b). For both GO Biological process and Cellular Compartment, we selected the BP- or CC-direct options.

For the analysis of the short-RNA-seq and long-RNA-seq data, initially FastQC (Babraham Bioinformatics, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was run for quality control of generated reads in fastq format. Subsequently, standard Illumina adaptor sequences were trimmed off using cutadapt-1.18 (<https://doi.org/10.14806/ej.17.1.200>). Short-RNA-seq reads were mapped to mouse reference genome (UCSC mm10) (November 2017) using bwa-0.7.17 in single-end mode with default aln parameters (Li and Durbin, 2009). Long-RNA-seq reads for each sample were mapped to reference genome ensembl GRCm38 (November 2018) primary assembly using hisat2-2.1.0, in single-end mode, with the following parameters: Report alignments tailored for transcript assemblers including StringTie, searches for at most one distinct, primary alignments for each read (Kim et al., 2019). SAM format files generated from mapping were converted to BAM format files using samtools-1.6 (Li et al., 2009), and to files in BED format with bamToBed utility from BEDTools-2.26.0 (Quinlan and Hall, 2010).

Models of distribution of 5' end read fragments within the B2 loci (B2_Mm1a, B2_Mm1t and B2_Mm2) were performed using an in house python script. In brief, the script constructs a read accumulation metagene model around a hypothetical set of genomic points, in our case the start site for all B2 elements (TSS), in which the numbers of reads (or read 5' ends) around each different TSS were calculated and attributed to defined points in the model. B2 element coordinates are based on the UCSC genome browser RepeatMasker track (as of November 2018). To calculate the B2 processing ratio, Babraham NGS analysis suite Seqmonk 1.38.2

(<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>) was used to obtain number of long reads overlapping with B2 loci (B2_Mm1a, B2_Mm1t and B2_Mm2), as well as the number of reads overlapping with tRNA loci from -5 to 15 bp. Processing ratio for each sample was calculated by processed B2 count obtained from the in house python scripts normalized by tRNA from -5 to 15 bp and small reads fastq read count, as well as B2 count and long reads fastq read count: $[\text{Small fragments (position 95–110)} / [\text{tRNAs/small RNA fastq}]] / [\text{B2RNA/long RNA fastq}]$. The full length of B2 consensus sequence is 188nt, and this is the one we use for the in vitro experiments. However, structure of the RNA has been resolved only for the 155nt (Espinoza et al., 2007), and this is the structure currently used in our figures. For the mapping of short fragments, we have used the same range tested in our previous study (Zovoilis et al., 2016) to maintain consistency of the results. The reason why this 120nt threshold was selected in the Cell paper was to exclude artifacts from short RNAs mapping partially in our metagene as well as downstream of those B2 elements that are shorter from the consensus sequence.

In long-RNA-seq, FPKM (Fragments Per Kilobase of transcript per Million) and TPM (Transcripts Per Million) for genes were generated using StringTie-1.3.4d (Pertea et al., 2015) with the following annotation: ensembl GRCm38 patch 94 gff3 file, and parameters limiting the processing of read alignments to only estimate and output the assembled transcripts matching the reference transcripts given in annotation and excluding non-regular chromosomes. Because TPM already includes scaling of the data it is unsuitable for the averaging of the gene expression levels of multiple genes (B2-SRGs) used in the boxplots of Figure 2. This does not apply in case of single genes as in Figure 2C (Trp53) or in the heatmap of the same figure, where each gene is presented in a separate row, and for which TPM values are used. For data visualization, statistics and

differential expression analysis, we employed R (version 3.4.3) (<https://www.R-project.org/>) and the package DESeq2 (Love et al., 2014). Differential expression analysis was implemented on transcript count data for 6-month-old mice between APP and wild type. Boxplots central line represents median and t-test was applied on the group numbers mentioned in the text. PCA plots for samples used and read count correlation matrix between 6-month-old mice samples are presented in Figure 6—figure supplement 1.

For Hsf1 metagene analysis, we used peak.txt files of Hsf1 peaks for ChIP-seq from Mahat et al., 2016. Peaks were analyzed with Seqmonk around Transcription start sites of genes (TSS) based on the Eponine annotation (Down and Hubbard, 2002), and filtered based on their overlap with B2 RNA regulated genes (Zovoilis et al., 2016). Also, B2 RNA-binding (CHART-seq) peaks were analyzed with Seqmonk around TSS (Eponine), then filtered by overlapping with learning-associated SRGs or all genes (Peleg et al., 2010). Relative density metagene plots of the distribution of the above peaks were generated using Seqmonk.

2.5 Results

2.5.1 B2 RNA regulated SRGs (B2-SRGs) are enriched in neural functions

In a previous study, we have identified genomic locations that are subject to regulation by SINE B2 non-coding RNAs during response to thermal stress (heat shock) through binding and suppression of transcription by the RNA Pol II at the pre-stimulus state. Upon induction of cellular stress through the application of a stimulus, SRGs in these locations become activated through B2 processing and release of RNA Pol II suppression (Figure 1A; Zovoilis et al., 2016). A list of B2-SRGs at these locations is available in Supplementary file 1. Response to thermal stress (heat

shock) has been used for years as a basic study model of cellular response to stress. Proteins and gene pathways initially identified in heat shock have been subsequently shown to play identical pro-survival roles in other biological systems. Thus, we questioned whether there are other known cellular functions, beyond response to heat shock, that are connected with B2-SRGs.

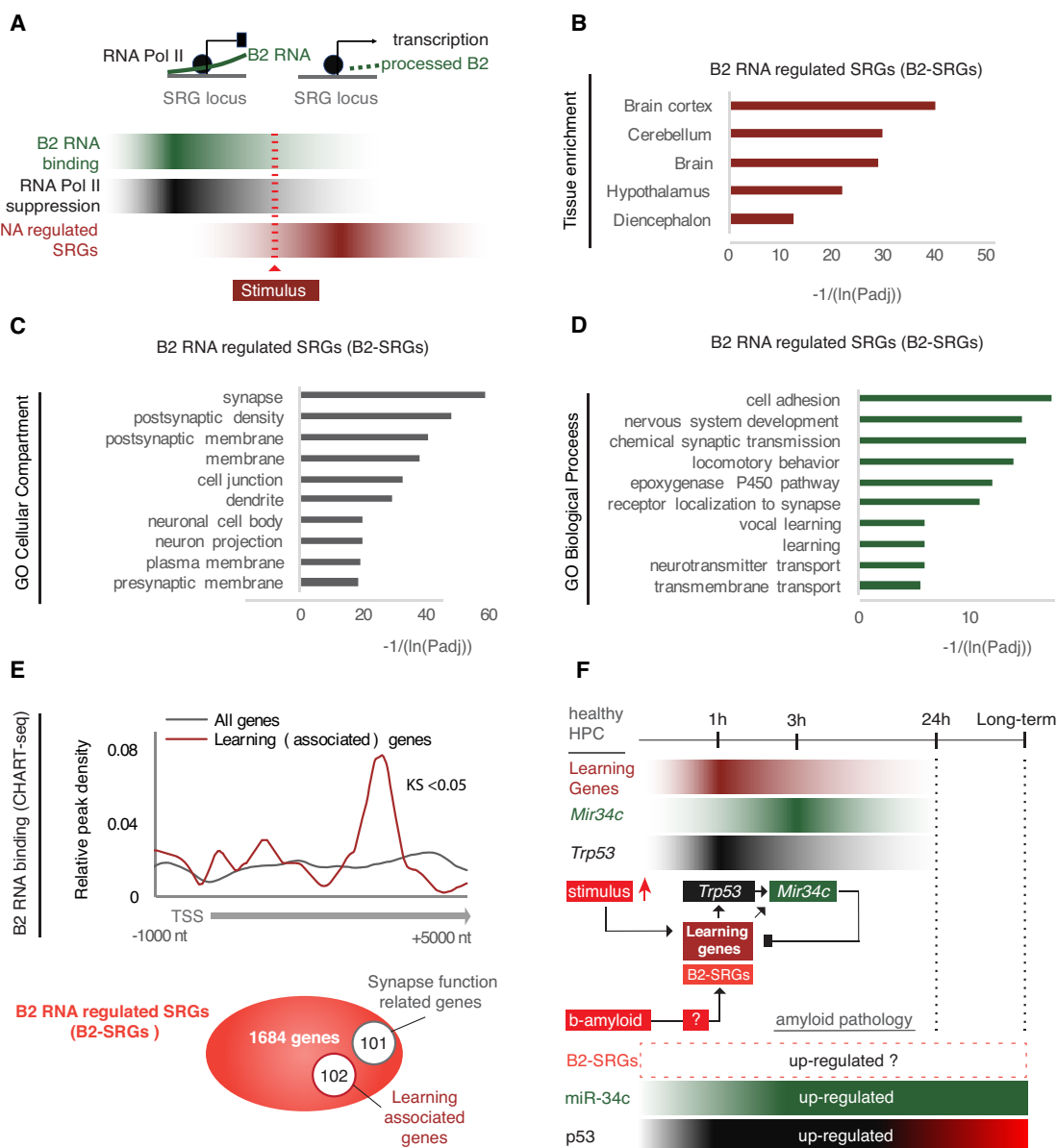


Figure 1. B2-SRGs are enriched in neural functions.

(A) Regulation mode of SRGs by B2 RNA processing based on previous works (Zovoilis et al., 2016; Yakovchuk et al., 2009; Ponicsan et al., 2010). Color intensity represents higher B2 RNA binding (green), Pol II suppression (black) or SRG expression (red). (B) Tissue enrichment analysis of B2-SRGs listed in Supplementary file 1 based on the DAVID functional annotation

platform (Huang et al., 2009a; Huang et al., 2009b). The adjusted p-values of top five ranking terms are plotted as a function with higher scores representing lower adjusted p-values (higher statistical significance) ranging from $\text{padj } 1.92\text{E-}04$ for ‘Diencephalon’ to $\text{padj } 8.91\text{E-}13$ for ‘Brain Cortex’. A complete list of all terms that pass the EASE 0.05 score threshold is presented in Supplementary file 1. (C) Gene ontology (GO) analysis (DAVID) of B2 RNA regulated genes on the basis of cellular compartments. The adjusted p-values of top 10 ranking terms are plotted as a function with higher scores representing lower adjusted p-values ranging from $2.23\text{E-}06$ for ‘Presynaptic membrane’ to $1.88\text{E-}18$ for ‘Synapse’. A complete list of all terms that pass the EASE 0.05 score threshold is presented in Supplementary file 1. (D) GO analysis (DAVID) of B2 RNA regulated genes on the basis of biological processes. As above, adjusted p-values of top ranking terms are plotted as a function, with higher scores representing lower adjusted p-values ranging from 0.02 for ‘Transmembrane transport’ to $5.44\text{E-}06$ for ‘Cell adhesion’. A complete list of all terms that pass the EASE 0.05 score threshold is presented in Supplementary file 1. (E) Upper panel: Metagene analysis of distribution of genomic B2 RNA-binding sites across the start of genes from Zovoilis et al., 2016, comparing learning-associated genes from Peleg et al., 2010 with all genes (Kolmogorov Smirnov test, $\text{KS} < 0.05$). TSS represents the Transcription Start Site of these genes. Lower panel: Representation of the number of B2-SRGs that are (i) learning associated based on Peleg et al., 2010 (Supplementary file 1) and (ii) synapse function related based on GO term enrichment (Supplementary file 1, right panel). (F) Regulatory loop of learning genes-Trp53-Mir34c in mouse hippocampus based on Peleg et al., 2010; Zovoilis et al., 2011; Yamakuchi and Lowenstein, 2009. A potential role of B2-SRGs, that include many learning genes, in affecting this loop in amyloid pathology remained unknown after these studies (noted with a question mark in the figure).

As shown in Figure 1B, after performing a tissue enrichment analysis to identify tissue terms that are over-represented in the list of our SRGs, we found a significant enrichment of neural tissue terms compared to other tissues in our list. Similarly, during Gene Ontology term enrichment analysis, cellular compartments closely related to neural functions top the list of enriched terms in these genes, including among the first 10 entries terms such as synapse, postsynaptic density and membrane, dendrites, neural projections, and pre-synaptic membrane (Figure 1C, Supplementary file 1). Most importantly, after performing the same analysis for Biological Processes GO terms, B2-SRGs were found to be enriched considerably in neural-function-related terms. GO terms such as learning, nervous system development, synaptic transmission, synapse receptor localization, and neurotransmitter transport were among the first 10 entries with the highest adjusted p value scores enriched in B2-SRGs (Figure 1D, Supplementary file 1).

Among the biological processes potentially affected by B2 RNAs that were identified above was learning. We and others have already shown that learning processes in the mouse hippocampus are connected with the transient activation of a number of learning-associated genes and pathways, including many known SRGs (Peleg et al., 2010). For this reason, we examined whether any learning-associated genes are among the binding targets of B2 RNAs. To this end, we compared the distribution of B2 RNA binding sites in the genome between learning-associated genes and all genes. Indeed, as shown in Figure 1E, B2 RNA-binding sites were found to be enriched in learning-associated genes compared to other genes. In particular, based on learning-associated genes previously identified in mouse hippocampus (Peleg et al., 2010), among the B2-SRGs (1684 genes), 102 genes (Figure 1E, Supplementary file 1) are associated with learning. In addition, biological process terms enriched in B2-SRGs included, among others, pathways implicated with response to cellular stress in neural cells, and various genes implicated with synaptic function (Figure 1E, Supplementary file 1).

This data suggests that B2-SRGs could potentially affect a wide spectrum of functions beyond heat shock including response to cellular stress in neural cells, synaptic function, and learning. All these biological processes are heavily impaired in AD. Given the role of hippocampus as a primary target of amyloid pathology in AD, we decided to expand our investigation towards the potential role of B2-SRGs regulation in this pathological process. We have previously shown that in hippocampus the pro-apoptotic miRNA Mir34c, which is induced among others by Trp53 and controls it in a negative feedback loop manner (Yamakuchi and Lowenstein, 2009), targets many learning-associated genes in order to help to restore their expression and Trp53 expression levels after

application of a stimulus (Figure 1F; Zovoilis et al., 2011; Yamakuchi and Lowenstein, 2009). In both mouse models of amyloid beta pathology and AD patient brains, we found persistently high levels of Mir34c. High Mir34c levels in hippocampi with amyloid pathology are indicative of a transcriptome-wide deregulation of Trp53 and associated genes, as many of these genes are either direct or indirect upstream regulators of Mir34c (Figure 1F; Rokavec et al., 2014). Since, as shown above, B2-SRGs include many learning genes as well as other genes involved in neural function, we questioned whether such transcriptome changes in response to amyloid pathology involve B2-SRGs.

2.5.3 A number of B2-SRGs get hyper-activated during the neurodegeneration phase of amyloid pathology

To test whether B2-SRGs are indeed deregulated in amyloid pathology, we employed a transgenic mouse model of amyloid pathology, APPNL-G-F (Saito et al., 2014), and the respective wild-type control (C57BL/6J). The same animal cohorts that were previously characterized through a battery of immunohistochemistry (IHC) and behavioral tests (Mehla et al., 2019) were used to isolate whole hippocampi for the transcriptome analysis conducted in the current study. Figure 2A depicts our experimental design while Figure 2B depicts the amyloid plaque deposition in the brains. The behavioral tests in these mouse cohorts are presented in our previous study (Mehla et al., 2019).

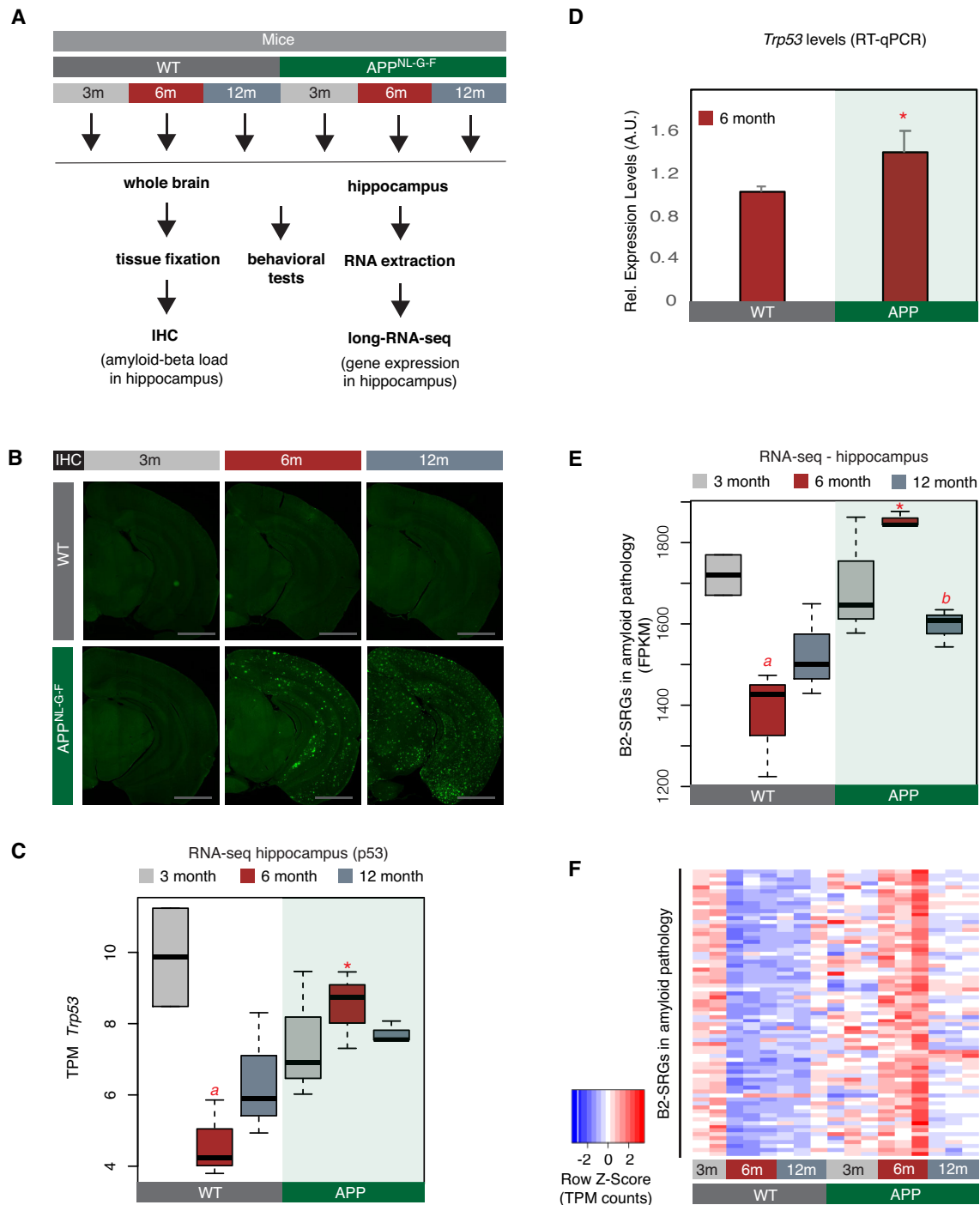


Figure 2. A number of B2-SRGs are hyper-activated in amyloid pathology.

(A) Experimental design for study of B2-SRGs in the hippocampus of the amyloid pathology mouse model (APP) and the respective wild type (WT) control. (B) Immunohistochemistry for identifying amyloid-beta load in the brain sections of mice from the same cohort as in (A) and our previous study (Mehla et al., 2019). Higher fluorescence intensity corresponds to higher amyloid load in APP mice from 6-month-old onwards. (C) Expression levels of the pro-apoptotic gene Trp53 (official symbol Trp53) as defined by long-RNA-seq. Trp53 transcripts per million reads

(TPM) from the APP mice compared to WT, grouped by age. Boxplot depicts distribution of expression levels of Trp53 gene among different age groups of mice (WT: wild type, APP: mice with amyloid pathology). Black line denotes median. Statistical significance (p value threshold 0.05) for the comparison between 6-month-old APP and 6-month-old WT ($p=0.01$) (depicted as asterisk, unpaired, non-directional t-test, $n = 3/\text{group}$) and the comparison between 3-month-old WT ($n = 2$) and 6-month-old WT ($n = 3$) ($p=0.04$) (depicted as a). No significance for the rest of the comparisons between APP and WT in the other two age groups, or between other different ages within the same group (APP or WT). (D) Expression levels of the pro-apoptotic gene Trp53 (official symbol Trp53) in 6-month-old mice as defined by RT-qPCR. Statistical significance (p value threshold 0.05) for 6-month-old APP greater than 6-month-old WT (depicted as asterisk, unpaired directional t-test, $n = 3/\text{group}$, error bars represent standard deviation from the mean). (E) Expression levels as defined by long-RNA-seq of B2-SRGs that are upregulated in amyloid pathology (72 genes, Supplementary file 1). Boxplot depicts distribution of expression levels (in FPKM/Fragments per Kilobase per Million) of B2-SRGs among different age groups of wild type and APP mice. Statistical significance (p value threshold 0.05) for the comparison between 6-month-old APP and 6-month-old WT ($p=0.02$) (depicted as asterisk, unpaired, non-directional t-test, $n = 3/\text{group}$), for the comparison between 3-month-old WT ($n = 2$) and 6-month-old WT ($n = 3$) ($p=0.03$) (depicted as a) and for the comparison between 6-month-old APP and 12-month-old APP ($p=0.005$) (depicted as b). No significance for the rest of the comparisons between APP and WT in the other two age groups, or between other different ages within the same group (APP or WT). (F) Gene expression levels of B2-SRGs that are upregulated in amyloid pathology (72 genes, Supplementary file 1) show strong association with amyloid pathology status in the hippocampus of APP and WT mice of different ages. Heatmap depicts gene expression with rows representing B2 RNA regulated genes and columns representing mouse samples. Expression values are normalized per row and correspond to TPM values. Red color represents higher expression.

We have focused on three different mouse ages that correspond to different phases of amyloid beta pathology and represent the: (i) pre-symptomatic stage with undetectable (very low) amyloid plaque load (3 months - 3 m, Figure 2B left panels), (ii) stage of symptom manifestation (6 months - 6 m) that coincides with the active neurodegeneration phase and appearance of amyloid plaques (Figure 2B, middle-panels), and (iii) terminal stage of the pathology (12 months - 12 m, Figure 2B right panels) when mice have already acquired the extensive brain atrophy due to neural cell death. Whole hippocampi from mice of these three groups were isolated and the extracted RNA was subjected to next-generation sequencing. We performed directional RNA sequencing for these samples and subsequently quantified gene expression levels. Sequenced samples depicted high

expression levels for known hippocampal markers such as *Gad1*, *Gad2*, *Slc17a7*, and *Mbp* (Figure 2—figure supplement 1).

In accordance with increased cell death during the active neurodegeneration phase at 6 months, levels of *Trp53* were elevated in APP mice of this age compared to controls of the same age (t-test, $p < 0.05$, $n = 3/\text{group}$) (Figure 2C). Differential expression analysis at this time point revealed a number of genes that are upregulated in 6-month-old APP mice (Supplementary file 1), including 72 B2-SRGs among them (Supplementary file 1), of which 13 are learning associated (Supplementary file 1). These 72 B2-SRGs were enriched in neural-function-related terms, such as neural development, learning, synapse function, and calcium signaling (Supplementary file 1). Consistent with a hypothesis of a transcriptome-wide deregulation that involves B2-SRGs, levels of the 72 B2-SRGs were found to be strongly upregulated during the active phase of neurodegeneration at 6 months (Figure 2D–F). In particular, in healthy hippocampi (WT control animals), levels of these genes are normally downregulated in 6-month and 12-month-old mice compared to 3-month-old mice ($p = 0.02$ and $p = 0.05$, respectively) (Figure 2E). This is in accordance with the higher level of neural synaptic activity and plasticity that younger mice have (Lilja et al., 2013). In contrast, levels of B2-SRGs in APP mice remain abnormally high in 6-month-old APP mice compared to the 6-month-old control mice ($p < 0.05$) (Figure 2E and F). A validation of our RNA-seq data through RT-qPCR for selected genes that we identified by RNA-seq to be upregulated in 6-month-old APP mice, is presented in Figure 2—figure supplement 2. In Figure 2, we only tested the expression of those B2-SRGs that overlapped with the upregulated genes revealed by DESeq in 6-month APP mice. When expression dynamics for all B2-SRGs are tested (Figure 2—figure supplement 3), these are similar to the ones observed for genes in Figure

2, while such dynamics are not observed when a random set of non-B2 RNA regulated genes is tested (Figure 2—figure supplement 3). Thus, as explained in Figure 2—figure supplement 3, it cannot be excluded that our findings may extend to additional B2-SRGs beyond the 72 identified through our differential gene expression analysis.

These findings show a transcriptome wide hyper-activation of certain B2-SRGs during the active neurodegeneration phase of amyloid pathology.

2.5.4 B2 RNA processing ratio increases during the neurodegeneration phase of in vivo amyloid pathology

Given the role of B2 RNAs in the regulation of SRGs, we hypothesized that the observed hyperactivation of B2-SRGs in the hippocampus of 6-month-old APP mice may reflect a similar upstream deregulation at the level of the B2 RNAs.

To test this, we employed a customized version of RNA sequencing and analysis used in our previous study (Zovoilis et al., 2016) that allows for enrichment and sequencing of the short SINE RNA fragments (<100 nt) produced by B2 RNA processing (short-RNA-seq). In contrast to standard long-RNA-seq protocols that include RNA fragmentation and, thus, may introduce bias, this approach circumvents this problem. Moreover, the long-RNA-seq protocols exclude short RNA fragments of <100 nt making the identification of B2 short fragments challenging. After short-RNA-seq (Figure 3A), mapping of the 5' ends of the sequenced fragments across the B2 loci enables the determination of processing points at B2 RNA (depicted as 'X' in Figure 3A), including those at the critical RNA Pol II binding region (depicted as a rectangle in Figure 3A).

Figure 3B depicts an example of these fragments from one of the samples. As shown in Figure 3A, B2 RNA is extensively processed in hippocampi with amyloid pathology. In particular, this data revealed a increased number of B2 RNA fragments in 6-month-old mice compared to controls (Figure 3A), a difference that cannot be explained by full-length B2 RNA levels between these groups of mice using long-RNA-seq (Figure 3C).

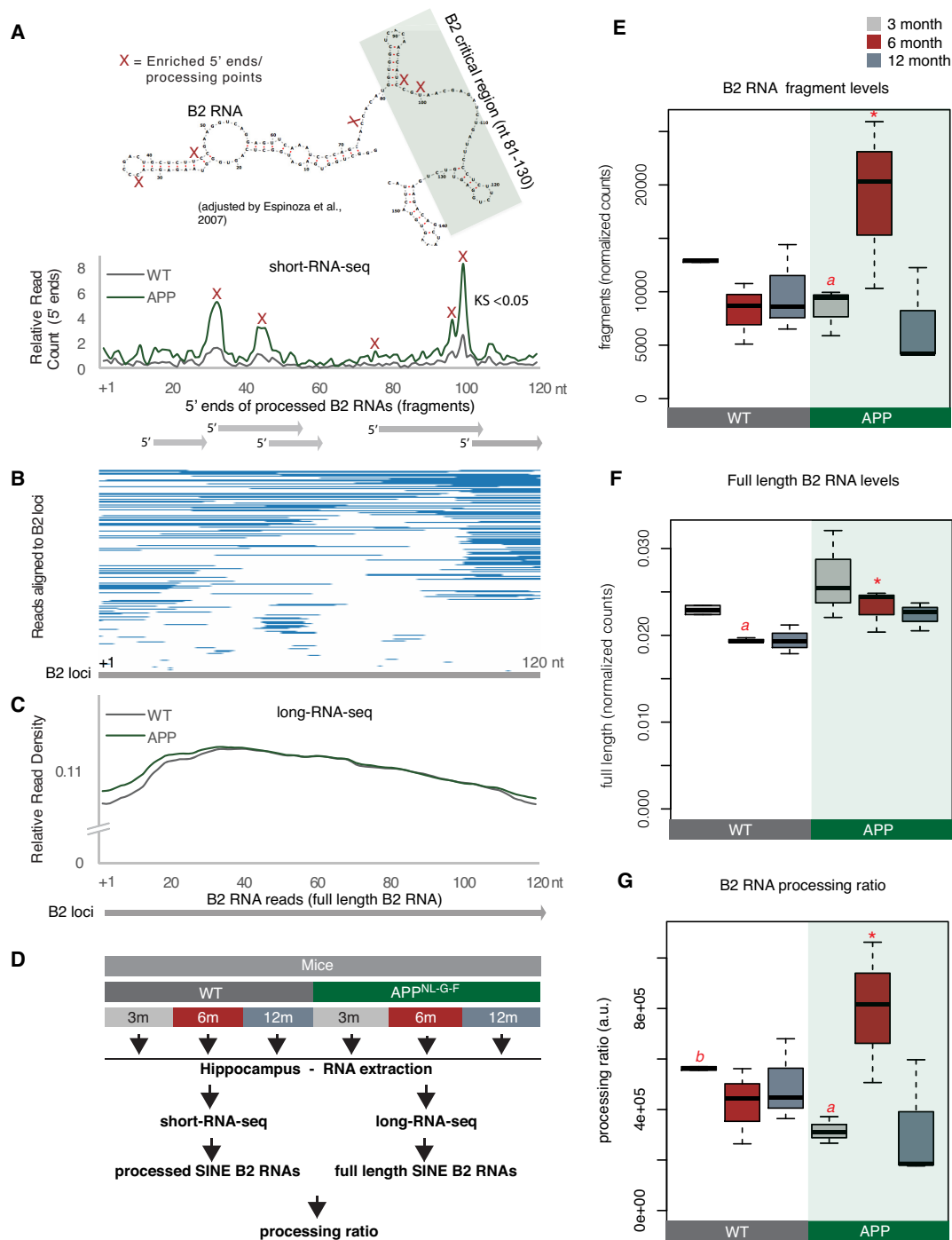


Figure 3. B2 RNA processing ratio is increased in 6-month-old APP mice.

(A) Plotting of the position of the first base (5' end) of B2 RNA fragments across the B2 loci to depict increased levels of B2 RNA fragments in 6-month-old APP mice. Upper panel: Secondary structure and processing points of B2 RNA. Secondary structure of B2 RNA adapted from Espinosa and colleagues (Espinosa et al., 2007). As in our previous study (Zovoilis et al., 2016), we depict the B2 RNA processing points based on short RNA-seq data and mapping of the 5' ends of B2 RNA fragments. X marks which cleavage sites of B2 RNA in the upper panel correspond to

enriched processing points (the peaks of 5' end fragments distribution) at the lower panel. The green rectangle depicts the critical region that binds and suppresses RNA Pol II (Yakovchuk et al., 2009; Espinoza et al., 2007; Ponicsan et al., 2010; Ponicsan et al., 2015) that may be affected by such processing points. Lower panel: Distribution of the 5' ends of B2 RNA fragments across the B2 RNA loci based on mapped short RNA-seq from the hippocampi of 6-month-old APP and WT mice. The x axis represents a metagene combining all B2 RNA loci aligned at the start site of their consensus sequence (position +1) and the y axis shows the relative 5' end count for B2 RNA fragments aligning to any position downstream of position +1. Figure depicts the difference between B2 RNA fragment distribution of 6-month-old APP mice (higher peaks) vs 6-month-old WT mice (lower peaks) (Kolmogorov Smirnov test, $KS < 0.05$). (B) Alignment of short-RNA seq reads mapped against the multiple B2 loci. Only B2 loci with at least one read are depicted. Short RNA-seq reads from one of the 6-month-old APP mice, whose 5' ends are plotted at (A) lower panel are used here as an example of the length and position of fragments across B2 loci. (C) Metagene analysis of long-RNA-seq reads of mice across B2 RNA loci that is enriched in full-length B2 RNAs but not in B2 RNA fragments. Comparative analysis shows low difference in relative read density of full-length B2 RNAs between WT and 6-month-old APP. X axis as in A. Y axis corresponds to read coverage across the B2 loci. (D) Experimental design for estimation of B2 RNA processing ratio based on short and long-RNA seq data for all APP mice and controls. (E) Boxplot depicts distribution of levels of processed SINE B2 RNA fragments among different age groups of mice between wild type and APP. Statistical significance (p value threshold 0.05) only for 6-month-old APP greater than 3-month-old APP ($p=0.04$), 12-month-old APP ($p=0.04$) and 6-month-old WT ($p=0.04$) (depicted as asterisk, $n = 3/\text{group}$, unpaired directional t-test). All the other comparisons between groups and different ages non-significant except 3-month-old WT less than 3-month-old APP ($p=0.03$) (depicted as a). (F) Boxplot depicts distribution of levels of full-length SINE B2 RNAs among different age groups of mice between wild type and APP. Statistical significance (p value threshold 0.05) for the comparison between 6-month-old APP and 6-month-old WT ($p=0.05$; depicted as asterisk, $n = 3/\text{group}$, non-directional unpaired t-test) and for the comparison between 3-month-old WT ($n = 2$) and 6-month-old WT ($n = 3$) ($p=0.004$; depicted as a). All the other comparisons between groups and different ages are non-significant. (G) Boxplot depicts distribution of SINE B2 RNA processing ratio among different age groups of mice between wild type and APP. Statistical significance (p value threshold 0.05) for the comparison between 3-month APP and 6-month-old APP ($p=0.04$; depicted as a, $n = 3/\text{group}$, non-directional unpaired t-test), for the comparison between 3-month-old WT ($n = 2$) and 3-month-old APP ($n = 3$) ($p=0.008$; depicted as b), and for 6-month-old APP greater than 6-month-old WT ($p=0.05$; depicted as asterisk, $n = 3/\text{group}$). All the other comparisons between groups and different ages are non-significant.

This data suggests that the processing ratio of B2 RNAs may be higher in APP mice of this age.

To estimate this ratio, we used the short RNA-seq data in combination with standard long-RNA-seq. As in our previous study, for normalization of short RNA 5' end values we used a class of

short RNAs that is not affected by B2 RNAs, the RNA Pol III-transcribed tRNAs and estimated the levels of B2 RNA processing fragments (Figure 3E). Moreover, the absolute numbers of fragmented B2 RNA may vary according to the underlying expression of full-length B2 RNA transcripts. Therefore, in order to factor in any differences in the amount of fragments due to basal expression levels of the full-length B2 RNA, full-length B2 RNA levels were calculated by the directional long-RNA-seq (Figure 3F). The long-RNA-seq approach excludes short fragments, and subsequently, B2 RNA fragment values from short RNA-seq were normalized to the levels of the full-length B2 RNAs to calculate the processing ratio. Consistent with our hypothesis and Figure 3A findings, 6-month-old APP mice were found to have substantially increased ratio of B2 RNA processing compared to control mice of the same age (Figure 3G) ($p < 0.05$, $n = 3/\text{group}$). This increase was observed only in the 6-month-old mice (Figure 3G, Figure 3—figure supplement 2) and coincides with the increase in Trp53 levels and B2-SRGs levels observed in this age.

Thus, APP mice at the active neurodegeneration phase are characterized by higher destabilization and processing ratio of B2 RNAs, consistent with the observed increase in B2-SRG levels in these same animals.

2.5.5 Hsf1 accelerates B2 RNA processing

We then focused on the molecular mechanism underlying the increased B2 RNA processing during response to amyloid toxicity. When we had previously examined the mechanism of B2 RNA processing in non-neural cells (NIH/3T3 cells during heat shock), a member of the PRC2 protein complex, Ezh2, was reported as being responsible for the B2-RNA-accelerated destabilization and processing during response to stress (Zovoilis et al., 2016). However, as shown in Figure 4—figure

supplement 1A–B, scant expression levels of Ezh2 levels in neural cells indicate that Ezh2 is not a key factor in B2 RNA processing in brain. Thus, the factors that may mediate destabilization for B2 RNAs during stress in neural cells remain elusive.

In our earlier study that described the induction of B2 RNA processing by Ezh2, it remained unclear how Ezh2 exerted its impact on B2 RNAs, since Ezh2 lacked any known RNase activity (Zovoilis et al., 2016). However, in subsequent experiments (Hernandez et al., 2020), we showed that instability is in fact inherent to the B2 RNA molecule while interaction with Ezh2 only accelerates this destabilization and Ezh2 does not cleave B2 RNA by itself. This finding suggests that other proteins may have a similar effect on B2 RNA stability. Therefore, we started searching for stress-related candidate proteins that could affect the B2 RNA processing.

We showed before that, during response to stress in NIH/3T3 cells, B2 RNA binding is enriched near stalled RNA polymerase genomic sites. These areas are known to be highly enriched in binding sites of various stress related proteins, among which Hsf1, a master regulator of stress response for various types of cellular stress (Pandey et al., 2011). Hsf1 has been previously connected with activation of SRGs through both transcriptional factor (TF) activities as well as other yet unknown TF-independent processes (Inouye et al., 2007). Interestingly, when we examined the proximity of Hsf1 binding sites, identified by the Lis lab (Mahat et al., 2016), to genes with B2 RNA binding sites (Supplementary file 1; Zovoilis et al., 2016), we found that increased number of Hsf1-binding sites were found near B2-SRGs (Figure 4—figure supplement 1C) and were further enriched after application of a heat-shock stimulus (KS-test <0.05). Moreover, as shown in Figure 4—figure supplement 1A–B, in contrast to Ezh2, Hsf1 is expressed

in neural tissues and especially in hippocampus. These findings provided an indication that Hsf1 may be implicated in B2 RNA biology and that it is also expressed in neural cells, which urged us to test further its expression pattern in our biological context. Hsf1 levels were found to be upregulated in APP 6-month-old mice compared to control group of the same age ($p < 0.05$, $n = 3/\text{group}$) (Figure 4A) in the RNA-seq data, a result confirmed also through RT-qPCR (Figure 4B). As mentioned above, at the same time, 6-month-old APP mice have increased B2 RNA processing ratio. Thus, this data suggests that Hsf1 may be a good candidate for accelerating B2 RNA processing in the context of amyloid pathology.

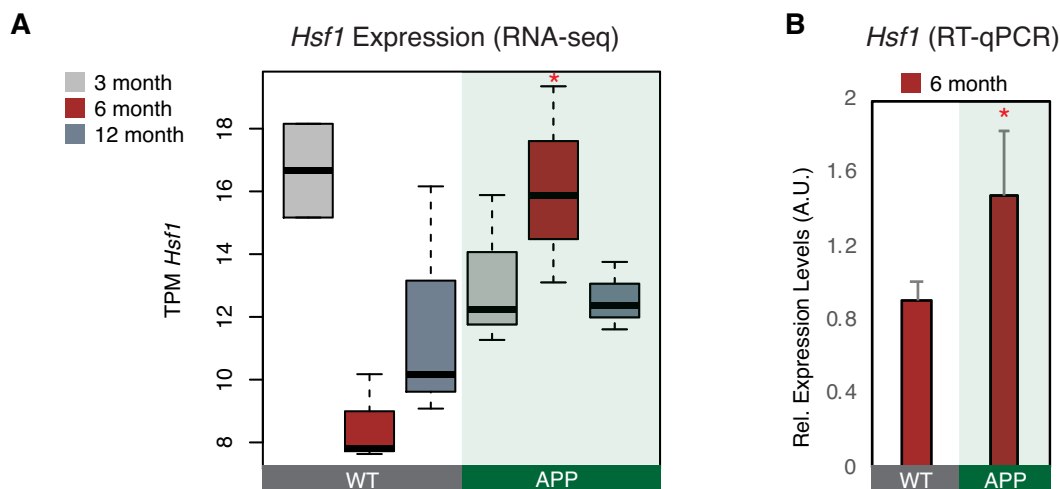


Figure 4. Hsf1 is upregulated in 6-month-old APP mice.

(A) Boxplot depicts distribution of expression levels of Hsf1 gene among different age groups of mice between wild type and APP. Values are based on TPM counts of long-RNA-seq data. Statistical significance (p value threshold 0.05) for the comparison between 6-month-old APP and 6-month-old WT ($p = 0.03$) ($n = 3/\text{group}$, unpaired non-directional t-test). All the other comparisons between groups and different ages are non-significant. (B) Expression levels of Hsf1 in 6-month-old mice as defined by RT-qPCR. Statistical significance (p value threshold 0.05) for 6-month-old APP greater than 6-month-old WT (depicted as asterisk, unpaired directional t-test, $n = 3/\text{group}$, error bars represent standard deviation from the mean).

To check this, we investigated whether the interaction between B2 RNA and Hsf1 can accelerate B2 RNA destabilization. In particular, we incubated full-length B2 RNA with the Hsf1 protein in

vitro. As shown in Figure 5A, in the presence of Hsf1, the destabilization of full-length B2 RNA was accelerated in contrast to the control protein (PNK) or in the absence of protein (no protein). No processing was observed in an RNA fragment co-incubated with B2 RNA and the protein (marked with an asterisk across Figure 5). We then questioned whether the fragments observed during B2 RNA processing in vitro are the same with those we observed in NIH/3T3 cells in our previous study and hippocampal cells in our current study (Figure 3A). To test this, we subjected the in vitro processed B2 RNA fragments to short RNA-sequencing as with the in vivo samples and compared the processing patterns between the two cases. As shown in Figure 5—figure supplement 1A, the in vitro processing points were the same for most of the positions identified in vivo including the most prominent ones around 99 and 33. The two processing patterns differed only at positions 90 and 47, suggesting that in vivo B2 RNAs may be protected or processed in these two positions by yet unidentified mechanisms.

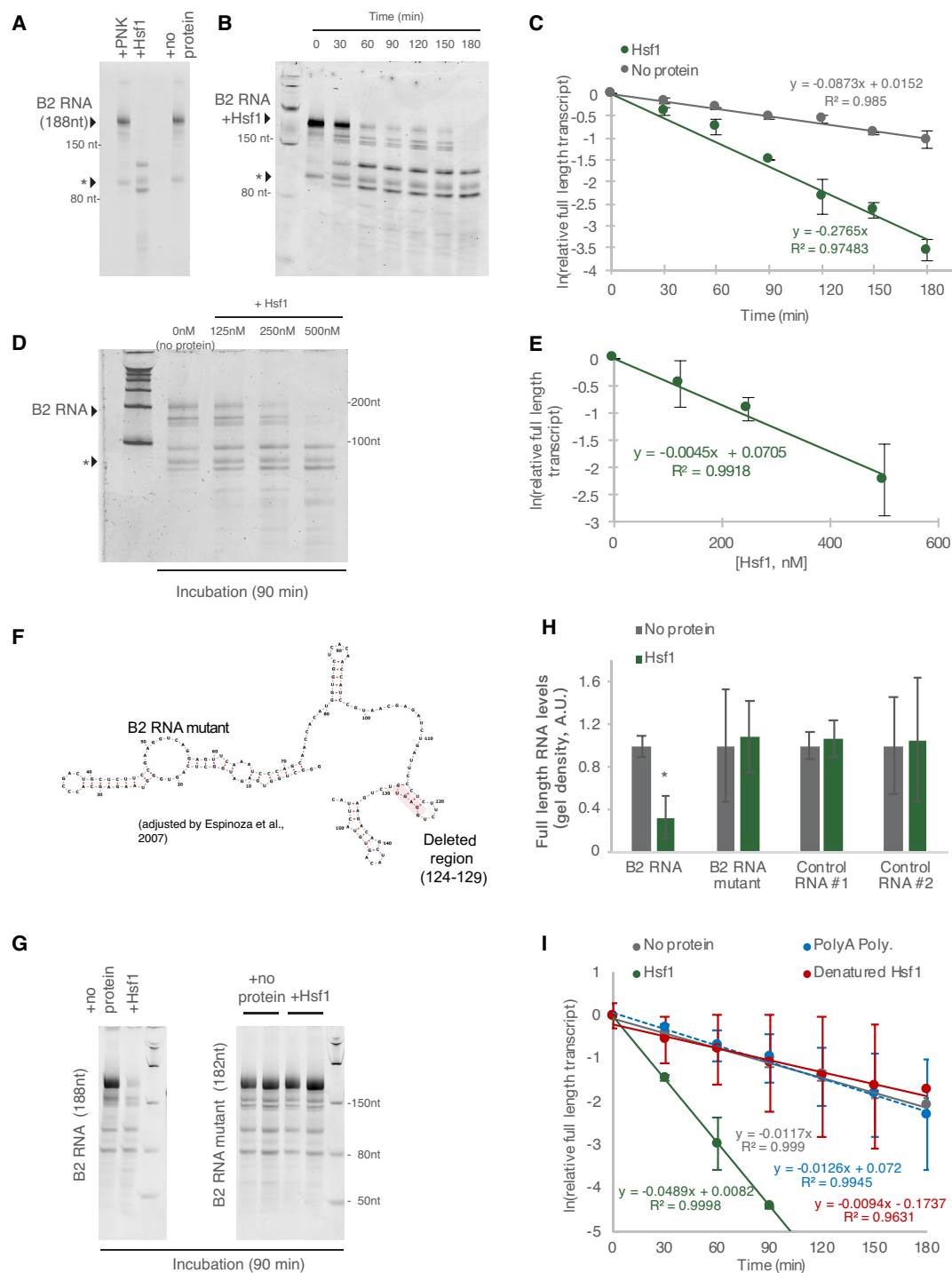


Figure 5. Hsf1 accelerates B2 RNA processing.

(A) In vitro incubation of B2 RNA. In vitro transcribed and folded B2 RNA at 200 nM incubated with PNK as a control (lane 1), 250 nM Hsf1 (lane 2) and without protein (lane 4). Incubations occurred for 6 hr at 37°C. (B) In vitro incubation of B2 RNA for different incubation periods. In vitro transcribed B2 RNA (100 nM) incubated at 37°C with 500 nM Hsf1 in the course of 180 min with time intervals of 30 min. (C) Relative full-length RNA remaining from (B) using ImageJ area

under the curve software over time. The full gels are available as source files. (D) Titration of Hsf1 protein in incubation with 200 nM in vitro transcribed B2 RNA. Concentrations of Hsf1 range from 0 to 500 nM (0 means incubation was performed in the absence of protein and only in the presence of TAP buffer, for the same duration as the other three conditions). (E) Relative full-length RNA remaining from (D) using ImageJ area under the curve software over time. The trendline is a linear fit, displaying standard deviation on the data points. The full gels are available as source files. (F) Graphical representation of the deletion in the B2 RNA mutant on the secondary structure of B2 RNA as adapted from Espinoza et al., 2007. (G) In vitro incubation of the B2 RNA and the B2 RNA mutant for 90 min at 37°C with either Hsf1 or no protein (just TAP buffer, see Materials and methods). (H) Comparison among B2 RNA, the B2 RNA mutant (6nt deletion) and two control RNAs regarding the full-length RNA levels remaining after in vitro incubation for 90 min at 37°C with Hsf1. Sizes of control RNAs are control for RNA #1, 143nt and for control RNA #2, 163nt. or no protein (just TAP buffer, see Materials and methods). Incubation in the absence of Hsf1 but presence of the same buffer (TAP) was used as control to take into account any non-Hsf1-specific RNA destabilization due to non-specific degradation. Asterisk represents statistical significance $p=0.005$ (unpaired, non-directional t-test) for the comparison between Hsf1 and no protein incubation (three replicates for B2 RNA). The full gels are available as source files. (I) Comparison among Hsf1 (~60 KDa), denatured Hsf1, Poly A polymerase (~56 KDa) and no protein (just TAP buffer) with regard to B2 RNA processing (estimating relative full-length RNA remaining as in C). The full gels are available as source files. For Hsf1, after 90 min no detectable full-length B2 RNA could be measured anymore and no further data points were included in the linear model.

Figure 5B shows B2 RNA destabilization in time. B2 RNA destabilization in the presence of Hsf1 was accelerated compared to no protein (Figure 5C and Figure 5—source datas 1 and 2). Consistent with the above finding, the rate of processing of B2 RNA was dependent on Hsf1 concentration in the reaction (Figure 5D and E) and increased upon increase of Hsf1 concentration. In Figure 5D and our analysis in Figure 5E, as a zero concentration point (no Hsf1 protein) we used RNA incubated in the same buffer and for the same time as the samples with the Hsf1 protein to take into account any RNA destabilization due to non-specific degradation, hydrolysis or B2 RNA endogenous self-cleavage. In order to confirm further the Hsf1-mediated acceleration of B2 RNA we transcribed a mutant B2 RNA that harbors a 6nt deletion in the hairpin at position 124 (Figure 5F). Then, we compared B2 RNA processing during incubation in the presence and absence of Hsf1 and observed an impairment in the ability of Hsf1 to accelerate B2 RNA processing (Figure

5G–H). Similar negative results were obtained when other control RNAs were tested (Figure 5H, control RNA #1, 143nt; control RNA #2, 163nt). Subsequently, we tested the impact on acceleration of B2 RNA processing by Hsf1 in the case of heat denaturation of Hsf1. As shown in Figure 5I, heat denaturation of Hsf1 resulted in abrogation of its ability to accelerate B2 RNA processing (Figure 5I). Similar results were obtained with the incubation of B2 RNA with another RNA binding protein of similar size to Hsf1 (Hsf1 ~60 KDa, Poly A polymerase ~56 KDa) (Figure 5I).

This data shows that Hsf1 has the potential to accelerate B2 RNA processing in vitro.

2.5.6 Hsf1 mediates increased B2 RNA processing in response to amyloid beta toxicity

The increased levels of Hsf1/B2 RNA processing in APP 6-month-old mice raised the question whether response to amyloid toxicity in hippocampal cells is connected with an Hsf1-mediated increase in B2 RNA processing. In order to test this, we employed a hippocampal cell culture model using the a HT-22 cell line, which has been used extensively in the past as hippocampal cell stress model (Davis and Maher, 1994; Liu et al., 2009). We incubated these cells with amyloid beta peptides (1–42 aa) and compared their transcriptome to the one of cells incubated with an inverted sequence control peptide (R, reverse 42–1) (Figure 6A). Incubation of these cells with 1–42 amyloid beta peptides results in upregulation of a number of genes (Supplementary file 1), 25 of which are also found upregulated in amyloid beta pathology (6-month-old APP mice) (Supplementary file 1). The increase in expression levels of genes associated with amyloid pathology (Figure 6B and C) suggests that this cellular model simulates to a certain extend the amyloid toxicity effect on the transcriptome of hippocampal cells and their response to cellular

stress. However, as any cell culture system, it has limitations regarding tissue level functions such as learning.

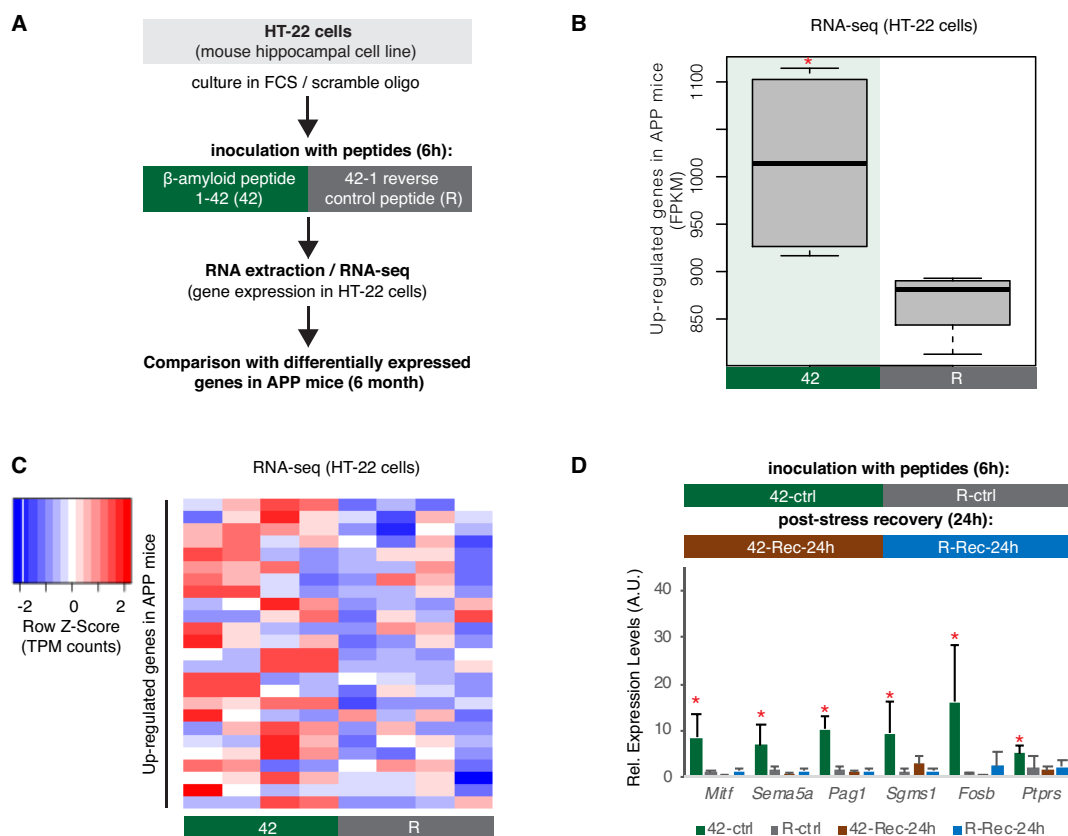


Figure 6. A hippocampal cell culture assay for tracking effects of amyloid beta toxicity on B2 RNA stability.

(A) Experimental design for the amyloid toxicity cell culture assay employing HT-22 cells. Cells culture media were supplemented with Fetal Calf Serum (FCS) and the scramble LNA described in Figure 8 to allow comparison with the LNA experiments. (B) Expression levels as defined by long-RNA-seq of genes upregulated in amyloid pathology (6-month-old APP mice) that are also upregulated in amyloid toxicity (25 genes, Supplementary file 1). Boxplot depicts distribution of expression levels (in FPKM/Fragments per Kilobase per Million) of these genes between the two conditions in HT-22 cells. $p=0.05$ for the comparison between HT22 samples transfected with 1–42 (Espinoza et al., 2007) vs 42–1 reverse control peptide (R) (depicted as asterisk, unpaired non-directional t-test, $n = 4/\text{group}$). (C) Heat map showing gene expression changes during incubation with amyloid peptides in HT22 cells for genes that are also upregulated in amyloid pathology (25 genes, Supplementary file 1) to test whether our cell culture system can simulate the transcriptome observed in amyloid pathology in vivo. Heatmap depicts increase in expression values calculated by long-RNA-seq (TPM values) in the HT22 samples transfected with 1–42 (Espinoza et al., 2007) vs 42–1 reverse control peptide (R) (heatmap columns), for these genes (Supplementary file 1) (heatmap rows). TPM values are normalized per row. Red color represents higher expression. (D) Upper panel: Experimental design for the amyloid toxicity assay and a control experiment testing recovery 24 hr after application of the stress stimulus to confirm return of gene expression levels to pre-stress levels. Lower panel: Confirmation through RT-qPCR of the expression differences of

selected B2-SRGs identified in RNA-seq data to differ between HT22 cells transfected with 1–42 (Espinoza et al., 2007) and cells transfected with the 42–1 reverse control peptide (R) for 6 hr (Supplementary file 1). Statistical significance (p value threshold 0.05) is depicted as asterisk for expression in treated cells (Espinoza et al., 2007) greater than the expression in controls (R) (unpaired directional t-test, $n = 4$, error bars represent standard deviation from the mean). In the same graph, we include also expression levels of cells from the same experiment that were left to recover in standard medium without any peptides for 24 hr after the initial 6 hr application of the peptides (42-Rec and R-Rec). Expression levels in all genes tested returned to pre-stress levels. Asterisk depicts statistical significance as described above also for the comparison between 42-ctrl and 42-Rec-24h.

Genes upregulated in our amyloid toxicity model included 25 B2-SRGs (Supplementary file 1). When testing for enriched terms in these 25 genes, biological processes related to apoptosis, such as regulation of apoptotic process and programmed cell death were at the top of the list (Supplementary file 1) and included, among others, genes such as *Fosb* and *Mitf* that have been connected with AD (Solé-Domènech et al., 2016; Gupta et al., 1986). Validation of our RNA-seq data in HT-22 cells through RT-qPCR for selected B2-SRGs identified to be upregulated in 42 vs. R is presented in Figure 6D. In the same figure, we also present the expression levels of these genes during recovery, 24 hr after application of the amyloid beta toxicity stimulus, confirming the return to pre-stimulus levels and the specificity of the treatment with amyloid beta peptides (Figure 6D).

Out of the 25 genes that are upregulated in both mice and our cell culture system, six are B2-SRGs (4932438A13Rik, *Fosb*, *Pag1*, *Ptprs*, *Sema5a*, and *Sgms1*) and include a well-known immediate early gene (*Fosb*), genes associated with sensitivity to amyloid toxicity (*Pag1*, *Sema5a*, *Sgms1*, *Fosb*) (Hadar et al., 2016; Lin et al., 2009; Hsiao et al., 2013), as well as genes associated with *Trp53* (*Ptprs*, *Fosb*) (Motiwala and Jacob, 2006; Liu et al., 2018).

We then questioned whether by inducing an artificial degradation of B2 RNA, we would be able to induce expression of B2-SRGs in our model system of HT22 cells in the absence of any stimulus such as amyloid beta. To achieve this, we employed a similar approach and the same LNAs against B2 RNA that we used in our previous study (NIH/3T3, heat shock). Application of the LNA against B2 (Figure 7A) was able to reduce levels of B2 RNA compared to the control LNA (Figure 7B). Similarly to NIH/3T3 cells, targeting of the B2 RNA in HT22 cells, resulted in the increase of the expression levels of selected B2-SRGs that are upregulated in amyloid beta pathology (see Figure 2—figure supplement 2, the first 10 genes and Figure 6D). The increase in gene expression occurred in the absence of the stress stimulus, in this case amyloid beta, suggesting that these genes are under the suppressive control of B2 RNAs in HT22 cells (Figure 7C). At the same time, B2 RNA destabilization did not affect expression of five non-B2-SRGs that were used as negative controls, including example genes, such as *Adcy1* and *Kalrn*, that are nevertheless upregulated in amyloid beta pathology (Figure 2—figure supplement 2, the last two genes).

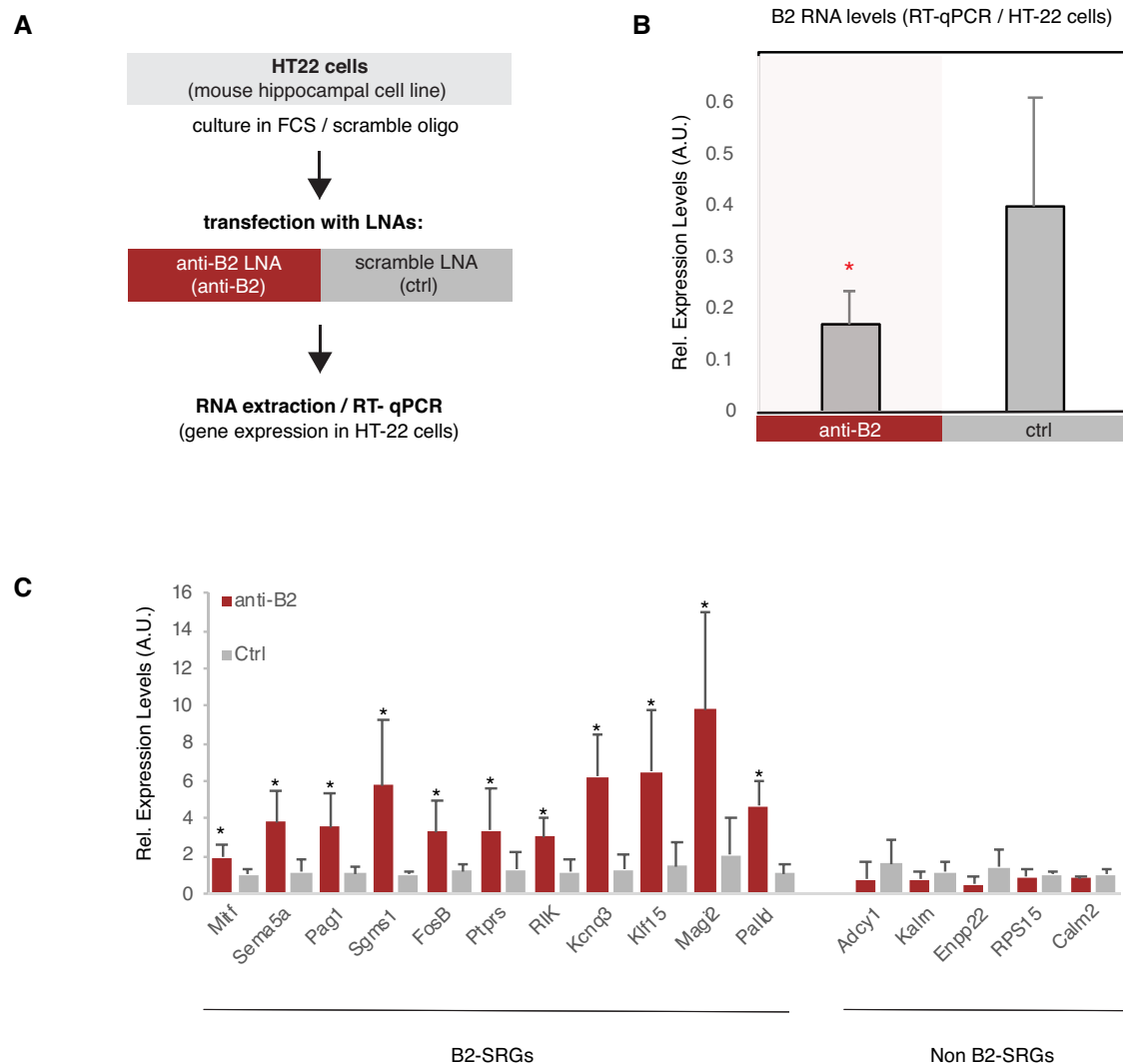


Figure 7. B2 RNA destabilization leads to increase in expression of B2-SRGs.

(A) Experimental design for the B2 RNA knock-down cell culture assay employing HT-22 cells. (B) Expression levels of full-length B2 RNA (RT-qPCR) in the B2 RNA KD experiment. Statistical significance (p value threshold 0.05) for anti-B2 less than control (depicted as asterisk, n = 3, unpaired directional t-test, error bars represent standard deviation from the mean). (C) Expression levels (RT-qPCR) of selected B2-SRGs (see Figure 2—figure supplement 2 and Figure 6D) and the respective negative controls during B2 RNA KD. Statistical significance (p value threshold 0.05) for anti-B2 greater than control (depicted as asterisk, n = 3/group, unpaired directional t-test, error bars represent standard deviation from the mean). Negative controls include five non-B2-SRGs that show no statistically significant difference between the two conditions.

Subsequently, in order to test the impact of amyloid toxicity to Hsf1-mediated B2 RNA processing we treated HT22 cells with either an LNA against Hsf1 or a scramble LNA (control) followed by

incubation with the 1–42 peptides, that subject the cells to amyloid toxicity stress, or the respective control peptide (Figure 8A, same experimental design and cells as in Figure 6). As shown in Figure 8B, treatment with the anti-Hsf1 LNA suppressed any increase in Hsf1 levels between 42-anti-Hsf1 and R-anti-Hsf1 cells, while this increase is observed between non-Hsf1 LNA-treated cells upon application of the amyloid beta peptides (42-ctrl vs R-ctrl). Treatment with the anti-Hsf1 LNA during transfection with the amyloid beta peptides also led to a decrease in protein levels of Hsf1 (Figure 5—figure supplement 1B). As observed in the APP mice, amyloid beta peptides resulted in increased B2 RNA processing only in cells without Hsf1 knock down, suggesting that it is the toxicity of these peptides that induces the SINE B2 RNA transcriptome changes (Figure 8C, Figure 8—figure supplement 1). In contrast, under anti-Hsf1 LNA, application of the 1–42 peptides was unable to increase B2 RNA processing in the cells compared to the control cells (Figure 8C). A similar effect was observed in B2-SRGs that are upregulated during amyloid toxicity (Supplementary file 1, Figure 8D–F). A selected number of genes was also tested with RT-qPCR to confirm the patterns observed in the RNA-seq data (Figure 8E). As in case of amyloid pathology, it cannot be excluded that our findings may extend to additional B2-SRGs (Figure 8—figure supplement 2).

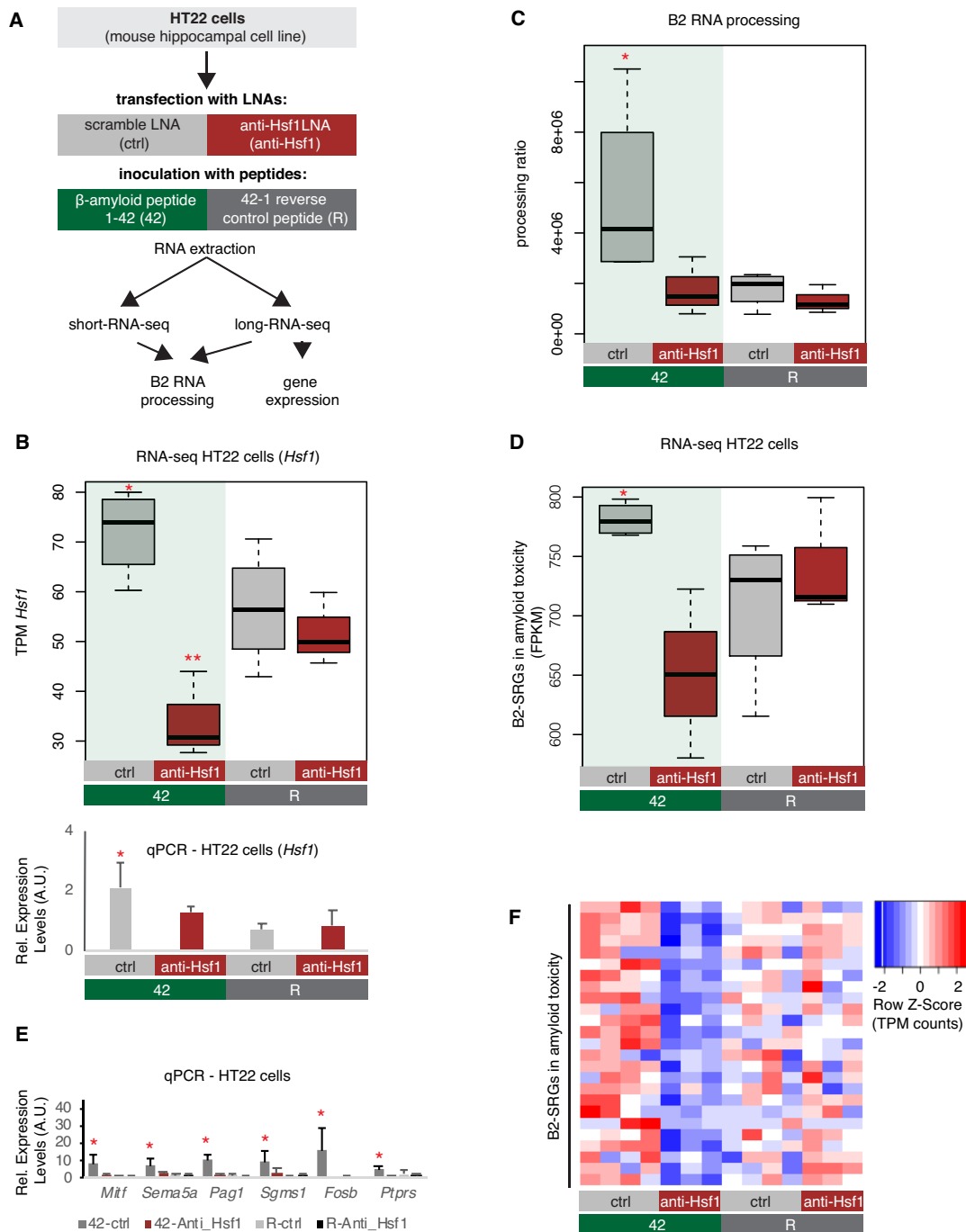


Figure 8. Hsf1 mediates B2 RNA processing in amyloid toxicity.

(A) Experimental design of the combined Hsf1 Knock Down – amyloid toxicity assay in HT22 cells followed by short and long RNA-seq. (B) Expression levels of Hsf1 as defined by long-RNA-seq (upper panel) and RT-qPCR (lower panel). Boxplots compare Hsf1 expression (TPM values) during incubation with the scramble LNA, and anti-Hsf1-specific LNA, incubated with either the 42 or R amyloid peptides. Statistical significance (p value threshold 0.05) for the comparison between 42/ctrl (n = 4) and 42/anti-Hsf1 (n = 3), unpaired non-directional t-test and for 42/ctrl (n = 4) greater than R/ctrl (n = 4) (both depicted as one asterisk), while two asterisks represent p<0.05

between 42/anti-Hsf1 and the other three groups. (C) B2 RNA processing ratio based on a combination of short and long-RNA-seq. Boxplot depicts distribution of total SINE B2 RNA processing ratio among different groups of HT22 cells between 42 and R. Statistical significance (p value threshold 0.05) for 42/ctrl greater than R/ctrl ($p=0.04$, $n = 4/\text{group}$, unpaired directional t-test, depicted as asterisk). No significant difference was observed between 42/anti-Hsf1 and R/anti-Hsf1 ($n = 3/\text{group}$) and between R/anti-Hsf1 and R/ctrl. (D) B2-SRGs expression levels in amyloid beta toxicity based on long-RNA-seq data (FPKM values) for genes of Supplementary file 1 (25 genes). Boxplot depicts distribution of expression levels in HT22 cells between 42 and R. $p=0.05$ for 42/ctrl greater than R/ctrl ($n = 4/\text{group}$, unpaired directional t-test, depicted as asterisk). No significant difference was observed between 42/anti-Hsf1 and R/anti-Hsf1 (NS, $n = 3/\text{group}$) and between R/anti-Hsf1 and R/ctrl. (E) Expression levels of selected B2-SRGs in the four conditions tested in our amyloid toxicity model though RT-qPCR. Statistical significance (p value threshold 0.05) for 42/ctrl greater than R/ctrl, unpaired directional t-test, n numbers as in subfigure B, with $p<0.05$ depicted as asterisk (error bars represent standard deviation from the mean). Samples and values depicted for non-Hsf1 LNA treated samples are the same as in Figure 6D and are used as controls as these samples were treated with a scramble LNA to allow comparison with the Hsf1 LNA treated samples. (F) Gene expression levels (long-RNA-seq) of B2-SRGs that are upregulated in amyloid toxicity (Supplementary file 1, 25 genes) show strong association with Hsf1 treatment during response to amyloid toxicity in HT22 cells. Heatmap depicts gene expression with rows as B2-SRGs in amyloid toxicity (Supplementary file 1) and columns as different HT22 cell treatments. TPM values are normalized per row. Red color represents higher expression.

These results suggest that amyloid beta toxicity induces B2 RNA processing also in vitro and Hsf1 comprises a necessary component in the upstream activation pathway of B2 RNA processing and, thus, of the genes regulated by B2 RNA.

2.6 Discussion

Fewer than five ribozymes have been identified in mammals, including our previous work on RNAs made from two retrotransposon families, murine SINE B2 RNAs and human SINE ALU RNAs, which are self-cleaving RNAs (Hernandez et al., 2020). We previously showed that cleavage in SINE B2 RNAs controls response to cellular stress through activation of stress response genes in heat shock. However, no connection between this novel molecular mechanism and pathologic processes was until now known. Moreover, since B2 RNAs are intrinsically

reactive, and contact with Ezh2 only accelerates cleavage, it remained plausible that other stress-related proteins may also have a similar effect on accelerating B2 RNA processing, which would link this ribozyme-like property to stress response through pathways other than Ezh2. This is especially relevant in mouse tissues, such as the brain, where Ezh2 expression is limited, an expression pattern observed also in human (Human Protein Atlas) (Uhlén et al., 2015).

Here, we unveil increased processing of SINE B2 RNAs as a novel type of transcriptome deregulation underlying amyloid beta neuro-pathology. Our data provides a new link in the murine hippocampal pathways connecting amyloid beta toxicity with transcriptome changes in SRGs through processing of B2 RNAs. In particular, the B2 RNA processing ratio increases upon progression of amyloid pathology both in mouse hippocampus and a hippocampal cell culture model, and B2-SRGs become hyperactivated. Consistent with the spatial proximity between B2-SRGs and Hsf1 binding sites, Hsf1 proved to be key for mediating B2 RNA processing in response to amyloid toxicity. This correlation is observed throughout all sequencing experiments performed in this study (Figure 8—figure supplement 1). Our work assigns to Hsf1 a new function that is independent of its long-established transcription factor function and includes the interaction with and processing of SINE B2 RNAs. The high levels of Hsf1 trigger a downstream cascade of events which are orchestrated into a cell-wide, SRG-mediated response to stress conditions. This axis is mediated by the ability of B2 RNA to get processed and, thus, act as a molecular switch. Although healthy cells and animals are able to restore the expression levels of Hsf1, SRGs and their regulating B2 RNAs upon removal of the stress-generating stimulus, the amyloid beta load in our biological models acts as a continuous stimulus that causes the Hsf1 - B2 RNA - SRG axis to

‘lock’ into an activated mode. Upregulation of SRGs results in increased Trp53 levels that induce neuronal cell death (Figure 9).

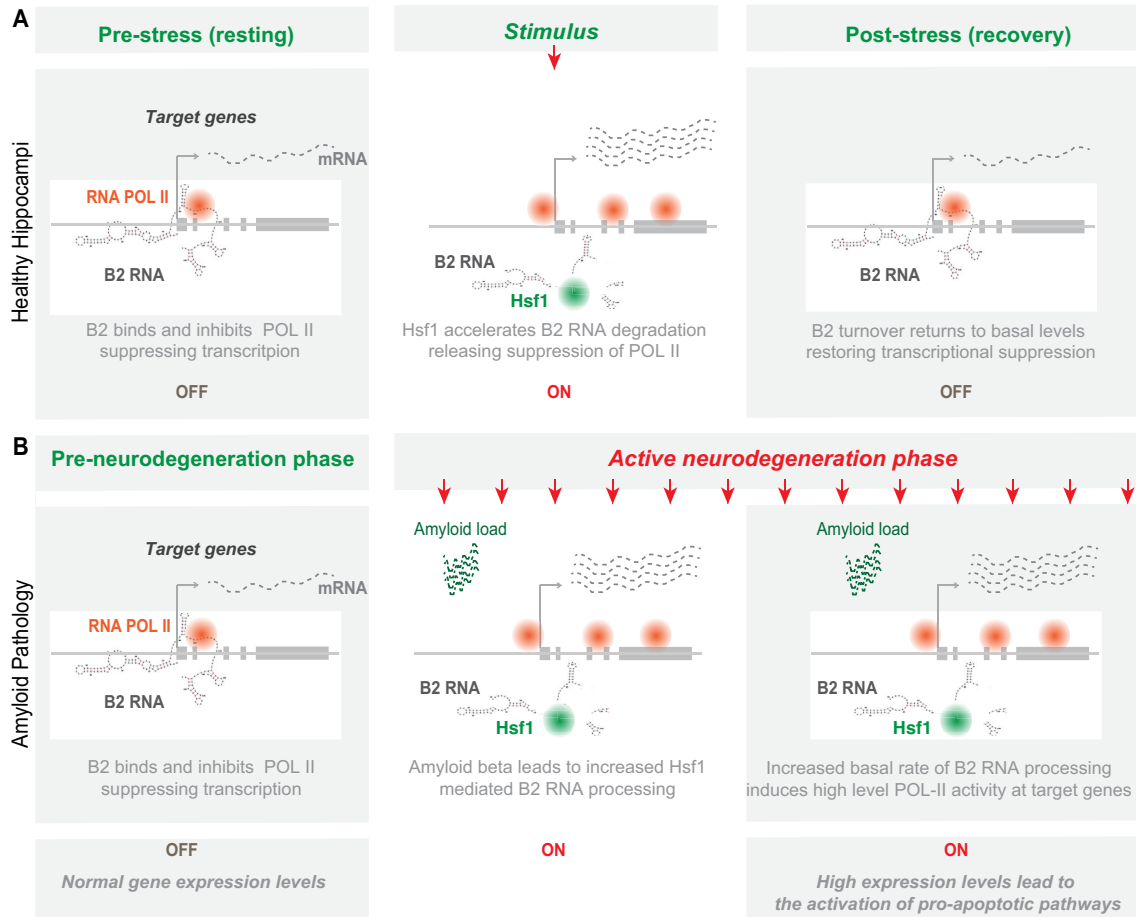


Figure 9. Representation of the role of B2 RNA processing in amyloid pathology.

Upon removal of the stress-generating stimulus, healthy cells restore the expression levels of Hsf1, specific B2 RNA regulated target genes and processing ratio of B2 RNAs returns to base levels. In contrast, in amyloid pathology, increased amyloid beta load acts as a continuous stimulus that causes the Hsf1 - B2 RNA – B2-SRG axis to ‘lock’ into an activated mode. ON/OFF represent active and suppressed SRG transcription, respectively.

Stress response genes, that constitute the basis of response to heat shock, have been shown in various studies from us and others to play a critical role also in hippocampal function (Peleg et al., 2010; Zovoilis et al., 2011). Thus, our findings here on a potential role of B2 RNAs in the context of neural response to stress constitute a natural continuation of our previous study in heat shock (Zovoilis et al., 2016). The fact that B2-SRGs identified in our previous study were found in the current study to be highly enriched in neuronal tissue related biological processes terms and compartments was what has compelled us to investigate further a potential role of B2 RNAs in neural tissue-associated pathologies.

In this study, we employed a mouse model of amyloid pathology in order to test the impact of increased amyloid beta load on B2 RNA processing in vivo. This mouse model has the NL- G-F mutations in the amyloid precursor protein knocked-in to a C57BL genomic background, with each mutation contributing to an increased severity and speed manifestation of the disease (Mehla et al., 2019). The combined effect of the APPNL–G–F mutations results in mice that experience rapid onset of AD-like symptoms at approximately 6 months (Mehla et al., 2019). This is exactly the time point that we observed the massive hyper-activation of B2 RNA SRGs, Trp53, and B2 RNA processing, suggesting that these changes constitute a molecular signature for the active neurodegenerative phase months, before the end state of 12 months that the mice develop terminal AD-like pathology and symptoms.

Interestingly, symptoms and molecular changes are not observed in younger APP mice suggesting the existence of a yet unknown protective mechanism against increased neuronal activity of SRGs in 3-month-old APP mice. This may be attributed to the increased neuronal plasticity observed in

younger brains (Lilja et al., 2013), which suggest that mechanisms in the younger brain may exist for counter-acting this excessive activity. In particular, stress response genes such as Fosb have been shown to participate in specification of cell-type-specific activity-dependent gene programs early in development (Yap and Greenberg, 2018). In contrast, during aging, B2-SRG activity in WT mice ramps down, which is not the case in APP mice (Figure 2), explaining at the molecular level the increased Trp53 levels and subsequent activation of cell death (Figure 1F). Our findings for B2 RNA SRG upregulation and B2 RNA increased processing coincide with the active neurodegeneration phase of amyloid pathology but in later stages the effect is not that prominent. This suggests that the initial active phase differs from the terminal stage regarding role of B2 RNA. As in the terminal stage a large number of cells have already died, B2 processing activation appears to be more connected with the initial response of cells to amyloid toxicity.

The B2-RNA-mediated regulation of gene expression during stress identified in one biological context may be relevant to a broad range of cellular types and disease pathophysiology. In the current study, to identify examples of genes that are subject to B2 RNA regulation in hippocampus molecular pathology we utilized the gene list of B2-SRGs identified in our previous study in heat shock. Of the previously identified 1684 B2-SRGs, we found here 72 genes that are deregulated specifically during the active neurodegeneration phase in hippocampal cells. However, this does not exclude the possibility that beyond these 72 genes, B2 RNA processing may also affect expression of additional B2-SRGs that are associated with other hippocampal functions that are independent of amyloid beta pathology. To get a better insight into this possibility, we tested the correlation between the gene expression of other B2-SRGs and the B2 RNA processing ratio, independently of amyloid beta status. In particular, for each of these genes we have calculated the

Pearson correlation coefficient between gene expression levels and B2 RNA processing ratio in the hippocampal samples of the current study. As shown in Figure 8—figure supplement 1E, we have been able to calculate a statistically significant correlation coefficient for at least 659 genes (the rest where either not sufficiently expressed in all samples or returned a p value > 0.05). From these 659 genes, expression of 344 genes (52.2%) showed a strong correlation with B2 RNA processing ratio ($r \geq 0.5$), while an additional 75 genes (11.3%) showed a weak correlation ($0.5 > r \geq 0.25$), and only 240 genes (36.4%) had an $r < 0.25$ and no correlation. This data suggests that for a large number of the B2-SRGs previously identified in the context of heat shock, correlation between gene expression and B2 RNA processing ratio holds true also in the context of hippocampal cells.

During our amyloid toxicity experiments, cells inoculated with the reverse peptide and the anti-Hsf1 LNA did not show any reduction in Hsf1 levels in contrast to those inoculated with amyloid beta (Figure 8B). This could be attributed to compensation during non-stress conditions. In contrast, under stress conditions, when Hsf1 is heavily used due to stress response, cellular needs surpass the available Hsf1 transcripts that are continuously depleted by the LNA. This is also in agreement with levels of B2-SRGs in this condition (Figure 8E), which are minimal in both R-ctrl and R-anti-Hsf1 conditions and only get activated during stress response to the amyloid beta in the 42-ctrl condition.

Our study leaves a number of open questions. Particularly, our study raises the question whether a similar mode of regulation of SRGs by SINE RNAs may exist also in human and which SINE RNAs could play such a role. Similarly to B2 RNA, such SINE RNAs would be able to bind and

inhibit RNA Pol II and would be subject to a similar RNA processing mechanism enabling the release of RNA Pol II. A number of studies have described that in human, repetitive SINE RNAs of the Alu class are also upregulated during cellular stress and can bind RNA Pol II inhibiting the transcription of target genes (Yakovchuk et al., 2009). Alu RNAs are widely regarded as the equivalent in human of B2 RNA. Most importantly, as we showed before, human Alu RNAs, alike B2 RNAs, are self-cleaving RNAs and can become destabilized in vitro (Hernandez et al., 2020). It remains unknown whether SINE RNAs and Hsfl play a similar role in amyloid pathology in the case of humans and whether we can extrapolate the generated conclusions in murine models to deduce that SINE RNAs are key components of the pathophysiological mechanisms underlying debilitating diseases such as AD. One major limitation compared to human pathophysiology is that the phenotype of amyloid pathology is not observed in mice even during aging. Nonetheless, a stress-central role of Alu RNAs, the human counterpart of B2 RNAs is plausible and, thus, future studies need to elucidate whether Alu RNA processing is also hyperactivated in the brain of patients with amyloid pathology in the context of AD.

Moreover, in the RNA-seq data one cannot distinguish between Pol III transcribed B2 RNA and Pol II transcribed B2 RNA (typically embedded within introns and UTRs of mRNAs). To get an indication whether such transcripts may contribute to our data, we have separated the B2 elements against which we map the RNA fragments into two categories, those that fall within exonic/genic regions and those outside these regions (Figure 3—figure supplement 2). Although B2 RNAs are produced by multiple copies in the genome, each copy does harbor multiple SNPs, insertions and deletions, which means that each B2 RNA fragment is mapped to a specific set of B2 elements and not to all of them. Thus, despite multiple mapping of the reads, a level of spatial specificity is

maintained. If the B2 RNAs we map were coming exclusively from either only Pol III B2 elements or mRNA-embedded B2 elements, we would expect at least some difference in the distribution of fragments between B2 elements of these two categories, as the second one overlaps with mRNAs. However, as shown in Figure 3—figure supplement 2, the fact that distribution models are very similar between the two categories supports the hypothesis that both types of B2 elements may contribute to B2 RNA processing. Thus, it cannot be excluded whether the regulatory role of B2 RNAs may extend from Pol III transcribed B2 RNAs into B2 RNAs embedded into mRNAs (likely nascent ones) that may be also under the same endogenous ribozyme activity of this sequence, may suppress Pol II and get processed in response to stimuli.

Our results suggest that B2 RNA regulation is a new process implicated in response to stress in amyloid pathology but it is definitely not the only one. Since B2-SRGs are highly interconnected and interweaved into various pathways, the impact of SRG hyperactivation by B2 RNA processing may extend to pathways that lie downstream of SRGs and affect various gene programs without binding directly B2 RNA. However, here it should be noted that high levels of Hsf1 are certainly expected to affect transcription also through Hsf1's conventional transcriptional factor function while there are still many SRGs that are not directly regulated by B2 RNA. Thus, B2 RNA processing described here does not constitute the only one but just one of the parameters in the equation of SRG regulation. It remains unclear which is the interplay between Hsf1 traditional transcription factor activities and its ability to affect B2 RNA processing and whether there is any overlap or synergies. For example, it remains unclear what percentage of activated genes are activated through Hsf1 DNA binding and what through binding of B2 RNA and subsequent release of suppressed Pol II activity.

Moreover, given how easily B2 RNA is processed in the presence of certain proteins, Hsf1 may be only one of the factors accelerating B2 processing in mouse hippocampus as we are just beginning to understand the implications of this form of SINE RNA regulation in cells. A broader role of SINE RNA processing in brain physiology and pathophysiology constitutes, thus, a significant possibility that could further revise our understanding of these RNAs as something more than just transcriptional noise and ‘junk DNA’ products.

2.7 References

- Bloom GS (2014) Amyloid- β and tau: the trigger and bullet in alzheimer disease pathogenesis *JAMA Neurology* 71:505–508. <https://doi.org/10.1001/jamaneurol.2013.5847>
- Cornutiu G (2015) The epidemiological scale of Alzheimer's Disease *Journal of Clinical Medicine Research* 7:657–666. <https://doi.org/10.14740/jocmr2106w>
- Cummings JLee GRitter ASabbagh MZhong K (2019) Alzheimer's disease drug development pipeline: 2019 Alzheimer's & Dementia: Translational Research & Clinical Interventions 5:272–293. <https://doi.org/10.1016/j.trci.2019.05.008>
- Davis JBMaher P (1994) Protein kinase C activation inhibits glutamate-induced cytotoxicity in a neuronal cell line *Brain Research* 652:169–173. [https://doi.org/10.1016/0006-8993\(94\)90334-4](https://doi.org/10.1016/0006-8993(94)90334-4)
- Down TAAHubbard TJ (2002) Computational detection and location of transcription start sites in mammalian genomic DNA *Genome Research* 12:458–461. <https://doi.org/10.1101/gr.216102>
- Espinoza CAGoodrich JAKugel JF (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription *RNA* 13:583–596. <https://doi.org/10.1261/rna.310307>
- Gallo FTKatche CMorici JFMedina JHWeisstaub NV (2018) Immediate early genes, memory and psychiatric disorders: focus on c-Fos, Egr1 and arc *Frontiers in Behavioral Neuroscience* 12:79. <https://doi.org/10.3389/fnbeh.2018.00079>
- Gao JWang WYMAo YWGräff JGuan JSPan LMak GKim DSu SCTsai LH (2010) A novel pathway regulates memory and plasticity via SIRT1 and miR-134 *Nature* 466:1105–1109. <https://doi.org/10.1038/nature09271>

- Gupta MA Gupta AK Haberman HF (1986) Psychotropic drugs in dermatology A review and guidelines for use *Journal of the American Academy of Dermatology* 14:633–645. [https://doi.org/10.1016/s0190-9622\(86\)70081-9](https://doi.org/10.1016/s0190-9622(86)70081-9)
- Hadar AMilanesi ESquassina ANiola PChillotti CPasmanik-Chor MYaron OMartásek PRehavi MWeissglas-Volkov DShomron NGozes IGurwitz D (2016) RGS2 expression predicts amyloid- β sensitivity, MCI and Alzheimer's disease: genome-wide transcriptomic profiling and bioinformatics data mining *Translational Psychiatry* 6:e909. <https://doi.org/10.1038/tp.2016.179>
- Hane FT Robinson M Lee BY Bai O Leonenko Z Albert MS (2017) Recent progress in Alzheimer's Disease Research, Part 3: Diagnosis and Treatment *Journal of Alzheimer's Disease* 57:645–665. <https://doi.org/10.3233/JAD-160907>
- Hernandez AJ Zovailis A Cifuentes-Rojas C Han L Bujisic B Lee JT (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2 *PNAS* 117:415–425. <https://doi.org/10.1073/pnas.1917190117>
- Hsiao J-HT Fu Y Hill AF Halliday GM Kim WS (2013) Elevation in sphingomyelin synthase activity is associated with increases in Amyloid-Beta peptide generation *PLOS ONE* 8:e74016. <https://doi.org/10.1371/journal.pone.0074016>
- Huang da W Sherman BT Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources *Nature Protocols* 4:44–57. <https://doi.org/10.1038/nprot.2008.211>
- Huang da W Sherman BT Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists *Nucleic Acids Research* 37:1–13. <https://doi.org/10.1093/nar/gkn923>
- Inouye S Fujimoto M Nakamura T Takaki E Hayashida N Hai T Nakai A (2007) Heat shock transcription factor 1 opens chromatin structure of interleukin-6 promoter to facilitate binding of an activator or a repressor *Journal of Biological Chemistry* 282:33210–33217. <https://doi.org/10.1074/jbc.M704471200>
- Ittner LM Ke YD Delerue FB I MGladbach Avan Eersel JWölfling H Chieng BC Christie MJ Napier IA Eckert A Staufenbiel M Hardeman EGötz J (2010) Dendritic function of tau mediates Amyloid- β toxicity in Alzheimer's Disease Mouse Models *Cell* 142:387–397. <https://doi.org/10.1016/j.cell.2010.06.036>
- Karijolic J Zhao Y Alla R Glaunsinger B (2017) Genome-wide mapping of infection-induced SINE RNAs reveals a role in selective mRNA export *Nucleic Acids Research* 45:6194–6208. <https://doi.org/10.1093/nar/gkx180>
- Kim D Paggi J M Park C Bennett C Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype *Nature Biotechnology* 37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>

Kramerov DAVassetzky NS (2011) SINEs Wiley Interdisciplinary Reviews: RNA 2:772–786. <https://doi.org/10.1002/wrna.91>

Li HHHandsaker BWysoker AFennell TRuan JHomer NMarth GAbecasis GDurbin R1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/Map format and SAMtools Bioinformatics 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

Li HDurbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform Bioinformatics 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Lilja AMRöjdner JMustafiz TThomé CMStorelli EGonzalez DUnger-Lithner CGreig NHNordberg AMarutle A (2013) Age-dependent neuroplasticity mechanisms in alzheimer Tg2576 mice following modulation of brain amyloid- β levels PLOS ONE 8:e58752. <https://doi.org/10.1371/journal.pone.0058752>

Lin LLesnick TGMaraganore DMIacson O (2009) Axon guidance and synaptic maintenance: preclinical markers for neurodegenerative disease and therapeutics Trends in Neurosciences 32:142–149. <https://doi.org/10.1016/j.tins.2008.11.006>

Liu JLi LSuo WZ (2009) HT22 hippocampal neuronal cell line possesses functional cholinergic properties Life Sciences 84:267–271. <https://doi.org/10.1016/j.lfs.2008.12.008>

Liu SYao XZhang DSheng JWen XWang QChen GLi ZDu ZZhang X (2018) Analysis of transcription Factor-Related regulatory networks based on bioinformatics analysis and validation in hepatocellular carcinoma BioMed Research International 2018:1–16. <https://doi.org/10.1155/2018/1431396>

Love MIHuber WAnders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 Genome Biology 15:550. <https://doi.org/10.1186/s13059-014-0550-8>

Mahat DBSalamanca HHDuarte FMDanko CGLis JT (2016) Mammalian heat shock response and mechanisms underlying its Genome-wide transcriptional regulation Molecular Cell 62:63–78. <https://doi.org/10.1016/j.molcel.2016.02.025>

Mehla JLacoursiere SGLapointe VMcNaughton BLSutherland RJMcDonald RJMohajerani MH (2019) Age-dependent behavioral and biochemical characterization of single APP knock-in mouse (APPNL-G-F/NL-G-F) model of Alzheimer's disease Neurobiology of Aging 75:25–37. <https://doi.org/10.1016/j.neurobiolaging.2018.10.026>

Motiwalla TJacob ST (2006) Role of protein tyrosine phosphatases in Cancer Progress in Nucleic Acid Research and Molecular Biology 81:297–329. [https://doi.org/10.1016/S0079-6603\(06\)81008-1](https://doi.org/10.1016/S0079-6603(06)81008-1)

Pandey RMandal AKJha VMukerji M (2011) Heat shock factor binding in alu repeats expands its involvement in stress through an antisense mechanism Genome Biology 12:R117. <https://doi.org/10.1186/gb-2011-12-11-r117>

- Peleg SSananbenesi FZovoilis ABurkhardt SBahari-Javan SAgis-Balboa RCCota PWittnam JLGogol-Doering AOpitz LSalinas-Riester GDettenhofer MKang HFarinelli LChen WFischer A (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice *Science* 328:753–756. <https://doi.org/10.1126/science.1186088>
- Pertea MPertea GMAntonescu CMChang TCMendell JTSalzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads *Nature Biotechnology* 33:290–295. <https://doi.org/10.1038/nbt.3122>
- Ponicsan SLKugel JFGoodrich JA (2010) Genomic gems: sine RNAs regulate mRNA production *Current Opinion in Genetics & Development* 20:149–155. <https://doi.org/10.1016/j.gde.2010.01.004>
- Ponicsan SLKugel JFGoodrich JA (2015) Repression of RNA polymerase II transcription by B2 RNA depends on a specific pattern of structural regions in the RNA *Non-Coding RNA* 1:4–16. <https://doi.org/10.3390/ncrna1010004>
- Quinlan ARHall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rokavec MLi HJiang LHermeking H (2014) The p53/miR-34 Axis in development and disease *Journal of Molecular Cell Biology* 6:214–230. <https://doi.org/10.1093/jmcb/mju003>
- Saito TMatsuba YMihira NTakano JNilsson Pitohara SIwata NSaido TC (2014) Single app knock-in mouse models of Alzheimer's disease *Nature Neuroscience* 17:661–663. <https://doi.org/10.1038/nn.3697>
- Sananbenesi FFischer ASchrick CSpiess JRadulovic J (2003) Mitogen-activated protein kinase signaling in the Hippocampus and its modulation by corticotropin-releasing factor receptor 2: a possible link between stress and fear memory *The Journal of Neuroscience* 23:11436–11443. <https://doi.org/10.1523/JNEUROSCI.23-36-11436.2003>
- Solé-Domènech SCruz DLCapetillo-Zarate EMaxfield FR (2016) The endocytic pathway in microglia during health, aging and Alzheimer's disease *Ageing Research Reviews* 32:89–103. <https://doi.org/10.1016/j.arr.2016.07.002>
- Uhlén MFagerberg LHallström BMLindskog COKsvold PMardinoglu ASivertsson ÅKampf CSjöstedt EAsplund AOlsson IEdlund KLundberg ENavani SSzigyarto CAOdeberg JDjureinovic DTakanen JOHober SAlm TEdqvist PHBerling HTegel HMulder JRockberg JNilsson PSchwenk JMHamsten Mvon Feilitzen KForsberg MPersson LJohansson FZwahlen Mvon Heijne GNielsen JPontén F (2015) Proteomics Tissue-based map of the human proteome *Science* 347:1260419. <https://doi.org/10.1126/science.1260419>
- Yakovchuk PGoodrich JAKugel JF (2009) B2 RNA and alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes *PNAS* 106:5569–5574. <https://doi.org/10.1073/pnas.0810738106>

Yamakuchi M Lowenstein CJ (2009) MiR-34, SIRT1 and p53: the feedback loop *Cell Cycle* 8:712–715. <https://doi.org/10.4161/cc.8.5.7753>

Yap EL Greenberg ME (2018) Activity-Regulated transcription: bridging the gap between neural activity and behavior *Neuron* 100:330–348. <https://doi.org/10.1016/j.neuron.2018.10.013>

Yutsudo N Kamada T Kajitani K Nomaru H Katogi A Ohnishi Y H Ohnishi Y N Takase K Sakumi K Shigeto H Nakabeppu Y (2013) fosB-null mice display impaired adult hippocampal neurogenesis and spontaneous epilepsy with depressive behavior *Neuropsychopharmacology* 38:895–906. <https://doi.org/10.1038/npp.2012.260>

Zovoilis A Agbemenyah H Y Agis-Balboa R C Stilling R M Edbauer D Rao P Farinelli L Delalle I Schmitt A Falkai P Bahari-Javan S Burkhardt S Sananbenesi F Fischer A (2011) microRNA-34c is a novel target to treat dementias *The EMBO Journal* 30:4299–4308. <https://doi.org/10.1038/emboj.2011.327>

Zovoilis A Cifuentes-Rojas C Chu H P Hernandez A J Lee J T (2016) Destabilization of B2 RNA by EZH2 activates the stress response *Cell* 167:1788–1802. <https://doi.org/10.1016/j.cell.2016.11.041>

Chapter 3 - Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes

3.1 Foreword

This chapter is presented in the form of the manuscript published in EmboRep (Cheng *et al*, 2021). As a co-first author on the manuscript, alongside Yubo Cheng, I contributed “testing of the next-generation-sequencing and in vitro assay pipelines, library construction, RT-qPCRs, in vitro assays, data visualization and writing of the manuscript” (Cheng *et al*, 2021). Yubo Cheng contributed “Bioinformatics analysis, statistical analysis, establishment and testing of the analysis pipelines, data visualization and writing of the manuscript” (Cheng *et al*, 2021). Additional work was performed by Babita Gollen (“establishment and testing of the next-generation-sequencing pipelines, library construction”), Ana Alvarez Veronesi (“Collection of tissues”), Majid Mohajerani (“data interpretation”), Jeffrey T. Joseph (Pathology analysis, provision of material, study design, data interpretation), and Athanasios Zovoilis (“conception and design, establishment and testing of data generation and analysis pipelines, data interpretation, data visualization and writing of the manuscript, overall supervision.”) (Cheng *et al*, 2021).

The research presented in this chapter investigates the phenomenon of aberrant Alu RNA processing in Alzheimer’s disease patients. We study RNA-seq data from two sources: the Calgary Brain Bank, led by coauthor Dr. Jeffrey T Joseph, and the publicly available ROSMAP data presented in Bennet *et al*, *J. Alzheimer’s Dis* (2018).

In addition to patient tissue, numerous *in vitro* models were employed to establish the mechanism by which Alu RNA processing was accelerated. Briefly, many of the genes upregulated in Alzheimer’s disease patients correlated with the upregulation of the same genes in the SH-SY5Y

cell line when Alu RNA was knocked down. When incubated with Hsf1, Alu RNA is more rapidly processed.

3.2 Abstract

Despite significant steps in our understanding of Alzheimer's disease (AD), many of the molecular processes underlying its pathogenesis remain largely unknown. Here, we focus on the role of non-coding RNAs produced by small interspersed nuclear elements (SINEs). RNAs from SINE B2 repeats in mouse and SINE Alu repeats in humans, long regarded as “junk” DNA, control gene expression by binding RNA polymerase II and suppressing transcription. They also possess self-cleaving activity that is accelerated through their interaction with certain proteins disabling this suppression. Here, we show that similar to mouse SINE RNAs, human Alu RNAs, are processed, and the processing rate is increased in brains of AD patients. This increased processing correlates with the activation of genes up-regulated in AD patients, while increased intact Alu RNA levels correlate with down-regulated gene expression in AD. In vitro assays show that processing of Alu RNAs is accelerated by HSF1. Overall, our data show that RNAs from SINE elements in the human brain show a similar pattern of deregulation during amyloid beta pathology as in mouse.

3.3 Introduction

As the human life expectancy continues to expand, the number of people with aging-associated diseases, including late-onset Alzheimer's disease (AD), is expected to increase. This will generate a wave of new cases in the upcoming years, putting a strain on societies worldwide (Cornutiu, 2015; Montgomery et al, 2018). Amyloid pathology is a required feature of AD, where amyloid beta peptides and their aggregated forms have been associated with AD pathogenesis in brains of

Alzheimer's disease patients as well as neural cell toxicity in mouse models of amyloid pathology (Ittner et al, 2010; Bloom, 2014; Gonzalez et al, 2018; Mehla et al, 2019). However, despite the significant steps made in understanding AD pathogenesis during the last decades, diagnostic tests and clinical trials for potential therapeutic factors in AD remain inconclusive, as AD remains a highly complex disease, with extreme heterogeneity and many of the molecular processes underlying its pathogenesis still elusive (Bertram & Tanzi, 2008; Korczyn, 2012; Toledo et al, 2013; Hane et al, 2017). Among such molecular mechanisms that have attracted only a little attention until now are transcriptome changes in a class of RNAs produced by retrotransposons called SINE RNAs.

Retrotransposons are the dominant repetitive transposable elements in the mammalian genome (Natt & Thorsell, 2016). SINE RNAs are non-coding RNAs generated by a class of non-autonomous retrotransposons called small interspersed nuclear elements (SINEs), which consist of 100–500 base pairs (Tatosyan & Kramerov, 2016). The most successful SINE retroelement in the human genome, Alu SINE, has an amplification rate of one new insertion per 20 human births leading to the current one million transcripts in the genome (Walters et al, 2009; Deininger, 2011). While most of retrotransposons are inactive through epigenetic regulatory mechanisms, Alu, together with some other retroelements such as L1 (a LINE element) can be active (Mu et al, 2016; Platt et al, 2018). For a long time, most SINEs have been regarded as part of the so called “junk DNA”, a part of the non-coding genome with no apparent function (Karijovich et al, 2017). More recently, it was shown that SINE genomic elements contribute to genomic diversity through their integration into multiple genomic sites, by providing novel splicing sites, promoter elements,

enhancers and transcription factor binding sites or disrupting existing ones (Kaaij et al, 2019; Zhang et al, 2019).

However, little is known about the role of the non-coding RNAs transcribed by SINE genomic elements themselves. Two of the most frequent subclasses of SINEs, the B2 and Alu retrotransposons, are present in millions of copies in mouse and humans, respectively, and are either transcribed by RNA polymerase III independently or by RNA polymerase II as parts of naïve transcripts in which they are embedded. RNAs from these elements have been known for years to be up-regulated during response to various types of cellular stress. In particular, levels of RNAs transcribed from SINEs such as B1 and B2 (mouse), Alu (humans), PRE-1 (swine), Bm 1 (silkworm), and C (rabbit) have been reported to increase during heat shock (Tatosyan & Kramerov, 2016; Funkhouser et al, 2017). However, their exact role in this process remained unknown until a number of studies have revealed that levels of SINE B2 and Alu RNAs are elevated during response to stress to suppress transcription through their binding of RNA polymerase II (Espinoza et al, 2007; Mariner et al, 2008; Yakovchuk et al, 2009; Ponicsan et al, 2010, 2015). This binding contributes to the genome-wide transcriptional repression observed during cellular stress by suppressing housekeeping genes, potentially facilitating redistribution of energy resources of the cell to support survival (Mariner et al, 2008). Subsequently, we showed that SINE RNAs can also mediate cellular response to stress in a different way. In particular, we showed that SINE B2 RNAs in mouse show a protein-accelerated self-cleavage activity (Zovoilis et al, 2016; Cheng et al, 2020; Hernandez et al, 2020), leading to their fragmentation and thus interfering with their capability of binding to and suppressing RNA polymerase II. As a result, SINE RNAs in mouse act as transcriptional switches during response to stress stimuli by binding

RNA polymerase II at several stress response genes in the pro-cellular stress state and suppressing their transcription. Upon application of a stress stimulus, B2 RNAs are processed, leading to the release of the delayed or stalled RNA polymerase at stress response genes and, thus, their transcription (Zovoilis et al, 2016). Thus, SINE RNAs, based on their processing status, play an important role in stress response through either suppressing (as full-length RNAs) or activating (during processing) gene expression (Appendix Fig S1 provides a model combining what is known for this mode of regulation of gene expression in mouse).

Interestingly, in a recent study we have shown that this processing and destabilization of B2 RNAs is connected with a pathologic process in the cell, namely amyloid pathology in mouse neural cells (Cheng et al, 2020). In particular, increased amyloid beta load during the active neurodegenerative phase acts as a continuous stimulus that causes an increase in a key stress response factor, called Hsf1. Increased Hsf1 levels accelerate B2 RNA processing and lead the Hsf1/B2 RNA/stress response genes axis to “lock” into an activated mode with accompanying high levels of pro cell death genes, such as p53 (Cheng et al, 2020).

The above findings have revealed a role for SINE RNA processing in amyloid beta pathology in mouse brain and raised the intriguing possibility that a similar mode of deregulation of SINE RNA processing may also exist in human brain. In humans, SINE Alu RNAs are also able to bind and inhibit RNA Pol II alike to B2 RNAs (Mariner et al, 2008). We have found that, in vitro, much alike B2 RNAs, Alu RNAs are also self-cleaving RNAs and can become destabilized (Hernandez et al, 2020). Thus, in vivo, SINE Alu RNAs in human brains could also be subject to a similar RNA processing as SINE B2 RNAs in mouse brains, assigning a broader role to SINE RNA

processing in brain molecular physiology and AD pathogenesis. Here, we investigate this hypothesis by examining whether Alu RNAs are processed in human brains, whether their processing ratio is deregulated in brains of AD patients compared to healthy aging individuals, and the potential mechanisms underlying such a deregulation.

3.4 Materials and Methods

3.4.1 Post-mortem hippocampal tissues

Hippocampal tissues were acquired by the Calgary Brain Bank following the respective institutional approvals. De-identified clinical metadata are available as a separate phenotype file accompanying the next-generation sequencing data. Aggregate clinical data were as follows: AD, seven males and four females, with a median age of 77; no AD, seven males, six females, with a median age of 64. Tissues from hippocampus (approx. 0.75 cm³ cubes) were homogenized in 5 ml TRIzol reagent: 15-min incubation and subsequent grinding using a pestle until nothing but insoluble connective tissue remained. The homogenized mix was pipetted up and down and the solution was stored at -80°C. 1 ml of homogenized mixture was phase separated by the addition of 200 µl of chloroform (Sigma, C2432) and mixed by inversion, incubated for 3 min and centrifuged at 12,500 g for 15 min at 4°C. The top (aqueous) layer was transferred to a new tube and mixed with 500 µl of isopropanol (Fisher, 67-63-0), followed by a 1-h incubation at -20°C and centrifugation at 12,500 g for 10 min at 4°C. The supernatant was removed and the pellet was washed and mixed with 1 ml of 75% ethanol, followed by a centrifugation at 7,600 g for 5 min at 4°C. The supernatant was removed and the pellet was allowed to dry for 1 min before eluting in 50 µl of nuclease free H₂O. The eluted RNA was heated at 55°C for 15 min and subsequently incubated with 1 µl of DNaseI (NEB, M0303), 10 µl of 10× DNaseI buffer (NEB, B0303), and 39

μl of nuclease free H₂O for 15 min at 37°C. The RNA was further cleaned using the Zymo Research RNA clean and concentrator kit -25 (R1017). The RNA was stored at –80°C. Estimation of RIN scores for the purified RNAs revealed no difference between the AD and no AD patients (Appendix Fig S10B).

3.4.2 RNA in vitro transcription and RNA-protein incubations

Alu consensus template was ordered as IDT g-blocks™ (lower case denotes the T7 promoter sequence):

AluY:

5'-

taatacgactcactatagGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
CGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGA
AACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGGGGCGCCTG
TAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGG
AGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAG
ACTCCGTCTCA; templates were amplified by PCR using a T7 promoter sequence as the forward

primer: 5'-TAATACGACTCACTATAG and a reverse primer: 5'-TGAGACGGAGTCTCGCTC.

The amplified g-blocks were then gel band excised and extracted using the BioBasic EZ-10 gel extraction kit (BS353). The gel excised DNA was used as an additional PCR template and the amplified g-block was in vitro transcribed by the Hiscribe T7 High Yield RNA Synthesis Kit (NEB, E2040S) for 2 h at 37°C. RNA was cleaned using the Qiagen RNeasy MinElute Kit. RNA

controls used are Control#1(G-44U):

5'GCCCCGUUGCAAUGGAAUGACAGCGGGUAUGUUAACAACCCCAUCCGUCAUGG
AGACAGGUGGACGUUAAAUAUAAACCUGAAGAUUAAACAUGACUGAAUCUUUUG
CUACUAGAAUGGUGAGCAAGGGCGAGGAGCUGUUC 3' and Control #2 (Zika 3'

scramble):

CCGCAGGCGAGUGAUGCAGAUAAAAACAUACUCGUCCAAUCCGAAAAUGAUCGGC
GGUCUUUCAGCGCACGCGAGCGAGCGACCGCCUUCGGCUGAUGCGCACGCCGCGC
GCAGGGACAGGUACAUCUUCAGAGGGAGCGGUCAGAAAUUGCACACGAUAGAAG
GAAAGACCCUAGAAUGCCCCACUCGGGGAUGUGCGCCACAAAGGACUCCGGCAGU
GAAGUUGAAUCGAGCGCCGAGGAGAUAGCCCGCCAGCACGACCACGGCGGACAGG
GUGCGGCGCGCCAUGGAUUUGUGGGACUUCCAAUAAGGCGUCGCUGAAGGCUU
ACUGAAUCUCUUCACAGAUGUGCAGGGAACUCUACAAGCAGGAUUAGGGCCUCCU
ACACUUGCGCACACGCACCGGUAGAAAUCAACGGGACGAUCUACGAUCA. Both

RNAs were generously donated by the Wieden and Patel laboratory respectively, from the University of Lethbridge.

In vitro experiments were done as described before (Zovoilis et al, 2016; Cheng et al, 2020; Hernandez et al, 2020) and included a refolding reaction for the RNA samples with 300 mM NaCl into nuclease free water. The incubation of 2 pmoles of Alu transcripts (0.33 pm for Fig 8A) occurred as follows: RNA was incubated for 1 min at 50°C and cooled at a rate of 1°C/10 s until 4°C. Samples were then incubated at 37°C with the addition of TAP buffer (final reaction concentrations: 5 nM Tris pH 7.9, 0.5 mM MgCl₂, 0.02 mM EDTA, 0.01% NP40, 1% glycerol, 0.2 mM DTT) or protein samples diluted in TAP buffer and folding buffer (reaction concentration: 300 mM NaCl, 50 mM Hepes). Degradation of Alu RNA was analyzed on 8 M urea 10% PAGE gels, stained with SYBR II (Invitrogen, S7564). Samples were mixed with Novex™ 2× TBE-Urea loading buffer (Invitrogen, LC6876) and denatured at 95°C for 5 min before loading. Electrophoresis occurred in 1× TBE buffer at 180 V. Gel analysis was performed using Amersham

Typhoon instruments. Band absorbance was analyzed using ImageJ area under the curve software and normalization by the ratio of experimental over initial.

HSF1 protein incubations were performed with phosphorylated, recombinant, His-tagged HSF1 (Enzo Life Sciences, ADI-SPP-902). HSF1 working concentrations were 250 nM unless otherwise specified, diluted in TAP buffer.

3.4.3 Short RNA-seq

5 µg total RNA was PNK phosphorylated (85 µl RNA, 10 µl 10× PNK buffer, 5 µl T4 PNK [NEB, M0201], incubated at 37°C for 1 h). The RNA was then cleaned using the MinElute RNA Cleanup Kit (Qiagen, 74204) with a customized protocol for total RNA recovery: in brief, 100 µl sample mixed with 350 µl supplied RPE buffer and 525 µl 100% ethanol. The samples were centrifuged and washed as recommended. 3,000 ng of remaining RNA was 2× rRNA depleted (NEB, E6130). The library was then prepared using NEBNext® Small RNA Library Prep Set for Illumina (NEB, E7330) with 1:10 adapter dilutions. Final PCR used 18 cycles and kit optimized index primers. Sample clean ups were performed as recommended using Omega NGS Total Pure Mag Beads (Omega) and following amplification, samples were twice 1.2× bead cleaned. Quality control analysis was performed using the Agilent bioanalyzer 2100 RNA pico kit and DNA HS kit. Quantification of libraries was done by qPCR using the NEBNext library quant kit for Illumina and library sizes were analyzed using the Agilent bioanalyzer 2100 HS DNA kit. Equimolar amounts were prepared for sequencing. Libraries were sequenced on the Illumina HiSeq platform in pairwise fashion. Appendix Fig S10A shows a representative Bioanalyzer electropherogram of one of these libraries.

3.4.4 SH-SY5Y cell culture

SH-SY5Y neuroblastoma cell line was obtained from ATCC (CRL-2266). Cells were cultured in 1% pen/strep treated EMEM (Sigma, M4655) and F12 Ham (Sigma, N6658) mixed 1:1, 1% l-glutamine (200 mM; Gibco, 25030149), 1% Non-essential amino acids (SAFC, M7145), and 10% FBS (Sigma, F4135). Before any assay, cells were thawed and plated in a T75 flask. During subculture, cell pellet was resuspended in 10 ml fresh medium and re-plated. Cells were maintained in 5% CO₂, 37°C. Cell passaging was done by incubating at 37°C with trypsin (Gibco, 12605028) for 5 min. Trypsin was then inhibited 1:1 with fresh medium, and cells are then centrifuged. The pellet was washed with 5 ml PBS, resuspended in fresh medium, and plated.

3.4.5 LNA transfections

500,000 cells were plated in a 6-well plate well. 15 min after plating, 100 µM of scramble LNA (Qiagen, custom) and 12.5 µM of each anti-Alu LNA (100 µM total pooled; IDT, custom) were prepared and applied dropwise in solution with 10.7% Hiperfect transfecting agent (Qiagen, 301704). Transfections occurred for 12 h before treating the cells with trypsin as described above to displace and pellet the cells. Cells were then added to a vial with 1 ml TRIzol and allowed at room temperature for 5 min. Pellets in TRIzol were then stored at -80°C. RNA extraction was performed as described with hippocampus tissue above. The anti-Alu LNAs are as follows (* denotes a phosphorothioate bond and +denotes an LNA bond):

- 1.+C**A**C**G**C**C**G**G**C**T**A**A**T**+T
- 2.+C**A**C**G**C**C**C**A**G**C**T**A**A**T**+T
- 3.+C**A**T**G**C**C**C**G**G**C**T**A**A**T**+T

4.+T*+C*+C*+T*G*C*C*T*C*A*G*C*+C*+T*+C*+C
5.+G*+G*+A*+G*T*C*T*C*G*C*T*+C*+T*+G*+T
6.+G*+G*+A*+G*T*C*T*C*A*C*T*+C*+T*+G*+T
7.+A*+G*+A*+G*T*C*T*C*G*C*T*+C*+T*+G*+T
8.+T*+C*+G*+G*C*C*T*C*C*C*A*A*+A*+G*+T*+G. The scramble LNA used is as follows: +C*+C*+T*+C*A*A*T*T*T*T*A*+T*+C*+A*+C.

3.4.6 RT-qPCR

Reverse transcription was performed by the standard Thermo-Fisher M-MLV reverse transcription protocol using the M-MLV reverse transcriptase (Invitrogen, 28025013) with 500 ng RNA. Quantitative PCR was performed using the Luna Universal qPCR mix (NEB, M3003). Reactions were set up so cDNA was diluted in water 1:20, constituting 20% of the overall reaction. Primers were diluted to 10 μ M, constituting 5% of the overall reaction, and the master mix was added to 50% of the overall reaction. Total reaction volume was 10 μ l. Standard dilution pools were set up in four standard dilutions from a pooled sample mixture, diluted 1:5, 1:10, 1:20, and 1:40. Thermocycler conditions were as follows: 3 min at 95°C (15 s at 95°C, 30 s at 54°C, 30 s at 66°C) $\times$ 40 cycles. qPCR was performed using the Bio-Rad CFX384 Real-time detection system and measurements were taken during extension steps. Primers used are included in Table EV1. Experiments were performed in three biological replicates and three technical replicates.

3.4.7 Bioinformatics analysis

For Calgary Brain Bank (CBB) short RNA-seq data analysis, standard Illumina adaptor sequences were trimmed off using cutadapt-1.18 (<https://doi.org/10.14806/ej.17.1.200>). For the construction

of the “ALUome”, we retrieved Alu sequences for all identified Alu elements (within the range of 280–320 bp in order to exclude truncated forms) from UCSC hg38 Repeat Masker (August 2018) and subsequently collapsed sequences present in our list more than once into one unique representative sequence, resulting only in unique Alu sequences. Alu RNA sequences mapping within known exons, upstream/downstream 300 bp of exons, snoRNAs, microRNAs, and tRNAs were filtered out (using as annotation: Hg38, UCSC NCBI RefSeq, sno/miRNA, tRNA Genes). Reads for each sample were mapped to the ALUome using bwa-0.7.17 (Li & Durbin, 2009) in single end mode, with parameters: maximum edit distance: 0, maximum number of gap opens: 0, disallow a long deletion within bp toward the 3' end: 100, disallow an indel within bp toward the ends: 100, maximum edit distance in the seed: 0, mismatch penalty: 4, gap open penalty: 22, gap extension penalty: 12. SAM format files generated from mapping were converted to BAM format files using samtools-1.6 (Li et al, 2009) and to files in BED format with bamToBed utility from BEDTools-2.26.0 (Quinlan & Hall, 2010). Then, start positions of reads mapping in the sense direction to each unique Alu of the ALUome were retrieved. For each sample, a matrix was generated with Alu sequences as rows and counts of read starts for each Alu position as columns, normalized to FASTQ file reads (per 100 million reads). Sum of counts from position above 130 were calculated for each Alu and patient, and then, the respective count matrices were generated combining each Alu (rows) with the counts in the patients (columns) for position above 130. R (version 3.4.3) (<https://www.R-project.org/>) package DESeq2 (Love et al, 2014) was applied to counts above 130 between AD and no AD patients to get differentially expressed processed Alus. To calculate the Alu RNA processing ratio, Babraham NGS analysis suite Seqmonk 1.45.4 (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>) was used to obtain the number of reads overlapping with RNA 7SL as well as the read coverage per base along all Alu loci.

Processing ratio for each sample was calculated by the processed Alu count obtained from mapping fragments to unique Alus normalized by the RNA 7SL counts and total Alu read coverage per base: [processed fragments (position above 130)/(Alu read coverage per base/RNA 7SL count)].

For ROSMAP study data, RNA-seq data from the AMP-AD Knowledge Portal (<https://adknowledgeportal.synapse.org>) for the MAP project (Synapse ID: syn3219045) (Bennett et al, 2012) were acquired and paired end FASTQ files of each sample were converted from original data in BAM format using utility bam2fq in samtools-1.6. Then, sequencing data were mapped against the ALUome and analyzed generating Alu count matrices. For each sample, a matrix was generated with Alu sequences as rows and counts of read starts for each Alu position as columns, normalized to FASTQ file reads (per 100 million reads). Sum of counts from position 0 to 15 (full-length Alu) and above 130 (processed Alus) were calculated for each Alu and patient, and then, the respective count matrices were generated combining each Alu (rows) with the counts in the patients (columns) for position 0 to 15 and above 130. To get differentially expressed processed Alus, the R package DESeq2 was applied to counts above 130 between no AD patients (with a final clinical consensus diagnosis of no cognitive impairment and a Braak stage: 0–III) and AD (final clinical consensus diagnosis: Alzheimer’s disease and no other cause of cognitive impairment, Braak stage: IV–VI). Processing ratio for each sample was calculated by dividing the processed Alu count obtained from mapping to unique Alus with the full-length Alu level count mentioned above.

The decision for the selection of the appropriate method for estimating Alu RNA processing ratio in short RNA-seq and RNA-seq data was based on whether the applied RNA selection and library construction protocol maintains both fragments generated by the cut at XR1 (one overlapping with left arm and one overlapping right arm) or only the one at the right arm. As shown in Appendix Fig S10C, short RNA-seq maintains both fragments as depicted by the existence of two major 3' end densities, one around XR1 and one around the expected end of the full-length Alu RNA. In contrast, RNA-seq of poly-A RNA filters out the fragment overlapping the left arm, as now increased density is observed around XR1 position. This is in accordance with the poly-A selection step used, which would select only the full-length fragment and the processed fragment at the right arm. Thus, position 0–15 is expected in the standard RNA-seq protocol to map only full-length Alu reads, which is why the count at this position is used for the estimation of the processing ratio at Fig 4. In contrast, for short RNA-seq, the 0–15 position count number includes both full-length and fragmented RNAs. Thus, for the estimation of the processing ratio in short RNA-seq samples, we used the total Alu read coverage per base that is independent from the position of the mapping of the 5' end of the reads. As in our previous studies in mouse, the SINE RNA levels were normalized to a Pol III transcript (7SL RNA) to account for non-specific degradation by nucleases. As shown in Appendix Fig S10D, as expected, we observed no difference in 7SL levels between the two groups of patients, supporting the use for the per sample normalization.

Regarding gene expression analysis of MAP patients, reads for each sample were mapped to the reference genome acquired from ensemble (GRCh38/November 2018, primary assembly) using hisat2-2.1.0, in single end mode, with the following parameters: Report alignments tailored for transcript assemblers including StringTie and searches for at most 1 distinct, primary alignments

for each read (Kim et al, 2019). SAM format files generated from mapping were converted to BAM format files using samtools-1.6 and to files in BED format with bamToBed utility from BEDTools-2.26.0. FPKM (Fragments Per Kilobase of transcript per Million) for genes were generated using StringTie-1.3.4d, with the following annotation: ensembl GRCh38 patch 94 .gff3 file, and parameters: limits the processing of read alignments to only estimate and output the assembled transcripts matching the reference transcripts given in annotation and excluding non-regular chromosomes, then normalized gene FPKM with RN7SL2. R package DESeq2 was applied to gene counts to perform differential gene expression analysis between no AD and AD individuals (with $P_{adj} < 0.05$ used as thresholds for reporting). For data visualization, statistics, and differential expression analysis, we employed R (version 3.4.3) with `r` in scatterplots corresponding to Pearson correlation coefficient (<https://www.R-project.org/>). In the boxplots, the central band (the line that divides the box into two parts) represents the median of the data, the ends of the box shows the upper (Q3) and lower (Q1) quartiles, the extreme line shows $Q3 + 1.5 \times IQR$ to $Q1 - 1.5 \times IQR$ (the highest and lowest value excluding potential outliers), while dots beyond the extreme line show potential outliers.

For the cellular deconvolution approach, we used the R (version 3.6.1) package MuSiC (Wang et al, 2019). This was implemented to estimate cell type proportions from MAP project data based on the Allen Brain Map Anterior Cingulate Cortex (ACC) SMART-seq dataset (<https://portal.brain-map.org/atlas-and-data/rnaseq/human-v1-acc-smart-seq>). Markers that are enriched in non-neuronal cells. For panel B image credit: Allen Institute. © [2015] Allen Institute for Brain Science. [Allen Mouse Brain Atlas]. Available from: [https://celltypes.brain-map.org/rnaseq/human_ctx_smart-seq] based on (Tasic et al, 2018; Hodge et al, 2019).

Gene expression analysis of microglia RNA-seq data (Synapse ID: syn11468526) from ROSMAP study was also conducted with the same analysis pipeline as described for MAP patients above, and subsequently, gene expression levels of the two datasets were compared.

Patients with available H3K9ac ChIP-seq data from ROSMAP study (Synapse ID: syn4896408) overlapping with tested MAP RNA-seq patients were identified, and filtered bam files against the provided peak file (H3K9acDomains.bed) were obtained with the intersectBed utility from BEDTools-2.26.0 (Quinlan & Hall, 2010). Then, metagene plots of the filtered bam files for the different gene lists were generated using Seqmonk.

3.4.8 Ethics approvals

CBB patient data analysis was performed under the following institutional approvals: University of Lethbridge Human Subject Research Committee (HSRC), protocol #2019-059 and Calgary Brain Bank provisions under the University of Calgary Conjoint Health Research Ethics Board Study ID: REB17-2289.

3.5 Results

3.5.1 Short RNA sequencing identifies processing at the right arm of Alu RNAs in human hippocampus

Much alike SINE B2 RNAs, SINE Alu RNAs can bind RNA polymerase II and suppress transcription. In the absence of an intact Alu RNA region binding Pol II, for example, after fragmentation of the Alu RNA, this binding and suppression of Pol II cannot take place (Fig 1A;

Yakovchuk et al, 2009; Ponicsan et al, 2010). This is of particular interest, given that we have recently shown that Alu RNAs, as in the case of B2 RNAs, are self-cleaving RNAs (Hernandez et al, 2020). In mouse, we have shown that processing of SINE RNAs in both NIH/3T3 cells and mouse hippocampus takes place at a position within the RNA region that binds and suppresses RNA polymerase II (Cheng et al, 2020). Interestingly, previous studies have identified a similar Pol II binding and suppression region also in Alu RNA. In particular, Alu RNA consists of two parts (arms) and the Alu RNA's 3' part (right arm), including the nucleotide bridge connecting the two arms, has been reported as the part of the Alu RNAs that is necessary for binding and suppression of RNA polymerase II (Mariner et al, 2008). Thus, we questioned whether Alu RNAs are processed within their Pol II binding and repression region also in human brain.

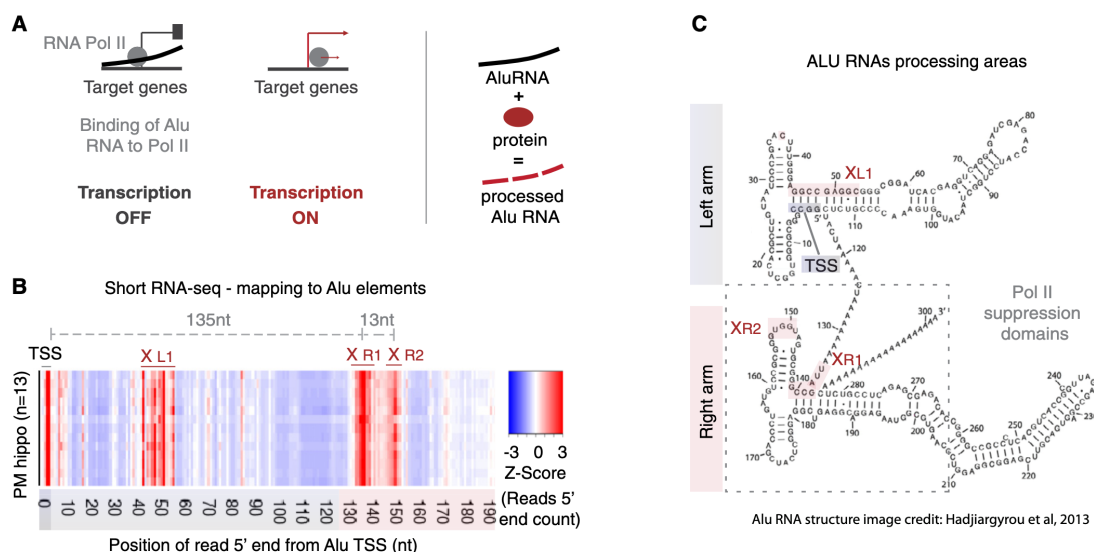


Figure 1. Processing areas of Alu RNAs in the human hippocampus as revealed by short RNA-seq

(A) Mode of regulation of transcription by Alu RNAs through suppression of RNA polymerase II (left panel) and protein-accelerated self-cleaving processing properties of Alu RNAs (right panel) as described in previous studies (Mariner et al, 2008; Yakovchuk et al, 2009; Ponicsan et al, 2010; Hernandez et al, 2020).

(B) Plotting of the position of the 5' end of Alu RNA fragments across the Alu metagene to depict potential processing areas of Alu RNAs in all post-mortem hippocampal tissues (PM hippo) from patients from the CBB. Each row in the heatmap depicts the distribution of counts of the 5' ends of reads mapped across the Alu metagene for each patient. The x-axis represents a metagene

combining all unique Alu RNA sequences (ALUome) aligned at the start site of their consensus sequence with numbers representing the distance from the transcriptional start site (TSS) area. Heatmap density corresponds to normalized counts of the 5' end of the reads with red corresponding to higher density of these 5' ends at a specific position. XL1, XR1, and XR2 denote the Alu processing areas defined by the high-density areas in the heatmap at specific positions of the Alu metagene, with the middle letter corresponding to the arm of the Alu RNA (see C, L = left, R = right), in which the area is located.

(C) Processing areas of Alu RNAs on the secondary Alu RNA structure. Secondary structure of Alu RNA adapted from Hadjiargyrou and colleagues (Hadjiargyrou & Delihis, 2013). As in our previous studies (Zovoilis et al, 2016; Cheng et al, 2020), we depict the SINE RNA processing areas (highlighted in pink) based on short RNA-seq data and mapping of the 5' ends of Alu RNA fragments. X mark the cleavage sites of Alu RNA that correspond to enriched processing areas (the high densities of 5' end fragments distribution) at the heatmap of (B). The rectangle depicts the critical region that binds and suppresses RNA Pol II based on (Mariner et al, 2008) that may be affected due to processing points.

To test this, we extracted RNA from post-mortem hippocampal tissues from 13 individuals with no clinical signs of Alzheimer's disease from the Calgary Brain Bank (CBB). We then employed an RNA sequencing approach that has been customized for the sequencing of SINE RNAs and their fragments (short RNA-seq). This library construction approach reduces any potential bias introduced through RNA fragmentation, that is used in standard long RNA-seq protocols, which is optimal for the identification of short SINE Alu RNA fragments (< 200 nt) that may be produced by Alu RNA processing. The version of short RNA sequencing we have used for Alu RNAs is similar to the version used in our previous work for SINE B2 RNAs (Zovoilis et al, 2016), but modified to include also the full-length Alu RNAs that are approximately 300 nt long (see Materials and Methods).

In contrast to B2 elements studied before in inbred mouse models and cell lines, Alu elements show increased genomic sequence variability when tested at the population level, for example, in humans. To account for this, after short RNA-seq, we performed mapping of the sequenced reads not against the human reference genome but against a generated list of all unique Alu sequences

that are available in the UCSC Genome Browser (genome.ucsc.edu) Repeat Masker track (repeatmasker.org). Sequenced reads were mapped against the generated unique “ALUome” with an aligner normally used for genomic mapping (no splicing or soft-clipping). Despite being computationally more intensive, this approach also helps to eliminate potential biases due to truncated Alu forms, Alu copy number variations, insertions, and deletions that may exist in the genome of tested individuals as well as sequences that are generated through splicing and could intervene with the mapping process.

Subsequently, for every patient, an Alu metagene model was constructed, plotting cumulatively the distribution of the 5' ends of all sequenced RNAs aligning to the ALUome with regard to their distance from the Alu elements' start. These 5' end read distribution models for each patient across the Alu metagene are presented one above the other in the form of a heatmap in Fig 1B. Increased color density in the heatmap represents increased number of 5' ends of sequenced reads mapping to specific positions in the Alu metagene, thus enabling the study of both full-length Alu RNAs (mapping within the first 15 nt to account for the variability in sequence among Alus) and their fragments (mapping either at the transcription start site—TSS or beyond) (Fig 1B). This approach enabled the determination in our short RNA sequencing data of three processing areas at Alu RNAs. These processing areas are highlighted in the previously described structure of the Alu consensus sequence (Hadjigargyrou & Delihias, 2013) depicted in Fig 1C and annotated as X L(left arm)1, X R(right arm)1 and XR2. The processing areas include two positions (XR1 and XR2) located within the right arm and the Pol II binding region (depicted as a rectangle in Fig 1C). Due to the very high prevalence of small insertions/deletions in the ALUome with regard to Alu consensus sequence, in the current study we opted for the definition of processing areas rather than

processing points to present more accurately this level of sequence variability. The distances presented in Fig 1B correspond to the distance between the peaks of distributions around each high-density area, with the peak in the TSS region marking position 0 in the distance from TSS.

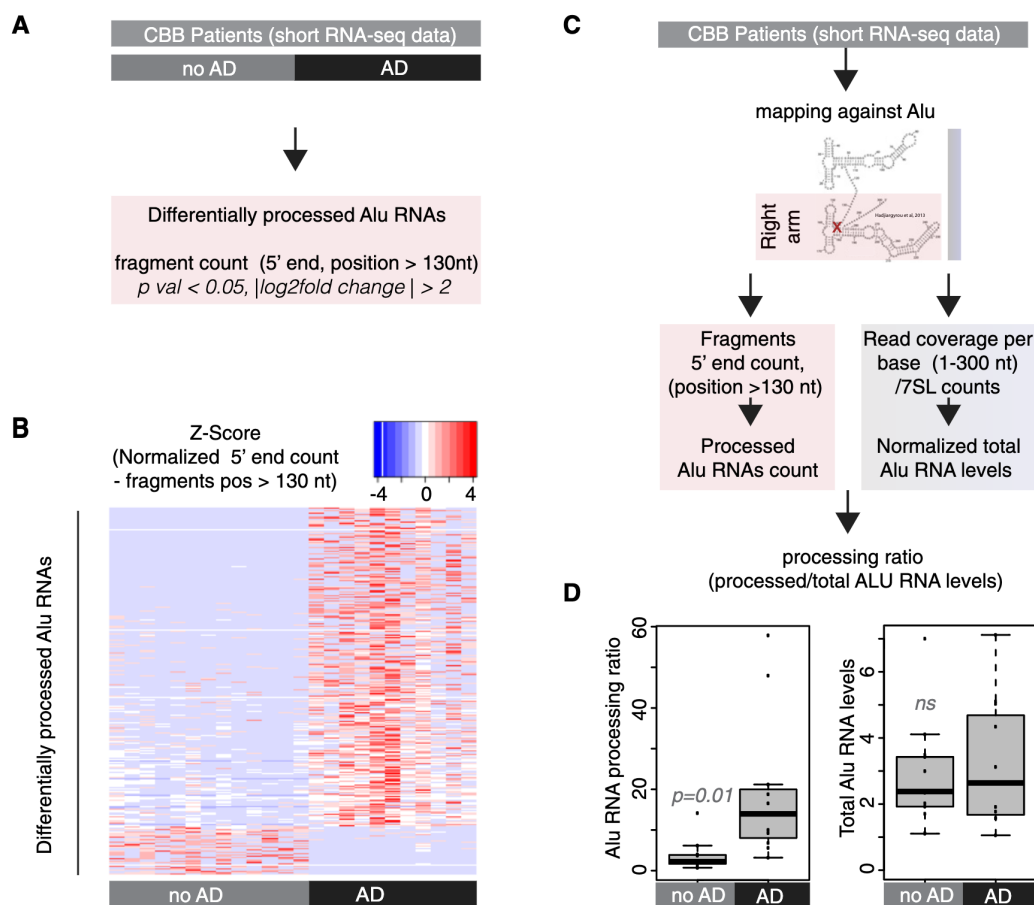
These data show that, in human hippocampus, SINE Alu RNAs are processed within their Pol II binding and suppression region, suggesting that the mode of regulation of gene expression through SINE RNA processing previously described in mouse may extend also to human brain.

3.5.2 Alu RNA processing is accelerated in hippocampi of AD patients

We have previously shown that in the hippocampi of a mouse model of amyloid beta pathology, SINE B2 RNA processing is accelerated. We questioned whether a similar mode of deregulation is present in humans. To this end, we employed post-mortem hippocampal tissue from a group of patients from the Calgary Brain Bank (CBB) clinically diagnosed with AD that was compared to the non-AD individuals tested above. Post-mortem delay was comparable between the two groups. De-identified clinical and pathological data for the CBB patients used in the current study are deposited as metadata in the controlled access next-generation sequencing repository (see Materials and Methods).

In total, we performed short RNA-seq in RNA from hippocampi from 24 patients, 13 with no clinical signs of AD and 11 diagnosed with AD (labeled as no AD and AD in Fig 2, respectively). RNAs were mapped against the ALUome as described above and 5' end counts of Alu RNA fragments mapping to the Pol II binding region of Alu RNA (right arm/positions > 130nt from Alu start) were calculated. Subsequently, we identified Alu RNAs that are differentially processed in

this region between AD and no AD patients (Fig 2A). These Alu RNAs are listed in Dataset EV1, while processing levels for these Alu RNAs for the two patient groups are presented as a heatmap in Fig 2B. As shown in Fig 2B, more than 80% of identified Alu RNAs are highly processed in hippocampi of AD patients, with these results being consistent within each group despite the



expected variability and heterogeneity among individuals.

Figure 2. Alu RNA processing ratio is increased in the hippocampi of AD patients in CBB cohort

(A) Analysis design for the identification of differentially processed Alu RNAs between no AD (n = 13) and AD (n = 11) patients of the CBB cohort using the DESeq2 R package. Statistical significance was estimated using the Wald significance tests within the DESeq2 package.

(B) Normalized counts of processed fragments mapping to the right arm of Alu RNAs that are differentially processed (rows) between AD and no AD for each CBB patient (columns) (Dataset EV1). Red corresponds to higher normalized counts of processed Alu RNA fragments.

(C) Analysis design for calculation of the Alu RNA processing ratio in CBB patients short RNA-seq data.

(D) Boxplots depict differences in hippocampi of AD and no AD patients regarding SINE Alu RNA processing ratio (left panel) (a P value of 0.05 was considered as threshold for statistical significance with $P = 0.01$, $n = 24$, unpaired non-directional t-test) and in total Alu RNA levels (right panel) (ns = not significant). In the boxplots, the central band (the line that divides the box into 2 parts) represents the median of the data, the ends of the box show the upper (Q3) and lower (Q1) quartiles, the extreme line shows $Q3 + 1.5 \times IQR$ to $Q1 - 1.5 \times IQR$ (the highest and lowest value excluding potential outliers), while dots beyond the extreme line show potential outliers.

These data revealed an increased number of Alu RNA fragments in AD patients, suggesting that the processing of Alu RNAs may be higher in these patients. In order to factor in any differences in the number of fragments due to changes in total Alu RNA levels, counts of Alu RNA processing fragments were normalized to total read coverage per base across the full-length Alu RNA sequence (from start to end) (Fig 2C). As in case of mouse short RNA-seq in our previous studies (Zovoilis et al, 2016; Cheng et al, 2020), in order to factor in any changes in these levels due to any SINE specific degradation of Alu RNAs, we normalized our data with a housekeeping Pol III non-coding RNA transcript (7SL RNA) the levels of which do not vary across AD and no AD patients (see Materials and Methods). This approach enabled us to estimate the Alu RNA processing ratio in AD and no AD patients, confirming substantially higher levels of Alu RNA processing in AD patients compared to no AD (Fig 2D, left panel) ($P = 0.01$, $n = 11$ (AD)/13 (no AD)), but not a change in total Alu RNA levels (Fig 2D, right panel).

Thus, hippocampi of AD patients are characterized by higher destabilization and processing ratios of Alu RNAs in the right arm including the RNA Pol II binding domain, consistent with the similar increased processing of B2 SINE RNAs previously described in the hippocampi of mouse models.

3.5.3 Standard RNA sequencing is able to detect the processing area in the right arm of Alu RNAs

Our strategy until now involved as a proof of principle the employment of short RNA sequencing as the method of choice for studying Alu RNA processing given its proven ability to detect SINE RNA fragments in mouse in the past. Our strategy also involved human hippocampi as the tissue of choice in order to maintain consistency with our previous SINE RNA studies data in mouse (Cheng et al, 2020; Fig 3A). However, when it comes to large-scale transcriptome studies in AD, this approach has a number of limitations. Firstly, short RNA-seq is not the first choice for study of mRNAs. In case of large transcriptome projects involving hundreds of patients compared to our small patient cohort, this approach would create the need for additional sequencing per sample, significantly increasing the cost. Moreover, a number of large-scale RNA-seq projects in AD patients have been completed using standard RNA-seq with a focus on protein coding RNAs. These studies have already generated an immense amount of publicly available data not only in hippocampus but also in other brain regions that are affected during AD, such as cortex. Thus, despite the increased transcript noise that RNA fragmentation in standard RNA-seq may cause, we were compelled to test whether available large-scale standard RNA-seq projects could enable identification of SINE Alu RNA processing at least at the Pol II binding region of the Alu RNA right arm.

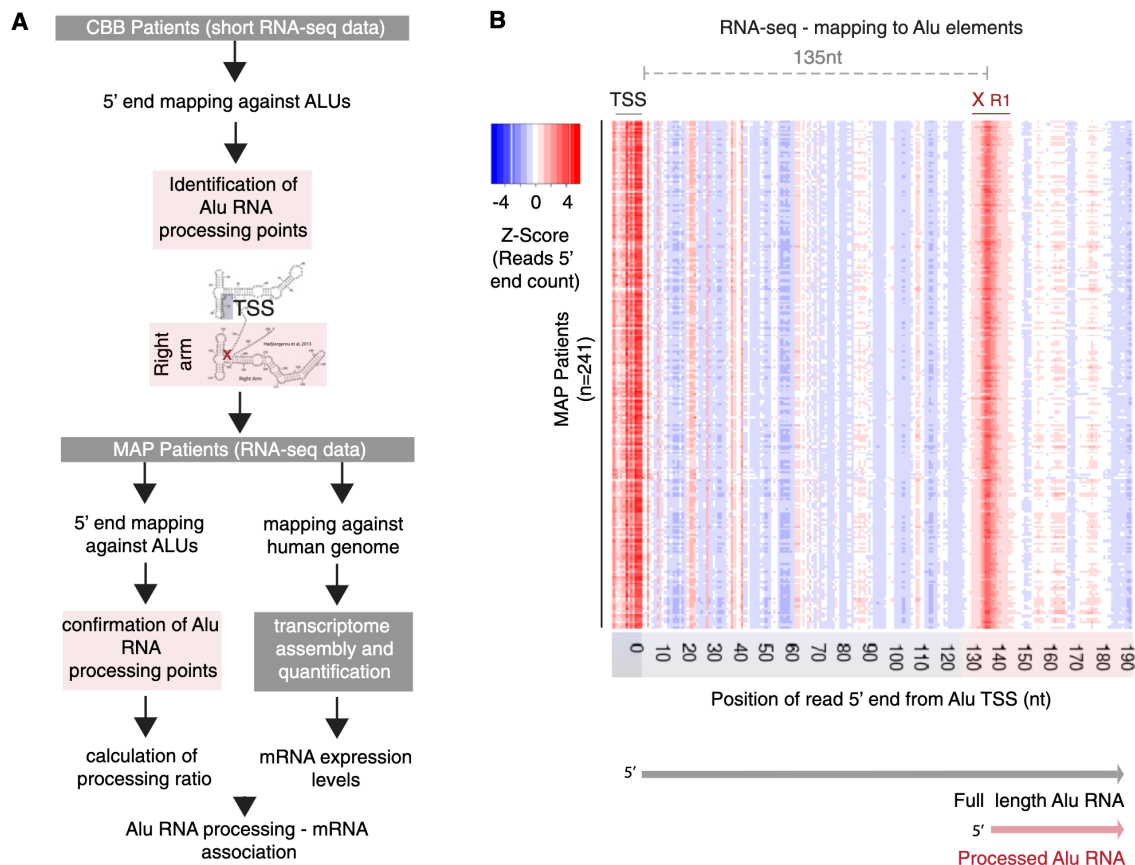


Figure 3. Processing areas of Alu RNAs in the human cortex as revealed by RNA-seq

(A) Two-tier analysis strategy and design for the confirmation of processing areas in Alu RNAs in human brain combining different sequencing approaches and patient cohorts (CBB cohort, n = 24; MAP cohort, n = 241).

(B) Plotting of the position of the 5' end of Alu RNA fragments across the Alu metagene to depict potential processing areas of Alu RNAs in all post-mortem cortex tissues from patients from the MAP cohort. As in Fig 1B, each row represents the distribution of normalized counts of the 5' ends of reads mapped across the Alu metagene for each patient. As in Fig 1B, x-axis numbers represent the distance from the transcriptional start site (TSS) area and heatmap density corresponds to normalized counts of the 5' end of the reads with red for higher density of these 5' ends at a specific position. XR1 as in Fig 1B.

In particular, we questioned whether publicly available poly-A selection-based RNA sequencing approaches in AD patients would allow the identification of Alu RNA processing in the Alu right arm. This could be possible in humans compared to mouse for a number of reasons. Firstly, Alu

RNAs in humans are much longer than B2 RNAs (300 nt vs. 180 nt). With a processing area at 135 nt downstream of TSS, the resulting Alu RNA fragment size would fall within the range of most standard RNA-seq protocols in contrast to shorter SINE B2 RNA fragments in mouse. Secondly, despite being non-coding, SINE RNAs contain long 3' end adenosine stretches. This would enable their sequencing in most standard RNA-seq projects that use poly(A) adapters or selection strategies. In addition, due to the poly(A) tail selection, this approach would enable filtering out processed transcripts at TSS and counting only of full-length Alu RNAs at this position in contrast to short RNA-seq which at TSS could include both types of transcripts (see Materials and Methods).

To test this, we employed RNA-seq data from one of the largest transcriptome projects in AD completed until now, the Memory and Aging Project (MAP) study that is part of the ROSMAP study. MAP is a longitudinal, epidemiologic clinical-pathologic cohort study with an emphasis on decline in cognitive and motor function and risk of Alzheimer's disease that began in 1997 and is run from Rush University (Bennett et al, 2012). The Broad Institute's Genomics Platform performed the RNA-seq (Mostafavi et al, 2018). The study employed poly(A)-based RNA sequencing for the study of mRNAs and includes RNA sequencing data from the gray matter of the dorsolateral prefrontal cortex, accompanied with various clinical and demographic metadata including clinical diagnosis, Braak staging, ApoE genotype, and age. Sequence data from this project were obtained from the AMP-AD Knowledge Portal (<https://adknowledgeportal.synapse.org>), and FASTQ files were reconstructed and then re-analyzed with the exact same pipelines used in our CBB patient cohort. In addition, given that these data offer the possibility to study also mRNA expression, we applied in parallel a typical

mRNA analysis pipeline, involving mapping against the genome (with splicing), transcriptome assembly, and quantification, in order to be able to associate our findings in SINE Alu RNAs with gene expression at a later stage (Fig 3A). Based on our preliminary quality control, we identified potential batch effects in samples sequenced within a specific time period (year 2013), and these samples were excluded from further analysis. In total, data from 241 MAP patients were analyzed.

As shown in Fig 3B, mapping of the 5' ends of reads aligning to the ALUome across the Alu metagene confirmed that Alu RNAs can indeed be poly(A) selected and sequenced through this approach generating clear strong read distribution density at the TSS region at the beginning of the Alu elements. Most importantly, it revealed the exact same processing area (XR1—135nt downstream of TSS) we have been able to identify through short RNA-seq above, suggesting that generation of this fragment is one of the first steps of processing of the initial poly(A) full-length Alu RNA. In contrast, both XL1 and XR2 areas identified in short RNA-seq were not prominent in standard RNA-seq, suggesting that these are processing fragments without the 3' poly(A) end, that can be recognized only with methods such as short RNA-seq that detect both poly(A) and non-poly(A) fragments. The level of transcriptional noise expected due to library RNA fragmentation was low and amenable, with none of the other density areas competing the density at XR1. Moreover, identification of the XR1 processing area was consistent across all MAP patients tested (Fig 3B).

These data show that study of Alu RNA processing is possible also when using standard RNA-seq techniques, expanding significantly our potential to investigate SINE RNAs in already completed large transcriptome studies in human brain.

3.5.4 Alu RNA processing is accelerated in the cortex of AD patients

As in the case of hippocampus, cortex regions are among the primary targets of amyloid pathology in AD. To this end, we investigated whether Alu RNA processing ratio is increased in the dorsolateral prefrontal cortex of AD MAP patients similarly to CBB hippocampal tissues. MAP includes patients with a wide range of clinical diagnosis, including no cognitive impairment (NCI), mild cognitive impairment (MCI) and AD, as well as of patients classified in all Braak stages (from 0 to VI). Similar to our CBB cohort, patients from the MAP cohort were separated into two categories, no AD and AD, based on clinical diagnosis and Braak staging (see Materials and Methods).

As in the case of hippocampus, based on RNAs mapped against the ALUome, we calculated 5' end counts of Alu RNA fragments mapping to the Pol II binding region of Alu RNA (right arm/position > 130nt) of MAP patients. Subsequently, we identified Alu RNAs that are differentially processed in this region between AD and no AD patients (Fig 4A). Resembling our observations in the hippocampus, the vast majority of differentially processed Alus (> 80%) depict increased processing levels in the dorsolateral prefrontal cortex of the AD patient group. These Alu RNAs are listed in Dataset EV2 and are presented for all MAP AD and no AD patients in the heatmap of Fig 4B. In contrast to short RNA-seq, in standard RNA-seq, due to poly(A) selection, 5' end counts at TSS correspond only to full-length Alu RNAs. Thus, we have been able to calculate the Alu processing ratio factoring in the full-length Alu RNA levels as described in Fig 4C. As in case of hippocampal samples, the Alu RNA processing ratio is significantly elevated in the dorsolateral prefrontal cortices of AD patients (Fig 4D left panel) ($P < 0.001$, $n = 67$ (AD)/51

(no AD)). Full-length Alu RNA levels were also slightly increased, but not to the extent of the processing ratio (Fig 4D, right panel, $P = 0.005$).

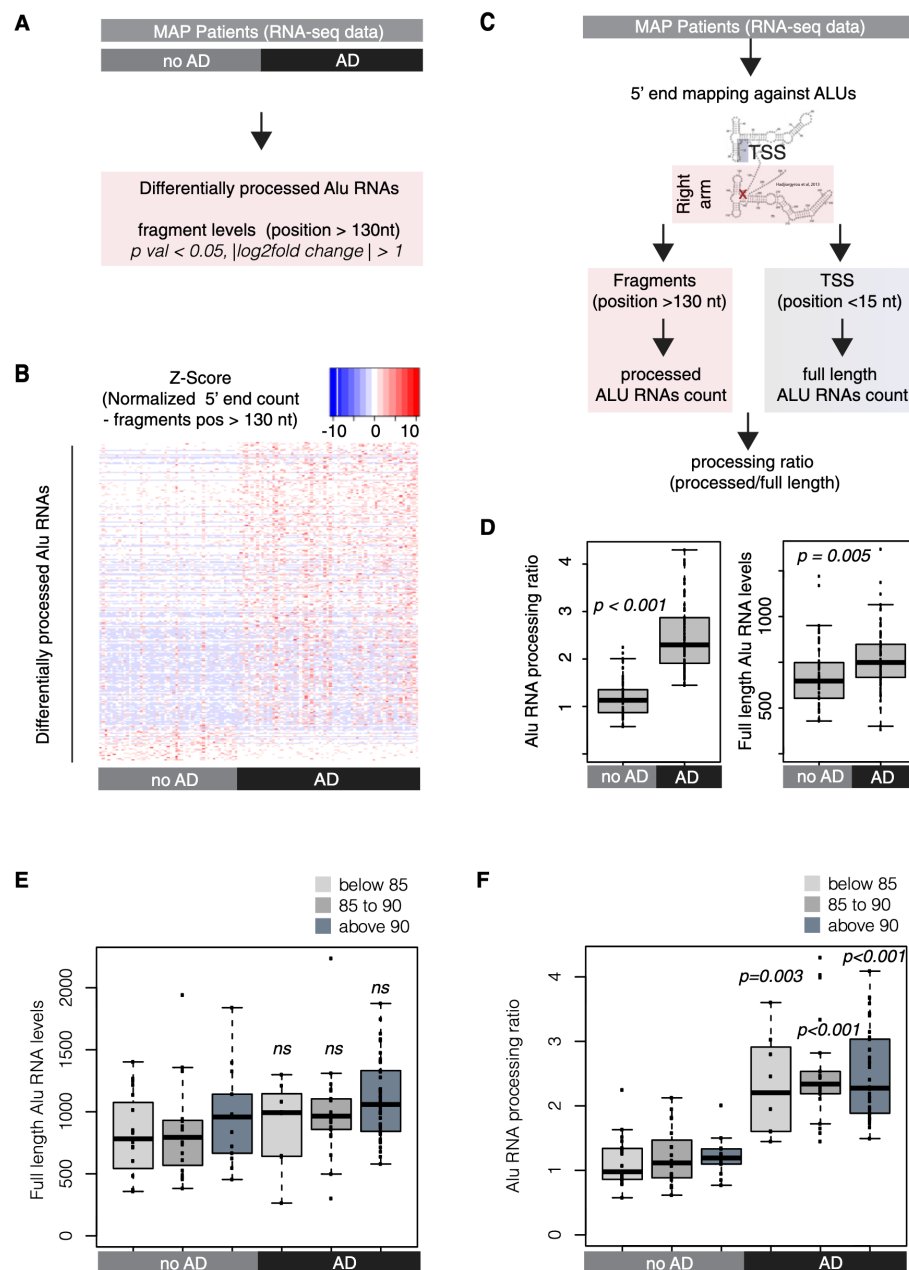


Figure 4. Alu RNA processing ratio is increased in the cortex of AD patients in the MAP cohort

(A) Analysis design for the identification of differentially processed Alu RNAs between AD (n = 67) and no AD patients (n = 51) of the MAP cohort using the DESeq2 R package. Statistical significance was estimated using the Wald significance tests within the DESeq2 package.

(B) As in Fig 2B, normalized counts of processed fragments mapping to the right arm of Alu RNAs that are differentially processed (rows) between AD and no AD for each MAP patient (columns) (Dataset EV2). Red corresponds to higher normalized counts of processed Alu RNA fragments.

(C) Analysis design for calculation of the Alu RNA processing ratio in MAP patients' RNA-seq data.

(D) Boxplots depict differences in cortex of AD and no AD patients regarding SINE Alu RNA processing ratio (left panel) (a P value of 0.05 was considered as threshold for statistical significance with $P < 0.001$, $n = 118$, unpaired non-directional t-test) and in full-length Alu RNAs ($P = 0.005$). Boxplot representation as in subfigure 2D.

(E) Boxplots depict full-length Alu RNA levels in cortex of AD and no AD patients separated into three age groups. No significant difference observed between the different age groups of either AD or no AD patients or for the comparisons between no AD and AD of each age group (unpaired non-directional t-test, with $P < 0.05$ considered the threshold for statistical significance, n for no AD = 19 (below 85), 19 (between 85–90), 13 (above 90); n for AD = 8 (below 85), 21 (between 85–90), 38 (above 90)). Boxplot representation as in subfigure 2D.

(F) Boxplots depict differences in SINE Alu RNA processing ratio in cortex of AD and no AD patients separated into three age groups. No significant difference observed between the different age groups of either AD or no AD patients. A P value 0.05 was considered as threshold for statistical significance for the comparisons between no AD and AD of each age group (unpaired non-directional t-test, with $P = 0.003$ for the comparison below 85, and $P < 0.001$ for the other two comparisons and n numbers as in (E)). Boxplot representation as in subfigure 2D.

We then questioned whether our results may be influenced by the patient's age. This is of particular interest given that aging constitutes one of the main risk factors for AD. As shown in Fig 4E and F, increased Alu RNA processing ratio is a distinct characteristic of AD patients independently of their age, while no difference was observed regarding full-length Alu RNA levels.

We then decided to investigate the association of Alu RNA processing ratio with additional clinical and molecular indicators of AD included in the dorsolateral prefrontal cortices of patients of the MAP study. In particular, the MAP patient cohort includes also a number of patients with clinical diagnosis or pathology data in-between the two categories of no AD and AD patients tested above (for example, patients with MCI, and/or intermediate Braak stage). Thus, we questioned whether changes in the Alu RNA processing ratio can mirror differences in the clinical diagnosis and the Braak staging not only of AD and no AD patient categories but of all MAP patients, including

those in the middle of the assessed clinical range. Indeed, as shown in Fig 5A and B, processing ratios are gradually being elevated as we move from low to higher severity cases both regarding clinical diagnosis and Braak staging. Moreover, the MAP study includes information about additional molecular genetics indicators, such as APOE genotypes. APOE gene is associated with the risk of developing late-onset Alzheimer's disease, with the variant APOE e4 (E4) conferring a higher risk (Green et al, 2009; Genin et al, 2011). Interestingly, as shown in Appendix Fig S2A, when all MAP patients are sorted based on their Alu RNA processing ratio (upper panel of Appendix Fig S2A), and the presence of the E4 variant in the genotype of each of them is plotted (lower panel of Appendix Fig S2A), we observed higher Alu RNA processing ratios in most patients with this variant.

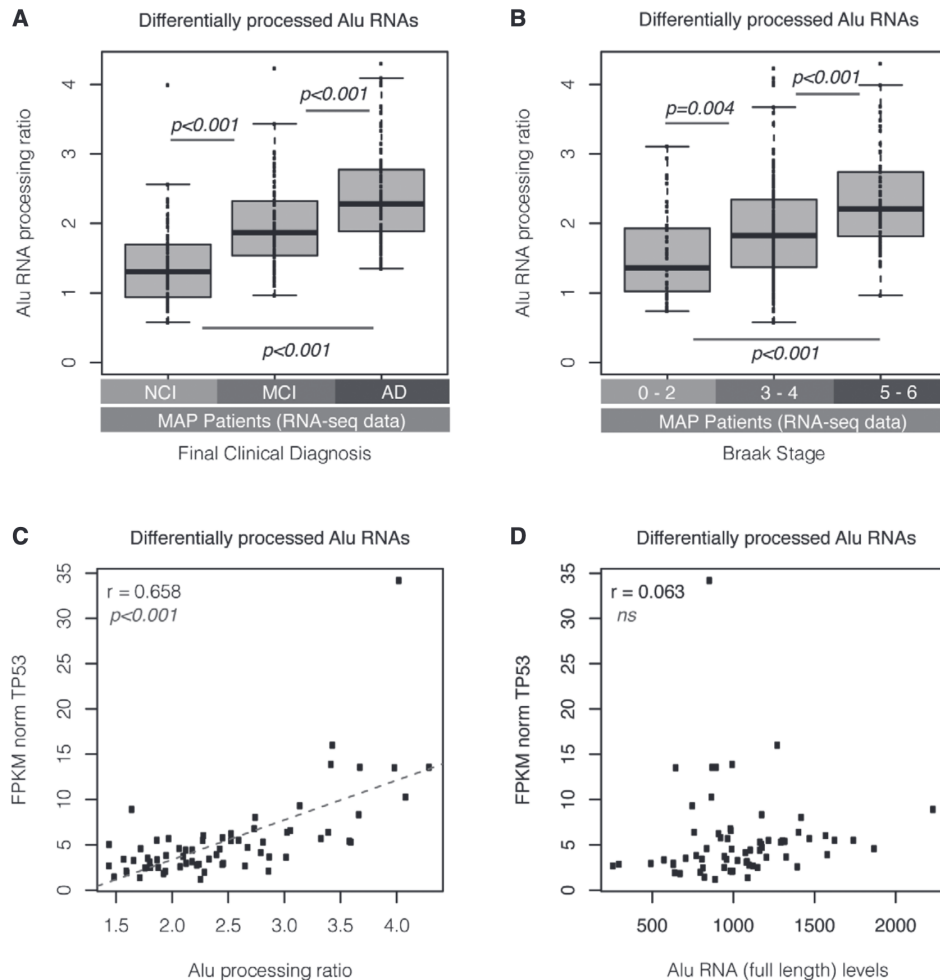


Figure 5. Alu RNA processing ratio is associated with clinical, pathology, genetic, and molecular markers of AD in MAP patients

(A) Boxplots depict differences in SINE Alu RNA processing ratio in all MAP patients with regard to final clinical diagnosis ($P < 0.01$ for the comparison between NCI and MCI, between NCI and AD, and between MCI and AD, unpaired, non-directional t-test). NCI = no cognitive impairment ($n = 71$), MCI = mild cognitive impairment ($n = 67$), AD = Alzheimer's disease ($n = 87$). Boxplot representation as in subfigure 2D.

(B) Boxplots depict differences in SINE Alu RNA processing ratio in all MAP patients ($n = 241$) with regard to Braak stage classified into three categories ($P = 0.004$ for the comparison between 0–2 and 3–4, and $P < 0.001$ for the comparisons between 3–4 and 5–6 and between 0–2 and 5–6, unpaired, non-directional t-test, $n = 39$ [0–2], $n = 148$ [3–4] and $n = 54$ [5–6]). Boxplot representation as in subfigure 2D.

(C) Scatterplot depicting the positive correlation between TP53 expression values and Alu RNA processing ratio in AD patients ($r = 0.65$, $P < 0.001$). Statistical test is based on Pearson's product moment correlation coefficient and follows a t distribution using the cor.test function in R package stats.

(D) Scatterplot depicting no correlation between TP53 expression values and full-length Alu RNA levels in AD patients ($r = 0.06$, no correlation, ns = non-significant). Statistical test as in subfigure 5C.

The above data suggest that Alu RNA processing is increased also in the cortex of AD patients and it is associated with major clinical, genetic, and molecular pathology markers of AD.

3.5.5 Changes in Alu RNA processing ratio are associated with changes in P53 levels

Neural cell death leading to brain atrophy is a hallmark of AD disease. P53 (TP53) constitutes one of the major markers of cell death, and P53 deregulation has been connected with AD molecular pathology, soluble amyloid beta oligomers, and tau hyperphosphorylation (reviewed at Jazvinskak Jembrek et al, 2018; Chang et al, 2012). In our previous study in the mouse (Cheng et al, 2020), we found that SINE B2 RNA processing is associated with high P53 levels, likely as a result of consistently high levels of stress response genes upstream of P53 that are activated by increased B2 RNA processing. Since the MAP patient data included levels of mRNAs in these tissues, we questioned whether a similar association between P53 (TP53) levels and Alu RNA processing ratio exists also in humans. The scatterplot of Fig 5C depicts this relationship, revealing a moderately strong correlation between P53 levels and Alu RNA processing ratio ($r = 0.65$, $P < 0.001$) in AD patients. This strong correlation was only observed with Alu RNA processing ratio (that is indicative of increased gene activation) and not with levels of full-length Alu RNAs (Fig 5D) (that are indicative of suppression of gene expression). Similar results were observed, as shown in Appendix Fig S2B, when we investigated the relationship between P53 and Alu RNA processing ratio in all MAP patients, which confirmed the association between the expression levels of this gene and Alu RNA processing ($r = 0.58$, $P < 0.001$) but no association with the full-length Alu RNA levels (Appendix Fig S2C).

These data raised the question whether gene expression of members of pathways that are upstream of P53 is also associated with Alu RNA processing ratio.

3.5.6 Changes in Alu RNA expression and processing are associated with changes in gene expression

As shown above, changes in the Alu RNA processing ratio correlate with P53 levels in MAP patients, which is consistent with our previous studies in mouse fibroblasts and neural cells that connected the activation of gene expression in pro-apoptotic/pro-survival pathways with SINE RNA processing (Zovoilis et al, 2016; Cheng et al, 2020). We decided to investigate the global gene expression in MAP AD patients and search for possible associations between SINE RNA processing ratio and transcriptome changes as it has been previously shown for mouse. To this end, we first analyzed the MAP mRNA data and identified genes that are differentially expressed between AD and no AD patients. These genes are divided into two lists including genes that are up-regulated in AD patients (up-regulated genes, Dataset EV3) and genes that are down-regulated in AD patients (down-regulated genes, Dataset EV4).

Should the mode of regulation for activated genes in humans be the same as the one described in mouse (Appendix Fig S1), we would expect a strong positive correlation between high Alu processing ratio and gene expression of up-regulated genes (Fig 6A, left panel). In the same way, should the mode of regulation for suppressed genes be the same as the one previously described in mouse, we would expect a negative correlation between high full-length (un-processed) Alu RNA levels and gene expression of down-regulated genes (Fig 6A, right panel). To test these two hypotheses, we sorted the patients based either on their Alu RNA processing ratio (Fig 6B, upper panel) or their full-length Alu RNA levels (Fig 6C, upper panel), respectively, while in the lower

panels, we depicted gene expression levels of up- and down-regulated genes, for each of these patients sorted in the same order as the sorted processing ratio or full-length levels (Fig 6B and C lower panels, respectively).

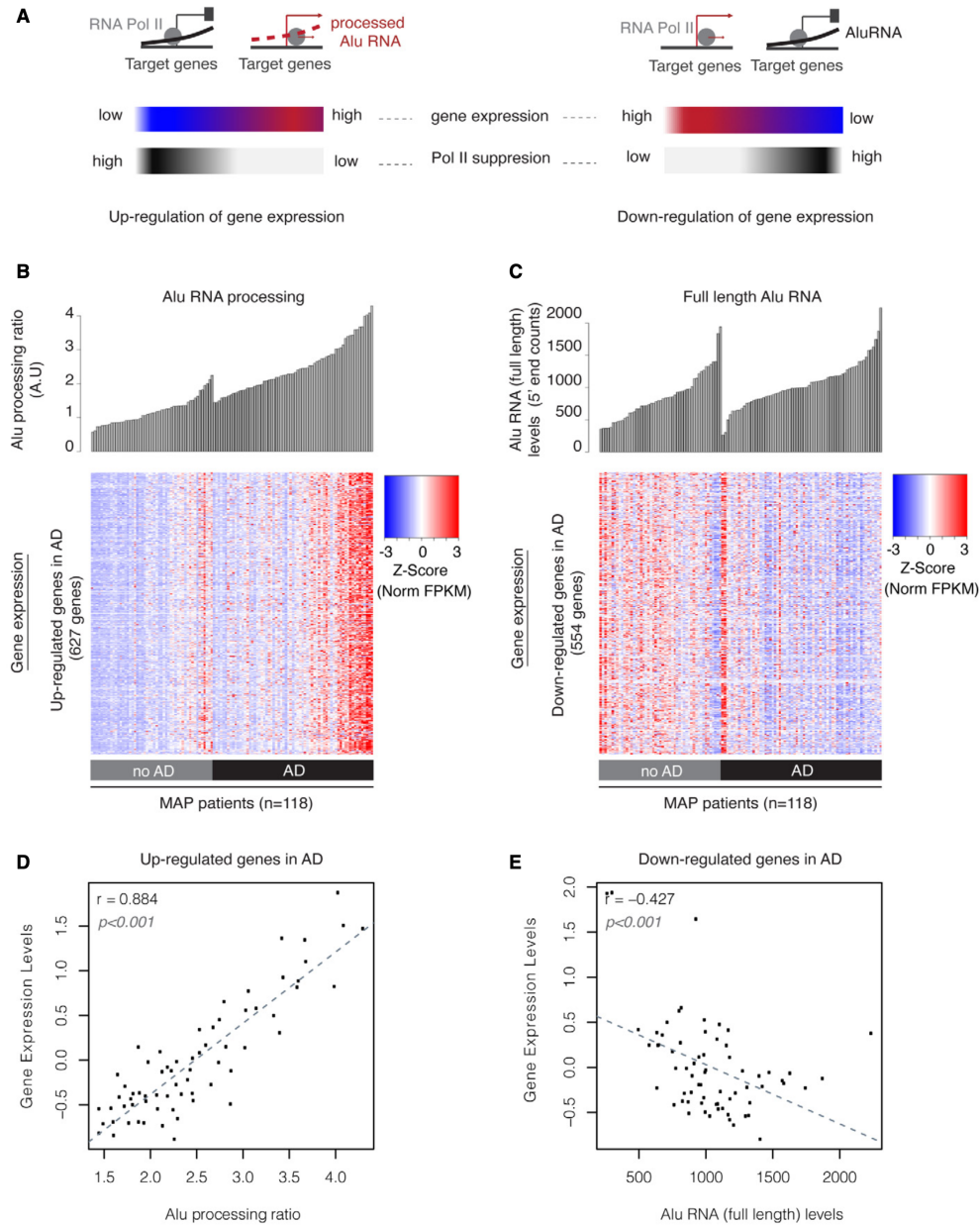


Figure 6. Alu RNA expression and processing levels are associated with transcriptome changes in the cortex of MAP patients

(A) Graphical representation of the expected transcriptome changes with regard to Alu RNA expression and processing levels based on previous findings in mouse (Appendix Fig S1) and prior

reports on the ability of Alu RNAs to control transcription (Mariner et al, 2008; Yakovchuk et al, 2009; Ponicsan et al, 2010) and become self-cleaved (Hernandez et al, 2020).

(B) Association between SINE Alu RNA ratio (upper panel) and gene expression levels (lower panel) of genes up-regulated in AD (Dataset EV3, $|\log_2\text{FoldChange}| > 0.5$). Every column in both panels corresponds to the same MAP patient of either the no AD or AD group. Patients in each group are sorted from left to right in an ascending order with regard to Alu RNA processing ratio. Every row in the heatmap corresponds to one gene, with color density representing normalized FPKM values from RNA-seq for this gene for each patient (columns) (red represents high and blue low expression levels).

(C) Association between full-length Alu RNA levels (upper panel) and gene expression levels (lower panel) of genes down-regulated in AD (Dataset EV4, $|\log_2\text{FoldChange}| > 0.5$). Every column in both panels corresponds to the same MAP patient of either the no AD or AD group. Patients in each group are sorted from left to right in an ascending order with regard to full-length Alu RNA levels. Every row in the heatmap corresponds to one gene, with color density representing normalized FPKM values from RNA-seq for this gene for each patient (columns) (red represents high and blue low expression levels).

(D) Scatterplot depicting the positive correlation between average gene expression values of up-regulated genes in Fig 6B and Alu RNA processing ratio in AD patients ($r = 0.88$, $P < 0.001$). Statistical test is based on Pearson's product moment correlation coefficient and follows a t distribution using the cor.test function in R package stats.

(E) Scatterplot depicting the negative correlation between average gene expression values of down-regulated genes in Fig 6C and full-length Alu RNA levels in AD patients ($r = -0.42$, $P < 0.001$). Statistical test as in subfigure 6D.

As shown in Fig 6B, consistent with our hypothesis, increase of Alu RNA processing ratio correlates with gradual increase in the expression levels of up-regulated genes. In fact, gene expression levels of up-regulated genes are strongly correlated with Alu RNA processing ratio in AD patients (Fig 6Dr $= 0.88$, $P < 0.001$). No such correlation is observed between up-regulated genes and full-length Alu RNA levels (Appendix Fig S3), in which sorting based on full-length Alu RNA levels generates a rather random distribution of gene expression densities across patients. In contrast, as shown in Fig 6C and E, full-length Alu RNA levels correlate negatively with gene expression levels of down-regulated genes ($r = -0.42$, $P < 0.001$), in accordance with our hypothesis, while as shown in Appendix Fig S4, there is a much weaker correlation when the relationship of these genes with Alu RNA processing ratio is tested ($r = -0.26$).

The changes in expression levels in AD relevant genes reported by RNA-seq correspond to final mRNA levels in the cell and not to the initial levels transcribed by Pol II. These initially transcribed levels may have been subject to a cascade of potential post-transcriptional modifications and processing that may have affected stability and half-life of the mRNAs. Thus, final mRNA levels reported by RNA-seq may not correspond completely to the elongation activity of RNA polymerase II at the chromatin level, for which we are mostly interested in the current study. To this end, we have also employed ChIP-seq data for histone 3 lysine 9 acetylation (H3K9ac) that are available for the same patients. H3K9ac is a chromatin mark for the switch of RNA Pol II from transcription initiation to elongation (Gates et al, 2017). Thus, testing this chromatin mark corresponds very well to RNA Pol II elongation activity. As shown in Appendix Fig S5, H3K9ac occupancy downstream of the transcription start site is significantly increased in AD patients for AD up-regulated genes compared to (i) down-regulated genes (Appendix Fig S5A) or (ii) a random set of non-AD differentiating genes (Appendix Fig S5B). This finding supports that there is an increase in Pol II elongation in AD up-regulated genes.

These data show that transcriptome changes in SINE RNA are associated with transcriptome and chromatin-wide changes observed in brains of AD patients. Targeting of Alu RNAs leads to activation of those AD up-regulated genes that are strongly associated with Alu RNA processing.

We then questioned to what extent the observed transcriptome-wide changes in AD patients are connected with Alu RNAs ability to suppress transcription, which would justify the increase of expression of potential target genes in AD due to increased Alu RNA processing and alleviation of such suppression.

Despite the overall positive correlation between gene expression and Alu RNA processing ratio in AD patients described in Fig 6, it is unlikely that this correlation applies to all 2,860 genes that were found to be up-regulated in AD (Dataset EV3). To clarify this point, we estimated the correlation coefficient (r) between gene expression and Alu RNA processing of each of these 2,860 genes as well as for a random set of genes that are not differentially expressed in AD as reference. Subsequently, we classified tested genes in two main categories: one including results either non-significant ($P \text{ val} < 0.05$) or with weak/no correlation ($r \leq 0.5$) and one including results that are both statistically significant and depict strong or very strong correlation (r more than 0.5). As shown in Fig 7A, in accordance with our findings in Fig 6, the vast majority of AD up-regulated genes (88%) were classified to the significant and strong correlation group compared to only 26% in the control set of genes.

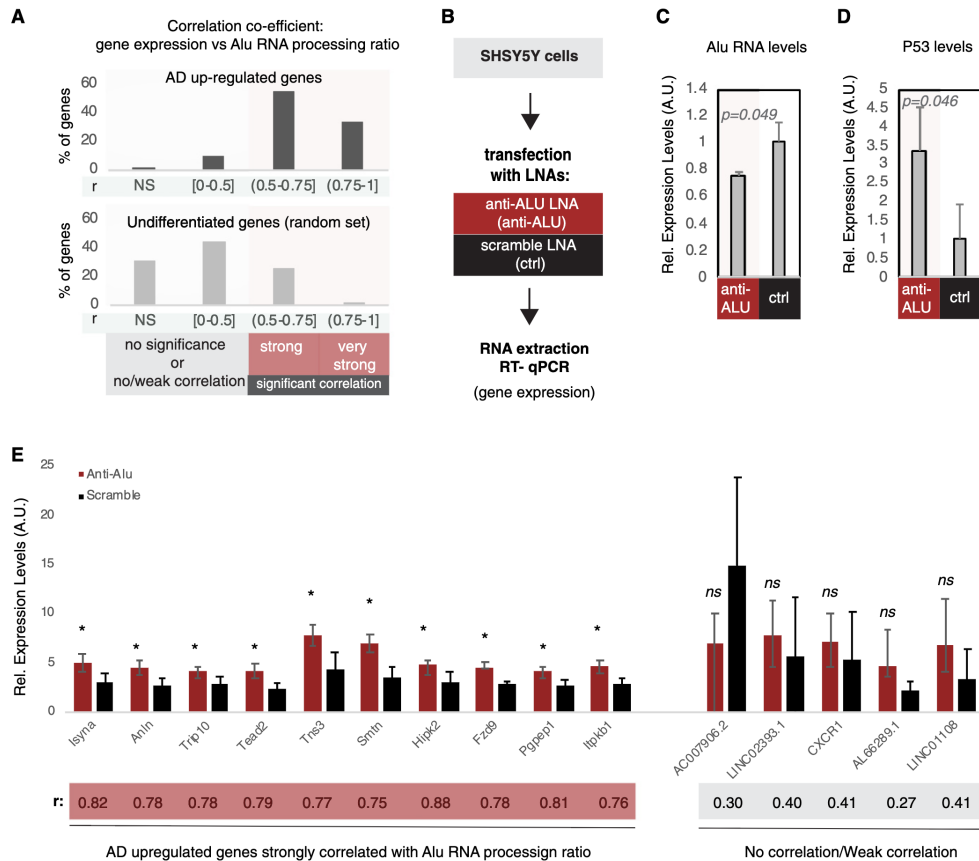


Figure 7. Alu RNA destabilization leads to increase in expression of Alu RNA processing correlated genes

(A) Correlation between gene expression and Alu RNA processing ratio for AD up-regulated genes (Dataset EV3) (upper panel) and a random set of non-differentially expressed genes (Dataset EV5) (lower panel). For every gene, a correlation coefficient was calculated (Pearson correlation) together with the respective P-value. Statistical test is based on Pearson's product moment correlation coefficient and follows a t distribution using the cor.test function in R package stats. Based on the r and P value, genes were classified into four categories: (i) NS, non-significant for P value > 0.05, (ii) no correlation or weak correlation ($r \leq 0.05$), (iii) significant ($P < 0.05$) and strong correlation ($0.5 < r \leq 0.75$), and (iv) significant ($P < 0.05$) and very strong correlation ($0.75 < r$). The bar graphs represent the percentage of each category for each set of genes. The exact r and P values of each gene are listed in Dataset EV6 for the upper panel and Dataset EV7 for the lower panel.

(B) Experimental design for targeting Alu RNAs in a cell culture assay employing SHSY5Y neural cells.

(C) Expression levels of full-length Alu RNA (RT-qPCR) in the Alu RNA KD experiment. Statistical significance (P value threshold 0.05) for the comparison between anti-Alu LNA-treated samples (anti-Alu) and samples treated with a scramble control LNA (ctrl) with $P = 0.049$ and $n = 3$, unpaired non-directional t-test. Error bars represent standard deviation from the mean.

(D) Expression levels of P53 (RT-qPCR) in the Alu RNA KD experiment. Statistical significance (P value threshold 0.05) for levels in anti-Alu LNA-treated samples (anti-Alu) greater than samples treated with a scramble control LNA (ctrl) with $P = 0.046$ and $n = 3$, unpaired directional t-test. Error bars represent standard deviation from the mean.

(E) Expression levels (RT-qPCR) of selected genes from panel A that are either strongly correlated with Alu RNA processing (left, $r > 0.5$) or only weakly/not correlated (right, $r < 0.5$). Statistical significance (P value threshold 0.05) for anti-Alu greater than control (depicted as asterisk, $n = 3$ /group, unpaired directional t-test, error bars represent standard deviation from the mean). Pearson correlation coefficient for each gene is depicted below the name of each gene.

Gene ontology and pathway analysis of the AD genes identified above to be strongly correlated with Alu RNA processing revealed a statistically significant enrichment ($P_{adj} < 0.05$) for terms directly related with response to cellular stress, cell proliferation, cell death, P53 function, and regulation of transcription and RNA polymerase activity (Dataset EV8). In particular, KEGG pathway analysis of these genes (Dataset EV8, middle) revealed enrichment of terms for a number of cell signaling pathways upstream of P53, originally described in cancer, but subsequently attributed also to regulation of learning and memory, chromatin and synaptic plasticity and AD pathogenesis in brain. Among these pathways are MAPK and PI3K-AKT signaling pathways (Appendix Fig S6) as well as the RAS and HIPPO signaling pathways (Appendix Fig S7). These results show that not only P53 but also members of upstream regulatory pathways of P53 are strongly associated with Alu RNA processing. Moreover, Gene Ontology term enrichment analysis of Alu RNA processing correlated genes (Dataset EV8, right) revealed among the top enrichment terms, multiple terms related with regulation of transcription from RNA polymerase II.

We then questioned whether by inducing an artificial degradation of Alu RNAs, we would be able to induce expression of these Alu RNA processing correlated genes in a neural cell culture model frequently used in AD molecular pathology experiments (SHSY5Y cells; Jazvinscak Jembrek et al, 2018). To achieve this, we employed a similar approach that we used in our previous study in

mouse (LNAs against SINE RNAs). Application of a set of LNAs against Alu RNAs was able to reduce levels of Alu RNAs compared to the control LNA (Fig 7B and C). Similarly to mouse neural cells and SINE B2 RNAs, targeting of the Alu RNAs in SHSY5Y cells resulted in the increase of the expression levels of selected AD up-regulated genes that are strongly correlated with Alu RNA processing ratio and also of P53 levels (Fig 7D and E). The increase in gene expression occurred in the absence of any stress stimulus, suggesting that these genes are under the suppressive control of Alu RNAs. At the same time, Alu RNA destabilization did not affect expression of 5 AD up-regulated genes with a weak or no correlation to Alu RNA processing that were used as negative controls.

3.5.7 HSF1 can accelerate Alu RNA processing in vitro

We have shown before that SINE B2 and SINE Alu RNA processing can be attributed to an endogenous self-cleaving ability of these RNAs that is accelerated in the presence of a protein called Ezh2 in a riboswitch manner (Hernandez et al, 2020). Subsequently, we expanded the list of B2 RNA processing accelerators to include a protein that is key in cellular response to stress called heat shock factor (Hsf1) and we showed that increased SINE RNA processing during amyloid beta toxicity in neural cells in mouse can be at least partially attributed to increased levels of Hsf1 (Cheng et al, 2020).

Thus, we questioned whether HSF1 in humans is also a potential accelerator of SINE Alu RNA processing as it is for SINE B2 RNAs in mouse. To test this, we incubated the Alu RNA in the presence of the HSF1 protein in vitro under the same conditions used in our previous studies. Then, we measured levels of the remaining full-length Alu RNA over time in the presence and the

absence of HSF1. As shown in Fig 8A, incubation of the full-length Alu RNA in vitro led to processing and destabilization of the Alu RNA in the presence of HSF1. We then tested the impact on acceleration of Alu RNA processing by Hsf1 in the case of heat denaturation of Hsf1. As shown in Fig 8B, heat denaturation of Hsf1 resulted in abrogation of its ability to accelerate Alu RNA processing (Fig 8B). Similar results were obtained with the incubation of Alu RNA with another RNA binding protein of similar size to Hsf1, poly-A polymerase (Hsf1 ~ 60 kDa, poly-A polymerase ~ 56 kDa) (Fig 8B). In order to take into account any RNA destabilization due to non-specific degradation, hydrolysis or Alu RNA endogenous self-cleavage, we also incubated Alu RNA in the absence of any HSF1 but in the same buffer and for the same time as the sample with the HSF1 protein, and we also used as controls other RNAs in parallel reactions. In the presence of HSF1, none of these RNAs depicted the destabilization observed in Alu RNAs (Fig 8C).

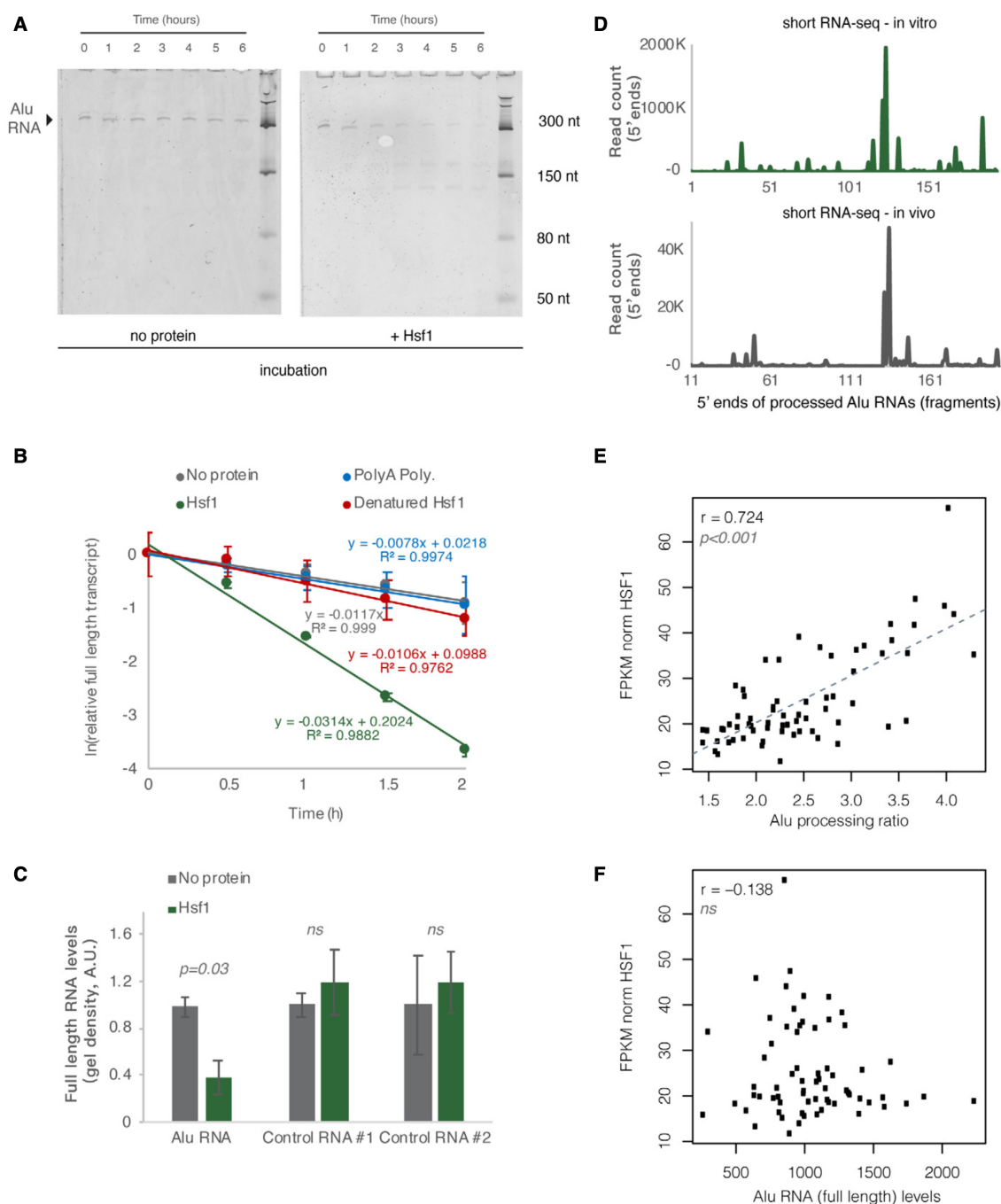


Figure 8. HSF1 accelerates Alu RNA processing in vitro

(A) In vitro incubation of one of the Alu RNA consensus sequences for different incubation periods. In vitro transcribed Alu RNA (67 nM) incubated at 37°C with 250 nM HSF1 in the course of 6 h with time intervals of 1 h.

(B) Comparison among Hsf1 (~ 60 kDa), denatured HSF1, poly-A polymerase (~ 56 kDa), and no protein (just TAP buffer) with regard to Alu RNA processing (estimating relative full-length RNA remaining) (two replicates). The full gels are available as Source Data for Fig 8. Relative full-

length RNA remaining was calculated using ImageJ area under the curve software over time. Error bars represent standard deviation from the mean.

(C) Comparison among Alu RNA (three replicates) and two control RNAs (two replicates) regarding the full-length RNA levels remaining after in vitro incubation for 90 min at 37°C with HSF1. Sizes of control RNAs are control for RNA #1, 143nt and for control RNA #2, 432nt. Incubation in the absence of HSF1 but presence of the same buffer (TAP) was used as control to take into account any non-HSF1-specific RNA destabilization due to non-specific degradation. Threshold for statistical significance was a P value of 0.05 with $P = 0.03$ (unpaired, non-directional t-test) for the comparison between HSF1 and no protein incubation ($n = 3$, for Alu RNA). The full gels are available as Source Data for Fig 8. Error bars represent standard deviation from the mean.

(D) Plotting of the position of the first base (5' end) of Alu RNA fragments across the Alu consensus sequence produced by Alu RNA that has been processed in vitro for 90 min at 37°C in the presence of HSF1 (upper panel) and compared with one of the in vivo samples of Fig 1 (lower panel). The x-axis represents an Alu RNA metagene aligned at the start site of the Alu consensus sequence, and the y-axis shows the 5' end count for Alu RNA fragments aligning to any position downstream of position +1. The in vivo sample x-axis depicts a 11nt shift compared to the in vitro one.

(E) Scatterplot depicting the positive correlation between HSF1 mRNA expression values and Alu RNA processing ratio in MAP AD patients ($r = 0.72$, $P < 0.001$). Statistical test is based on Pearson's product moment correlation coefficient and follows a t distribution using the cor.test function in R package stats.

(F) Scatterplot depicting lack of correlation between HSF1 mRNA expression values and full-length Alu RNA levels in MAP AD patients (ns = non-significant, with P value 0.05 as the significance threshold). Statistical test as in subfigure 8E.

Then, we purified the Alu RNA fragments derived in vitro from the incubation with HSF1 and subjected them to the same short RNA sequencing protocol we used for the post-mortem samples in order to compare the Alu RNA fragmentation pattern in vitro with that in vivo. As shown in Fig 8D, we have been able to identify also in vitro the major cut at Alu RNA right arm we had observed in vivo. Interestingly, the position of the cut was slightly shifted for 11nt toward the 5' end, across the poly-A bridge. This finding implies that the initial cut in vivo may be subsequently subjected to a yet unclear form of maturation that trims the 5' end of the generated fragment. Alternatively, given the great genomic sequence diversity of Alu RNA elements among humans, we cannot exclude that the single Alu consensus sequence we used for the in vitro synthesis of the RNA may not correspond completely to the most frequently represented Alu RNAs in the samples we tested

from the patients that may include deletions compared to this consensus. However, the lack of whole-genome sequencing data from these patients makes clarifying this point challenging.

These data show that B2 RNA and Alu RNA may share a common factor, HSF1, as a potential accelerator of their processing. To assess this potential in our context, we tested the association of Alu RNA processing ratio with regard to HSF1 mRNA levels in the brains of MAP AD patients. Indeed, as shown in Fig 8E a strong correlation was found between them in AD patients ($r = 0.72$, $P < 0.001$). In contrast, when an association between HSF1 levels and full-length Alu RNA levels was tested, no correlation was found (Fig 8F).

These data suggest that similarities observed between mice and humans in vitro may also extend to SINE RNA processing in human brain.

3.3 Discussion

Elucidation of the molecular mechanisms underlying AD pathogenesis comprises an important part of efforts currently underway to understand better this debilitating disease. To this end, recent high-throughput approaches and advances in our understanding of novel elements of human transcriptome's architecture such as RNAs from SINE transposable elements may offer novel perspectives regarding molecular pathology of neurodegenerative diseases.

SINE Alu elements constitute a substantial part of the repetitive non-coding genome in humans, being present in millions of copies in human DNA (Walters et al, 2009). Although SINE elements had been regarded for long as genetic parasites and part of "junk" DNA, numerous studies in recent

years have started shedding light into their potential functions in increasing genomic regulatory capabilities through generation of new genomic regulatory sites (Deininger, 2011). However, little is known about the role in human health and disease of the non-coding RNAs generated by these SINE repeats. The identification, some years ago, of the critical ability of SINE B2 and SINE Alu RNAs to regulate transcription through suppression of RNA polymerase II, both in humans and in mouse, opened a totally new avenue of possibilities regarding their potential role in cellular function as well as their potential involvement in molecular pathologies (Espinoza et al, 2007; Mariner et al, 2008; Walters et al, 2009; Yakovchuk et al, 2009; Ponicsan et al, 2010, 2015). Subsequently, our recent discovery of the SINE RNA ribozyme/riboswitch potential in response to cellular stress expanded further these possibilities (Zovoilis et al, 2016; Hernandez et al, 2020) and resulted in the identification of a connection between transcriptome deregulation observed in amyloid beta pathology in mouse brain and increased processing of SINE B2 RNAs accelerated by high Hsf1 levels (Cheng et al, 2020).

Here, we show that SINE Alu RNAs are processed also in human brains and reveal increased processing of SINE RNAs as a novel type of transcriptome deregulation observed in AD (Fig 9). These findings allow us to extend to human AD patients our previous conclusions in murine models of AD and deduce that SINE RNAs may be important components of the molecular mechanisms underlying this disease. To this end, this is the first human disease identified to be connected with the ribozyme-based processing of SINE RNAs and also with a ribozyme-based activity in humans in general.

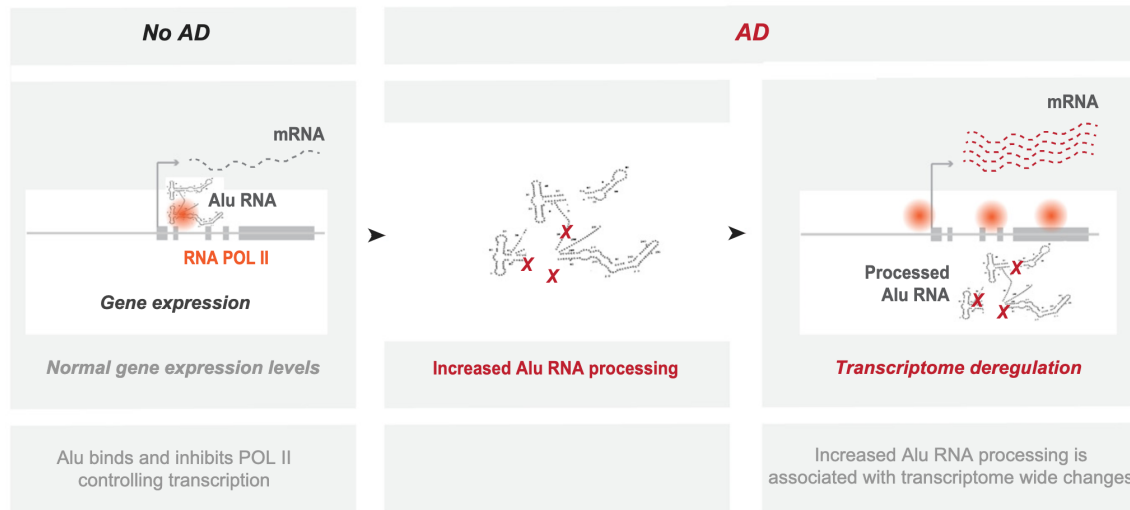


Figure 9. Representation of the observed changes in Alu RNA processing in AD

In no AD brains, basal Alu RNA processing levels maintain control of gene transcription through regulation of RNA Pol II. In contrast, in AD pathology, various, yet unknown, factors lead to an increase in Alu RNA processing associated with transcriptome-wide deregulation of gene expression.

The hypothesis that transposable elements such as Alu elements may play a role in neurodegeneration is not new. Non-coding RNAs and transposable elements in the context of neurodegenerative diseases have been shown to play interesting roles in how neural cells regulate pro-survival and pro-death pathways. For example, tau has been shown to activate transposable elements in Alzheimer's disease, while pathogenic tau-induced piRNA depletion has been shown to promote neuronal death through transposable element dysregulation in neurodegenerative tauopathies (Sun et al, 2018). Moreover, a described protein-RNA interaction between the engrailed homeoprotein b and the LINE-1 retrotransposon has been shown to prevent neurodegeneration in adult dopaminergic neurons (Blaudin de The et al, 2018). A potential connection of Alu elements with neurodegeneration has been documented in a number of reports, and it is very well outlined in a review by Larsen and colleagues (Larsen et al, 2017) that points into the impact of insertions of Alu elements into genes such as TOMM40 that control

mitochondrial function. In addition, more than 95% of Adenosine-to-Inosine RNA editing sites have been discovered within Alu regions (Huang et al, 2018) and it is now well established that the major targets for these modifications are Alu elements embedded in mRNAs, especially the retro-transposition-incompetent ones (Nakahama & Kawahara, 2020). In particular, RNA editing in protein coding genes has been shown to be connected with recoding (changes in the protein coding sequence) of protein coding RNAs associated with neurodegenerative diseases including Alzheimer's disease (Singh, 2012; Khormesh et al, 2016; Annese et al, 2018; Krestel & Meier, 2018; Larsen et al, 2018; Kanata et al, 2019).

However, until now the role of Alu elements in neurodegeneration has been described mainly as a passive one, via disruption of genomic integrity through their integration into critical regions in the genome or as mediators of RNA editing of the mRNAs in which they are embedded. Here, we present evidence about a more active role of SINEs in regulation of gene expression in neurodegenerative diseases, through an independent ribozyme-related function of the transcripts generated by these elements. Moreover, a pioneering study by the Ambati laboratory (Kaneko et al, 2011) revealed a critical role of Alu RNAs in pathogenesis of macular degeneration, through cellular toxicity effects of Alu RNAs. Thus, in addition to the potential direct impact that Alu RNA processing can have on gene transcription through abolition of Pol II suppression, it is likely that increased Alu RNA processing we observe in AD patients may impact cells also in other ways such as through the generation of toxic Alu RNA species.

In this study, we employed RNA-seq data from two different sequencing approaches, one customized for these RNAs (short RNA-seq) and a standard one used in one of the larger

transcriptome studies in AD, in order to test whether existing standard RNA-seq datasets can be used to study Alu RNA processing beyond the more specific short RNA-seq approach. Our findings indeed indicate that these data can be used in case of humans for the study of Alu RNA processing. Being able to use standard RNA-seq to study Alu RNA processing has significant advantages with regard to the more specific short RNA-seq approach given that existing publicly available data have been generated using standard protocols. Moreover, the list of Alu RNA sequences that were found to be differentially processed in hippocampus of CBB patients was different from that in the prefrontal cortex of MAP patients. This difference could be attributed to either the different tissue source (different transcription levels in different tissues) or to different demographics between the two patient cohorts (genetic variability/CNVs, SNPs, insertions/deletions), or to a combination of both. Thus, despite the end result being the same (i.e., increased processing of Alu RNAs), the source of these RNAs remains a population of a quite diverse set of Alu genomic elements reflecting human genomic variability, which denotes the need for cost-effective ways to study Alu RNA processing in order to include large patient cohorts that are representative of such variability.

For our analysis, we have used RNA-seq data from bulk tissue specimens. Despite their undisputable value, such type of data cannot directly distinguish between neuroglia and neurons. As shown in Appendix Fig S8, RNA-seq data used in this study express both markers that are predominantly neuronal (such as SLC17A7) and markers that are mainly non-neuronal (such as AQP4) in contrast to RNA-seq dataset from exclusively non-neuronal cells (Appendix Fig S8A–C). Application of a cellular deconvolution approach revealed that approx. 43% of the cells that contribute to the MAP RNA-seq data are neuronal cells (Appendix Fig S8D and E). Moreover,

both neuronal and neuroglia cells express Alu RNAs (Appendix Fig S8F). These data exclude the possibility that our findings could be attributed exclusively to either neuronal or non-neuronal cells within this brain tissue, though the exact contribution of each cell type to our results remains unknown. The same applies to the observed correlation between P53 and Alu RNA processing ratio. It is currently unclear whether this is result of microgliosis, astrogliosis neuronal death, or combination of these, since our RNA-seq is not single cell specific and all the identified pathways (Dataset EV8, Appendix Figs S6 and S7) are cellular stress response and signaling pathways that are universally expressed in all these cell types.

In our analysis, we found no correlation between HSF1 levels and full-length Alu RNAs. We have included the HSF1 vs. full-length plot in Fig 8 as a control to the HSF1 vs. processing ratio plot. In that way, we wanted to exclude the possibility that Alu RNAs may be under the direct transcriptional control of HSF1, which would confound our findings by causing an increase in the denominator of our estimated Alu RNA processing ratio. Our results show that HSF1 is unlikely to be an upstream direct regulator of Alu RNA transcription, as there is practically no correlation with Alu RNA full-length levels, and that HSF1 rather exerts its action on Alu RNAs by increasing the proportion of fragmented Alu RNAs. As with many gene circuits and pathways involved in cellular response to stress, there are usually a number of compensatory cellular homeostasis mechanisms in place that through positive and negative feedbacks regulate RNA levels. In our case, it would be reasonable to expect that in vivo there are compensatory pathways that would respond to the reduction of Alu RNA levels by increased processing through an increase of Pol III Alu RNA transcription. This could even result in an increase of total Alu RNA levels after chronic exposure to such stimulus as observed in Fig 4D.

Due to the repetitive nature of Alu RNAs, it is currently difficult to make any direct conclusions regarding how many of the Alu sequences tested here originate from Pol III transcripts and how many from transcripts embedded into mRNAs (likely nascent ones). As shown in Appendix Fig S9, we tried to address this question indirectly by repeating our mapping against the genome, and separating the Alu elements, against which we map the RNA fragments, into two categories: (i) Alu elements that fall within gene regions, and (ii) Alu elements outside of gene regions. Despite Alu multiple mapping, a level of information regarding genomic positions that generate Alu RNAs is expected to be maintained through this approach. Therefore, if the mapped Alu RNAs originated exclusively from either only Pol III Alu elements or mRNA embedded Alu elements, we would expect at least some difference in the distribution of fragments between Alu elements of these two categories, as the genic ones overlaps with mRNAs. As shown in Appendix Fig S9, the fact that distribution models are very similar between the two categories implies that both types of Alu elements may contribute to Alu RNA processing. However, given the limitations posed by the repetitive nature of Alu RNAs mentioned above, it remains difficult to provide an exact number regarding the portion of B2 RNA fragments produced by each category.

In the current study, we have justified higher levels of Alu RNA fragments in AD patients as a result of higher processing and destabilization of Alu RNAs, using as reference the full-length Alu RNA levels estimated through the poly(A) selection-based RNA-seq in MAP patients. However, it is important to note that acceleration of Alu self-cleavage is likely only one of the ways through which the cell may control Alu RNA levels. Other pathways may also be at play including nucleases or protection through A-I editing. In fact, the 11nt shift in the cutting point we observe

in vivo compared to the in vitro processed RNA in Fig 8D implies an additional Alu RNA processing step in vivo beyond its self-cleaving activity. We are at the moment unaware of any nuclease defects in AD patients that may be specific to Alu RNA, but such a possibility can also not be excluded.

Our study leaves also some additional open questions. This study is the first to confirm an association between SINE RNA transcriptome and genome-wide transcriptome changes in a human tissue. We also showed that in the context of a neural cell line, targeting SINE Alu RNAs in humans elicits the up-regulation of various genes in agreement with what was previously described in mouse. However, it remains unclear whether this applies also in vivo since performing such studies in vivo is not possible.

Moreover, our data tested in humans the potential role of HSF1 in SINE RNA processing previously observed in mouse only in vitro. Future studies should examine that impact also in cell systems that may resemble neural cells in human brain. Finally, it currently remains unknown which other factors may be implicated in the increased processing of Alu RNAs in AD beyond HSF1. Elucidating the factors that are upstream of this process may provide additional insight regarding the mechanisms underlying molecular pathology of AD and help advance further our understanding of these RNAs as potential novel therapeutic targets and not just transcriptional noise and “junk DNA” products.

3.7 References

- Annese A, Manzari C, Lionetti C, Picardi E, Horner DS, Chiara M, Caratozzolo MF, Tullo A, Fosso B, Pesole G et al (2018) Whole transcriptome profiling of Late-Onset Alzheimer's Disease patients provides insights into the molecular changes involved in the disease. *Sci Rep* 8: 4282
- Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS (2012) Overview and findings from the religious orders study. *Curr Alzheimer Res* 9: 628–645
- Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nat Rev Neurosci* 9: 768–778
- Blaudin de The FX, Rekaik H, Peze-Heidsieck E, Massiani-Beaudoin O, Joshi RL, Fuchs J, Prochiantz A (2018) Engrailed homeoprotein blocks degeneration in adult dopaminergic neurons through LINE-1 repression. *EMBO J* 37: e97374
- Bloom GS (2014) Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* 71: 505–508
- Chang JR, Ghafouri M, Mukerjee R, Bagashev A, Chabrashvili T, Sawaya BE (2012) Role of p53 in neurodegenerative diseases. *Neurodegener Dis* 9: 68–80
- Cheng Y, Saville L, Gollen B, Isaac C, Belay A, Mehla J, Patel K, Thakor N, Mohajerani MH, Zovoilis A (2020) Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. *Elife* 9: e61265
- Cornutiu G (2015) The epidemiological scale of Alzheimer's disease. *J Clin Med Res* 7: 657–666
- Deininger P (2011) Alu elements: know the SINEs. *Genome Biol* 12: 236
- Espinoza CA, Goodrich JA, Kugel JF (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* 13: 583–596
- Funkhouser SA, Steibel JP, Bates RO, Raney NE, Schenk D, Ernst CW (2017) Evidence for transcriptome-wide RNA editing among *Sus scrofa* PRE-1 SINE elements. *BMC Genom* 18: 360
- Gates LA, Shi J, Rohira AD, Feng Q, Zhu B, Bedford MT, Sagum CA, Jung SY, Qin J, Tsai MJ et al (2017) Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J Biol Chem* 292: 14456–14472
- Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, Bullido MJ, Engelborghs S, De Deyn P, Berr C et al (2011) APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 16: 903–907
- Gonzalez C, Armijo E, Bravo-Alegria J, Becerra-Calixto A, Mays CE, Soto C (2018) Modeling amyloid beta and tau pathology in human cerebral organoids. *Mol Psychiatry* 23: 2363–2374

Green RC, Roberts JS, Cupples LA, Relkin NR, Whitehouse PJ, Brown T, Eckert SL, Butson M, Sadovnick AD, Quaid KA et al (2009) Disclosure of APOE genotype for risk of Alzheimer's disease. *N Engl J Med* 361: 245–254

Hadjiargyrou M, Delias N (2013) The intertwining of transposable elements and non-coding RNAs. *Int J Mol Sci* 14: 13307–13328

Hane FT, Robinson M, Lee BY, Bai O, Leonenko Z, Albert MS (2017) Recent progress in Alzheimer's disease research, part 3: diagnosis and treatment. *J Alzheimers Dis* 57: 645–665

Hernandez AJ, Zovoilis A, Cifuentes-Rojas C, Han L, Bujisic B, Lee JT (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc Natl Acad Sci USA* 117: 415–425

Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, Close JL, Long B, Johansen N, Penn O et al (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573: 61–68

Huang Y, Cao Y, Li J, Liu Y, Zhong W, Li X, Chen C, Hao P (2018) A survey on cellular RNA editing activity in response to *Candida albicans* infections. *BMC Genom* 19: 43

Ittner LM, Ke YD, Delerue F, Bi M, Gladbach A, van Eersel J, Wolfing H, Chieng BC, Christie MJ, Napier IA et al (2010) Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models. *Cell* 142: 387–397

Jazvinscak Jembrek M, Slade N, Hof PR, Simic G (2018) The interactions of p53 with tau and Ass as potential therapeutic targets for Alzheimer's disease. *Prog Neurobiol* 168: 104–127

Kaaij LJT, Mohn F, van der Weide RH, de Wit E, Buhler M (2019) The ChAHP complex counteracts chromatin looping at CTCF sites that emerged from SINE expansions in mouse. *Cell* 178: 1437–1451

Kanata E, Llorens F, Dafou D, Dimitriadis A, Thune K, Xanthopoulos K, Bekas N, Espinosa JC, Schmitz M, Marin-Moreno A et al (2019) RNA editing alterations define manifestation of prion diseases. *Proc Natl Acad Sci USA* 116: 19727–19735

Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA et al (2011) DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* 471: 325–330

Karijolich J, Zhao Y, Alla R, Glaunsinger B (2017) Genome-wide mapping of infection-induced SINE RNAs reveals a role in selective mRNA export. *Nucleic Acids Res* 45: 6194–6208

Khmermesh K, D'Erchia AM, Barak M, Annese A, Wachtel C, Levanon EY, Picardi E, Eisenberg E (2016) Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* 22: 290–302

- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907–915
- Korczyn AD (2012) Why have we failed to cure Alzheimer's disease? *J Alzheimers Dis* 29: 275–282
- Krestel H, Meier JC (2018) RNA editing and retrotransposons in neurology. *Front Mol Neurosci* 11: 163
- Larsen PA, Lutz MW, Hunnicutt KE, Mihovilovic M, Saunders AM, Yoder AD, Roses AD (2017) The Alu neurodegeneration hypothesis: a primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. *Alzheimers Dement* 13: 828–838
- Larsen PA, Hunnicutt KE, Larsen RJ, Yoder AD, Saunders AM (2018) Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Res* 26: 93–111
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550
- Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 29: 499–509
- Mehla J, Lacoursiere SG, Lapointe V, McNaughton BL, Sutherland RJ, McDonald RJ, Mohajerani MH (2019) Age-dependent behavioral and biochemical characterization of single APP knock-in mouse (APP(NL-G-F/NL-G-F)) model of Alzheimer's disease. *Neurobiol Aging* 75: 25–37
- Montgomery W, Ueda K, Jorgensen M, Stathis S, Cheng Y, Nakamura T (2018) Epidemiology, associated burden, and current clinical practice for the diagnosis and management of Alzheimer's disease in Japan. *Clinicoecon Outcomes Res* 10: 13–28
- Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, Taga M, Klein HU, Patrick E, Komashko V et al (2018) A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* 21: 811–819
- Mu X, Ahmad S, Hur S (2016) Endogenous retroelements and the host innate immune sensors. *Adv Immunol* 132: 47–69

Nakahama T, Kawahara Y (2020) Adenosine-to-inosine RNA editing in the immune system: friend or foe? *Cell Mol Life Sci* 77: 2931–2948

Natt D, Thorsell A (2016) Stress-induced transposon reactivation: a mediator or an estimator of allostatic load? *Environ Epigenet* 2: dvw015

Platt RN, Vandeweghe MW, Ray DA (2018) Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res* 26: 25–43

Ponicsan SL, Kugel JF, Goodrich JA (2010) Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* 20: 149–155

Ponicsan SL, Kugel JF, Goodrich JA (2015) Repression of RNA polymerase II transcription by B2 RNA depends on a specific pattern of structural regions in the RNA. *Noncoding RNA* 1: 4–16

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842

Singh M (2012) Dysregulated A to I RNA editing and non-coding RNAs in neurodegeneration. *Front Genet* 3: 326

Sun W, Samimi H, Gamez M, Zare H, Frost B (2018) Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat Neurosci* 21: 1038–1048

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S et al (2018) Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563: 72–78

Tatosyan KA, Kramerov DA (2016) Heat shock increases lifetime of a small RNA and induces its accumulation in cells. *Gene* 587: 33–41

Toledo JB, Shaw LM, Trojanowski JQ (2013) Plasma amyloid beta measurements – a desired but elusive Alzheimer’s disease biomarker. *Alzheimers Res Ther* 5: 8

Walters RD, Kugel JF, Goodrich JA (2009) Invaluable junk: the cellular impact and function of Alu and B2 RNAs. *IUBMB Life* 61: 831–837

Wang X, Park J, Susztak K, Zhang NR, Li M (2019) Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 10: 380

Yakovchuk P, Goodrich JA, Kugel JF (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci USA* 106: 5569–5574

Zhang XO, Gingeras TR, Weng Z (2019) Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res* 29: 1402–1414

Zovoilis A, Cifuentes-Rojas C, Chu HP, Hernandez AJ, Lee JT (2016) Destabilization of B2 RNA by EZH2 activates the stress response. *Cell* 167: 1788–1802

Chapter 4 - NERD-seq: A novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs

4.1 Foreword

This chapter is presented in the form of the manuscript submitted for review at the pre-print peer review service *Review Commons*. It is also published in the form of a preprint on the biorxiv server: <https://www.biorxiv.org/content/10.1101/2021.05.06.442990v1.full#ref-29> (Accessed Mar 10th, 2023). As co-first author of the manuscript alongside my colleagues Yubo Cheng, Babita Gollen and Liam Mitchell, my contributions were “testing of the next-generation-sequencing, library construction, sequencing, writing of the manuscript,” Yubo Cheng contributed “Bioinformatics analysis, statistical analysis, establishment and testing of the analysis pipelines, data visualization,” Babita Gollen contributed “testing of the next-generation-sequencing, library construction, sequencing,” and Liam Mitchell contributed “Bioinformatics analysis, statistical analysis, establishment and testing of the analysis pipelines, data visualization.” Additional contributions include Travis Haight (“testing of the next-generation-sequencing pipeline”), Matthew Stuart-Edwards (“testing of the base calling pipeline”), Majid Mohajerani (“data interpretation”) and Athanasios Zovoilis (“conception and design, establishment and testing of data generation and analysis pipelines, bioinformatics analysis, data interpretation, data visualization and writing of the manuscript, overall supervision”).

The methodology presented in this chapter is a means to enrich primarily short ncRNAs in the Oxford Nanopore Technologies’ direct RNA-seq protocol. The primary motivatory behind being able to enrich these species of RNA is to enable the study of RNA modifications for which short ncRNAs are some of the most sophisticated recipients of RNA editing. The standard approach for direct RNA-seq is unable to capture many of these RNAs because it relies on poly(A) capture on the 3’ end. However, our protocol: NERD-seq is capable of capturing many different ncRNAs.

As a proof of principle we primarily use the RNA from the hippocampus of the Wild type mice from Chapter 2. The methodology presented in Chapter 4 is important, for much of the analysis in chapter 5 relies on enriching short ncRNAs.

4.2 Abstract

The new next-generation sequencing platforms by Oxford Nanopore Technologies for direct RNA sequencing (direct RNA-seq) allow for an in-depth and comprehensive study of the epitranscriptome by enabling direct base calling of RNA modifications. Non-coding RNAs constitute the most frequently documented targets for RNA modifications. However, the current standard direct RNA-seq approach is unable to detect many of these RNAs. Here we present NERD-seq, a sequencing approach which enables the detection of multiple classes of non-coding RNAs excluded by the current standard approach. Using total RNA from a tissue with high known transcriptional and non-coding RNA activity in mouse, the brain hippocampus, we show that, in addition to detecting polyadenylated coding and non-coding transcripts as the standard approach does, NERD-seq is able to significantly expand the representation for other classes of RNAs such as snoRNAs, snRNAs, scRNAs, srpRNAs, tRNAs, rRFs and non-coding RNAs originating from LINE L1 elements. Thus, NERD-seq presents a new comprehensive direct RNA-seq approach for the study of epitranscriptomes in brain tissues and beyond.

4.3 Introduction

Since the advent of RNA sequencing (1,2), our understanding of the transcriptome and its regulation has grown substantially(3). By improving the context of transcriptome analysis, RNA

sequencing allows for an improved understanding of the effects of cellular regulation(4,5), the environment(6,7), and disease pathology(8,9), on transcription changes and regulation.

Since RNA sequencing's introduction as an essential molecular biology tool, many technological iterations have occurred. From the initial iteration of sequencing by synthesis of cDNA, through improved technologies such as 454 pyrosequencing(10), and eventually the Illumina sequencing platform(11), a number of limitations have been identified. In particular, although short read sequencing is both powerful and exceptionally high throughput on the Illumina sequencing platform, the need to fragment longer RNA polynucleotides and the need to reverse transcribe the RNA and PCR amplify the cDNA before the sequencing makes identification of RNA modifications(12), and variations in RNA structure and RNA splicing(13,14) challenging, complicated and expensive, while increasing the probability of sequencing and analysis errors(15–17).

Recent technological iterations show promise in addressing many of the above challenges associated with RNA sequencing. Third-generation sequencing is dominated by the PacBio and Oxford Nanopore Technologies (ONT) sequencing platforms and has made long read – whole molecule – polynucleotide sequencing possible(18–22). Most promising, the Nanopore platforms enable us to sequence native RNA polynucleotides in their whole form, without the need for replacement by cDNA and subsequent amplification(23–25). Although Nanopore sequencing is capable of sequencing cDNA and cDNA amplicons(25), the sequencing of the native RNA strand (direct RNA-seq) allows the resolution of modified RNA nucleotides to provide context to the epitranscriptome(19,26), while reducing the library preparation and analysis complexity(27,28).

Over 100 unique RNA modifications have been described(29), including adenosine to inosine (A-to-I) edits, pseudouridylation, and methylation on multiple sites on the nucleotide base(30–35). Non-coding RNAs (ncRNAs) constitute frequent targets of these modifications with emerging significance for human health and disease. For example, higher A-to-I editing ratios in SINE RNAs have been linked to reduced severity in some viral infections such as that by SARS-CoV-2(36), and reduced A-to-I editing ratios have been linked to multiple sclerosis(37). Additionally, m6A modifications are thought to encourage circular RNA formation by back splicing(38) and inosines in tRNAs allow for wobble base pairing for redundant codon recognition, dysfunction of which may cause intellectual disability(39). Direct RNA-seq has exhibited the ability to detect these RNA modifications in recent research studies(28,40,41) by analyzing perturbations in the ionic trace plots produced by Nanopore sequencing.

Despite the promise of Nanopore sequencing to deconvolute the above native edits, the process of direct RNA-seq using standard, polyadenylation (poly(A))-selecting Nanopore library preparations, eliminates the possibility for capturing many of the ncRNAs that constitute the vast majority of editing substrates without modifying the library preparation method(23,32,42–44). Some of the substrates that may be omitted include among others tRNAs, snoRNAs, snRNAs, scRNAs and other cellular non-poly(A) RNAs as well as viral RNAs. Here, we present an approach, called NERD-seq, to expand the ncRNA representation in Nanopore direct RNA-seq to include these classes of ncRNAs.

4.4 Materials and methods

4.4.1 Post mortem hippocampal tissues

Mice were raised and had tissue extracted as described previously(45). Left and right mouse hippocampus tissue were homogenized separately in 1.0mL Trizol reagent: 15-minute incubation and subsequent grinding using a pestle until nothing but insoluble connective tissue remained. The homogenized mix was pipetted up and down and the solution was stored at -80°C . 0.5mL of homogenized mixture was phase separate by the addition of 100 μL of chloroform (Sigma, C2432) and mixed by inversion, incubated for three minutes and centrifuged at 12,500xg for fifteen minutes at 4°C . The top (aqueous) layer was transferred to a new tube and mixed with 250 μL of isopropanol (Fisher, 67-63-0), followed by a 1-hour incubation at -20°C and centrifugation at 12,500xg for ten minutes at 4°C . The supernatant was removed and the pellet was washed and mixed with 0.5mL of 75% ethanol, followed by a centrifugation at 7 600xg for five minutes at 4°C . The supernatant was removed and the pellet was allowed to dry for 1 minute before eluting in 30 μL of nuclease free H_2O . The eluted RNA was heated at 55°C for fifteen minutes, and subsequently incubated with 1 μL of DNaseI (NEB, M0303), 10 μL of 10x DNaseI buffer (NEB, B0303), and 39 μL of nuclease free H_2O for 15 minutes at 37°C . The RNA was further cleaned using the Zymo Research RNA clean and concentrator kit –25 (R1017) and combined in equal densities (left and right RNA) The RNA was stored at -80°C .

4.4.2 Direct RNA sequencing using Nanopore

For the standard approach we used the ONT SQK_RNA002 kit and the direct RNA-seq protocol, listed at the Nanopore Community portal (<https://nanoporetech.com/community>) with 1.5ug total RNA as starting material. In both the standard and the NERD-seq protocol, ligations were

performed for 15min instead of 10min, to increase yield. For NERD-seq, the standard protocol was further modified as follows: 1.5ug total RNA was separated into two fractions (short and long RNA) using the Invitrogen MirVana kit (AM1561) with a modified protocol as described before in our short RNA-seq approach for illumina (45–48). 90uL of the short RNA elution fraction was polyadenylated using 12uL 10x polyA polymerase buffer (NEB, B0276), 12uL 10mM ATP (NEB, B0756A), 6uL polyA polymerase (NEB, M0276) and incubated at 37C for 30min. The polyadenylated short fraction and the long fraction were combined and purified and concentrated using the RNeasy MinElute kit (Qiagen, 74204). 12uL of the RNA was adaptor ligated using 4uL NEBNext Quick ligation buffer (B6058), 0.66uL RNA CS (Nanopore), 1.3uL RTA adaptor (Nanopore), 2uL 2 000 000U T4 DNA ligase (NEB, M0202). The reaction was incubated at room temperature for 10 minutes. The sample was subsequently reverse transcribed using 1.5uL of the Lucigen OmniAmp polymerase (F831942-1), 10uL 10x OmniAmp buffer (Lucigen, F883707-1), 8uL dNTP (NEB, N0447), 6uL 100mM MgSO₄ (Lucigen, F98695-1), 10uL Betaine (Lucigen, F881901-1), 5uL random primer mix (NEB, S1330), and 39.5uL RNase free H₂O. The mixture was incubated for 50C – 10min, 70C – 20min. Samples were cleaned by 2.88x Omega Mag-Bind® TotalPure NGS beads. The 20uL library was ligated to the Nanopore RMX adaptor using 6uL RMX adaptor (Nanopore), 8uL NEBNext quick ligation buffer, 3uL RNase free H₂O, and 3uL 2 000 000U T4 DNA ligase and incubated for 10 minutes at room temperature. The adapted library was 1x bead cleaned using the Omega Mag-Bind® TotalPure NGS beads, eluted in Nanopore elution buffer. Library was loaded onto Nanopore PromethION according to manufacturer's instructions (flow cells were version 9.4.1).

4.4.3 Bioinformatics analysis

The RNA sequencing data was aligned using minimap2 (86) version 2.17 against the GRCm38/mm10 Mouse Genome. Minimap2 was used with the options: -x sr to optimize for short RNA reads. Aligned sam files were converted to bam format using samtools version 1.7(49). Bam to bed conversion was performed using bedtools version v2.26.0(50). Intersections were performed using bedtools v2.26.0 and with the intersect options: -s -wa -u -e -f 0.8. Annotations were retrieved from UCSC Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>); specifically, GENCODE V23, GENCODE V25, and RepeatMasker (as of Jan 2021); with subsequent filtering for specific families of elements. The number of lines in the intersected bed files (corresponding to number of reads which occur within respective annotation file) are taken as a percentage of the total number of reads present in the pre-intersection bed file, thus giving proportion of reads which fall into the category represented by each annotation. TSS annotation is based on the Eponine annotation (51), and rRNA annotation was done using the Seqmonk default annotation. Models of read distribution around the TSS of various genomic elements were performed using the Babraham NGS analysis suite Seqmonk 1.38.2 (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>). In brief, we constructed metagene models around a hypothetical set of genomic points, such as the transcription start site and subsequently plotted the distribution of read counts from all contributing elements for each position. Then the numbers of reads at the same strand with the element around each different TSS were calculated and attributed to defined points in the model. Relative density or cumulative distributions at the metagene plots were generated using Seqmonk.

4.4.4 Data access

NERD-seq data have been deposited to GEO with access number GSE171312 (RNA sample used in the current study is named as “ctrl-6m_3”).

The external dataset from Sessegolo et al, 2019 was accessed from the European Nucleotide Archive, accession: PRJEB27590 (https://www.genoscope.cns.fr/ont_mouse_rna/datasets_RNA_LR.html). Samples Brain C1 and C2 were used for comparison.

4.5 Results

4.5.1 The modified library construction protocol used in NERD-seq bypasses the limitations posed by the standard Nanopore direct RNA-seq approach

Standard, direct RNA-seq relies on poly(A) selection by using a poly(T)-tethering adaptor that base-pairs canonically to the motor protein linked adaptor, facilitating the movement of the polynucleotide string through the protein pore (Figure 1A). Nucleotide detection is performed by analyzing perturbations of the ionic current in the sequencing membrane(25). By doing so, Nanopore sequencing provides a moderately simple method to sequence the poly(A) transcriptome, which predominantly represents post-transcriptionally modified mRNAs(23). Unfortunately, this method omits many ncRNAs that do not have a poly(A) tail.

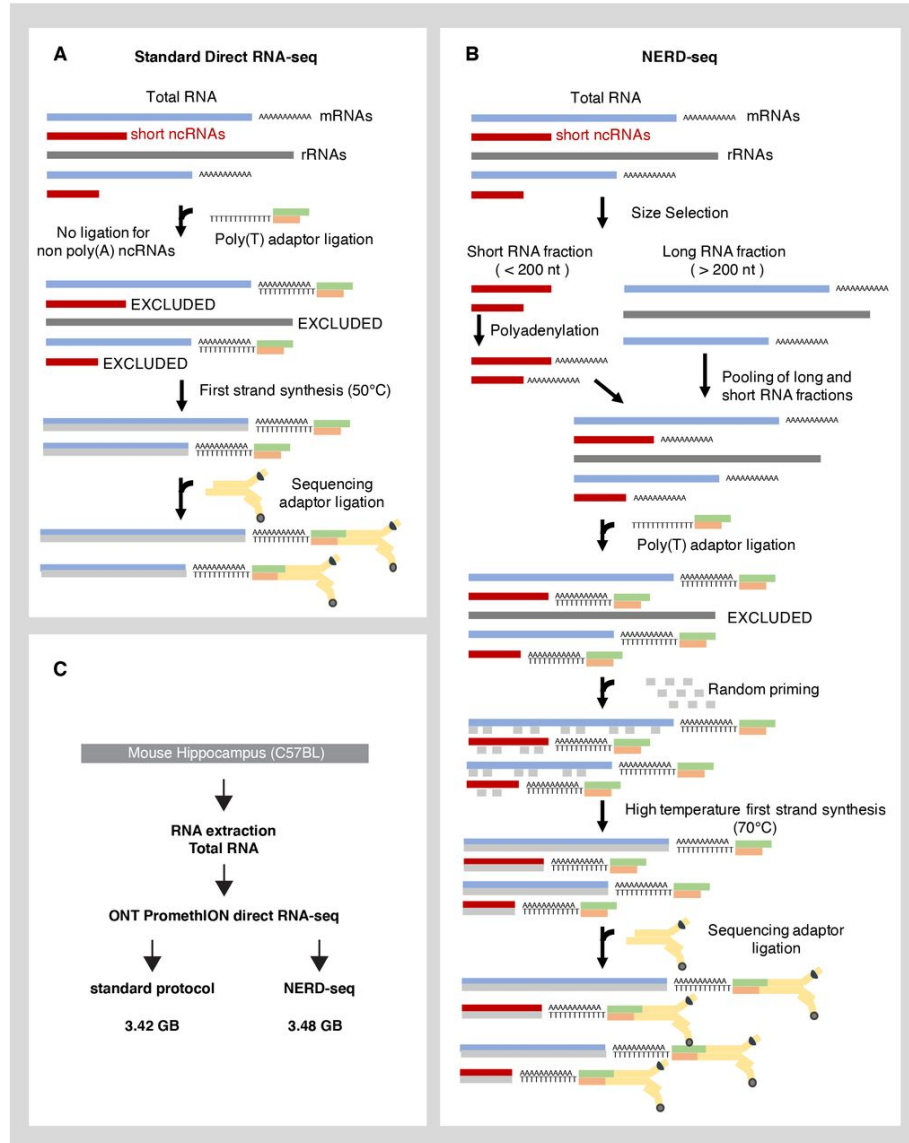


Figure 1. Description of the NERD-seq approach.

(A) Illustration of the standard Nanopore direct RNA-seq approach and the types of RNAs it can theoretically detect. ncRNA: non-coding RNA; rRNA: ribosomal RNA; mRNA: messenger RNA

(B) Illustration of the NERD-seq approach and the additional types of RNAs (short non poly(A) ncRNAs) it can theoretically detect.

(C) Experimental design of the current study.

To address this, we established a Non-coding Enriched RNA Direct sequencing approach called NERD-seq. In this modified approach (summarized in Figure 1B), we first separated total RNA into two fractions (long and short RNA fraction) using a column-based size enrichment approach,

which we have used in the past for similar approaches to enrich for short (approx. <200nt) ncRNAs with Illumina sequencing (short-RNA-seq)(45,48,52). This column-based approach enables for the significant enrichment of short transcripts (although it does not completely eliminate very long RNAs)(53). To the short fraction, we added a polyadenine tail using a poly(A) polymerase to allow in subsequent steps the short ncRNAs to form canonical base-pairing to the poly(T) sequencing adaptor and, thus, facilitate its ligation. Additionally, this poly(A) addition helps achieve higher accuracy in Nanopore sequencing of short RNAs because the beginning of the sequence usually requires some voltage adjustment for accurate basecalling and, proportionally, the error rate becomes higher for short reads(54,55). At the same time, full length rRNAs that remain at the long RNA fraction, and are usually not desirable due to their extremely high read numbers that decrease yield for the other RNA classes, are not subjected to the polyadenylation, enabling their exclusion at later protocol steps. After deactivating the poly(A) polymerase, the two fractions were pooled again together. Thus, the new sample contains all the initial RNAs with a fraction of the short ones being poly-adenylated.

An additional problem that arises in both Illumina and Nanopore sequencing protocols is the ability of highly structured RNA regions to hinder the sequencing process in various ways. In the case of Nanopore sequencing, highly structured RNA regions may prevent the pulling of the RNA molecule through the pore(25). To this end, although the direct RNA-seq protocol does not sequence cDNA, an optional first-strand cDNA synthesis step through a reverse transcriptase is recommended to stabilize the native RNA strand and resolve highly structured RNA regions commonly observed in SINE RNAs, snoRNAs and other ncRNAs (the cDNA strand is not sequenced)(56–58). However, the direct RNA-seq protocol utilizes reverse transcriptases with a

temperature optimum between 45°C-50°C. At this temperature range some highly structured RNA regions remain unresolved resulting in reverse transcriptase pauses and subsequent decrease in the length of sequenced transcripts (Figure 2A, grey distribution). To this end, we modified the complementary strand synthesis in two ways: Firstly, we reverse transcribed using the OmniAmp polymerase (Lucigen, F831942-1). Omniamp is a mutated form of the PyroPhage 3173 DNA polymerase(59,60). Omniamp is capable of loop-mediated isothermal amplification that enables reverse transcription at a temperature as high as 70°C, which facilitates further RNA unfolding. Moreover, to prevent RNA degradation at this temperature, we added random primers and initially allowed first-strand synthesis by OmniAmp for 10 minutes at 50°C to “coat” RNAs with protective cDNA strand fragments in multiple locations along their whole length. Since OmniAmp has a strong strand displacement activity(56), all these short first-strand fragments are subsequently displaced by the reverse transcribed cDNA strand that is initiated by the poly(T) adaptor.

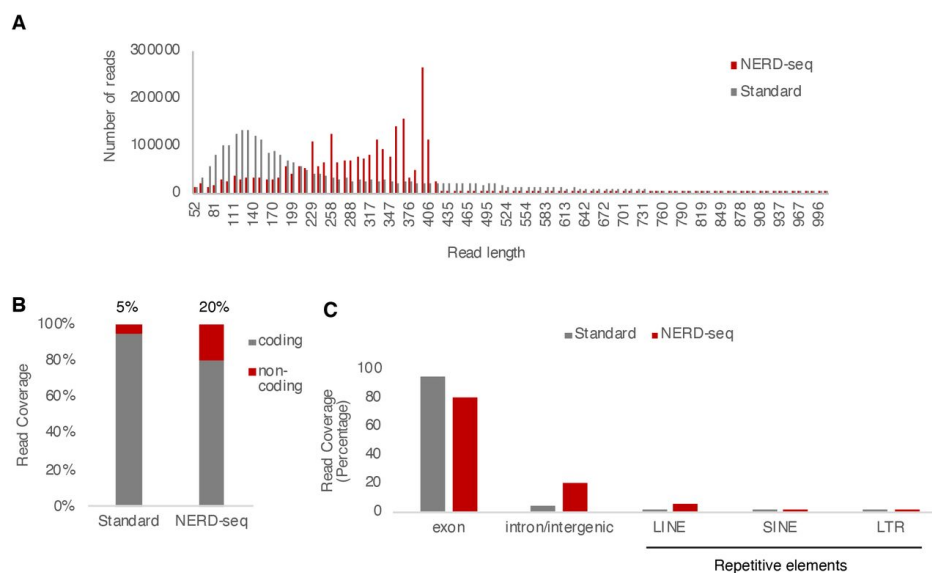


Figure 2. NERD-seq enables the generation of reads with higher coverage for the non-coding genome.

(A) Comparison of read length distribution between the standard direct RNA-seq and the NERD-seq approach.

(B) Percentage of reads aligning to the coding (exons) and non-coding portion of the genome in standard and NERD-seq approaches.

(C) Coverage (percentage) across various coding and non-coding genomic elements for standard and NERD-seq reads.

4.5.2 NERD-seq enables the generation of reads with higher coverage for the non-coding genome, while still detecting mRNAs and poly(A) ncRNAs

To assess the potential of NERD-seq we selected total RNA from mouse hippocampus, a tissue that is very active at the transcriptome and epitranscriptome level(61–63), and performed both NERD-seq and standard direct RNA-seq on the same total RNA from this tissue (Figure 1C). To our knowledge, there is only one study that has performed direct RNA-seq in neural tissues (whole brain)(64), and data from this study have been used as an external dataset for the validation of our findings (see supplement). While in that study Sessegolo et al. used the first developed platform by ONT, called MinION, we used here our in-house ONT PromethION platform, that has approx. 10-time higher yield compared to MinION. This enabled for a first time a more in-depth Nanopore sequencing in a neural mouse tissue, and the first direct RNA-seq in hippocampus. Both the standard approach and NERD-seq produced very high and comparable yields at 3.42 GigaBases (GB) and 3.48 GB, respectively. Consistent with the protocol for NERD-seq that ensures a higher representation of short RNAs, this similar yield in the overall number of sequenced bases translates to different read numbers, (3.6M for standard compared vs. 9.2M for NERD-seq). In other words, NERD-seq is able to produce more reads without this affecting its overall base yield. Moreover, NERD-seq achieved an even larger N50 number than the standard protocol. Read N50 refers to a value where half of the data is contained within reads with alignable lengths greater than this. N50 for NERD-seq was 1.84 KB compared to 1.16KB for the standard approach. Consistent with the

high-temperature first-strand synthesis approach, these results confirmed the potential of NERD-seq to increase the average length of produced reads at the range around 400nt, to which many ncRNAs belong. Indeed, as shown in Figure 2A, the center of the NERD-seq read length distribution (red distribution) is skewed towards reads of higher average length (around 400nt). These results show that the inclusion of short and long non-poly(A) RNAs does not affect the overall yield metrics but does enrich for reads with length towards 400nt.

Then, we questioned how NERD-seq reads compare to those produced by the standard approach with regards to their alignment to some of the most prevalent coding and non-coding genome elements. To this end, we calculated the percentage of the sequenced reads overlapping (>80% of their length) to coding genes (known exons in mm10) or to regions outside of them (introns, intergenic regions). As shown in Figure 2B, standard direct-RNA-seq reads predominantly (95% of the total reads) represent coding regions and particularly known exons, confirming that the standard approach is primarily tailored towards detecting and studying mRNAs and protein-coding genes. In contrast, this percentage falls to 80% in the case of NERD-seq, with the portion of reads coming from non-coding regions climbing from 5% to 20%. Finally, LINE elements are among the major non-coding elements that are overrepresented in NERD-seq compared to the standard approach (Figure 2C).

Despite the increase of the percentage of reads representing introns and intergenic regions in NERD-seq, reads originating from exons still constitute 80% of overall reads, suggesting that NERD-seq remains efficient in detecting mRNAs. To test this, we constructed metagene models of all known genes and compared the relative read density around their transcription site (i.e. read

numbers normalized to the total number of sample reads and elements that constitute the metagene). The generation of distribution plots of relative read densities allows comparisons of read coverage among different samples for the same set of genomic elements that construct the metagene model. As shown in Figure 3A, NERD-seq has been able to detect mRNAs as documented by the peak in the read distribution directly downstream of Transcription Start Site (TSS). Consistent with the higher enrichment in mRNAs in standard direct RNA-seq vs. NERD-seq mentioned above, the relative height of the peak of the distribution at TSS is higher in the standard than in NERD-seq (Kolmogorov–Smirnov test (KS test) < 0.05). However, a closer look to the characteristics of these distributions reveals that, despite being lower at the TSS, the NERD-seq distribution consists of proportionally longer transcripts that extend further downstream of TSS. This is more obvious when we scale both distributions (Supplementary Figure S1), revealing that reads from mRNAs detected by NERD-seq around TSS indeed tend in average to be longer than in the standard approach.

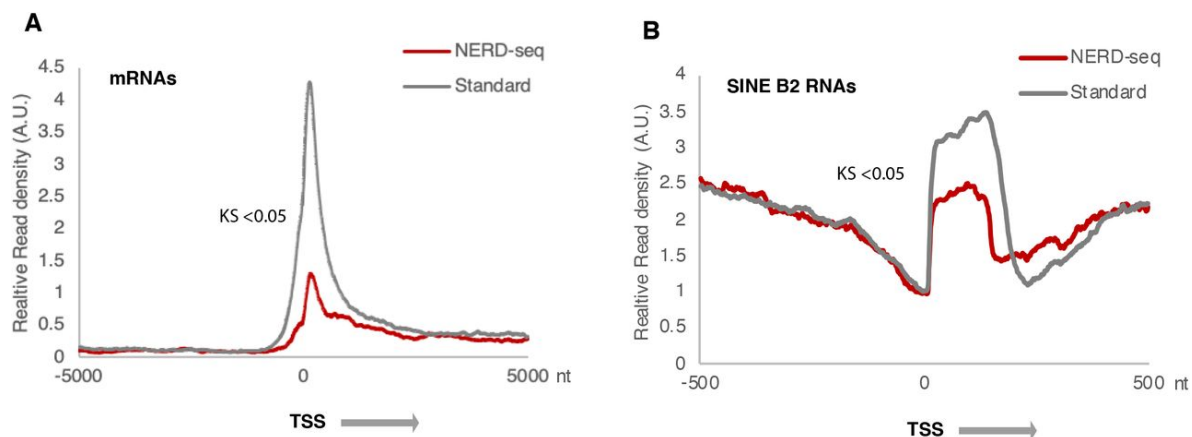


Figure 3. NERD-seq can detect mRNAs and poly(A) ncRNAs.

(A) Relative read density distribution around the Transcription Start Site (TSS) of known genes for standard and NERD-seq. A metagene model has been constructed by aligning all known genes

at their Transcription Start Site (as estimated by the use of Eponine at mm10(51)). Distances at X axis correspond to absolute distance (in nucleotides; nt) 500nt upstream (left) and downstream (right) from TSS. The arrow next to TSS depicts the direction of transcription. Read density is calculated by dividing the number of reads aligning to each position, divided to the total number of reads and elements (genes) that construct the metagene. KS: Kolmogorov–Smirnov test (KS) < 0.05 for the comparison between the two distributions.

(B) Relative read density distribution around the Transcription Start Site (TSS) of SINE B2 RNAs (repeat masker mm10) for standard and NERD-seq.

Next, we asked whether NERD-seq can detect also ncRNAs that are known to be polyadenylated.

To test this, we checked for the expression of ncRNAs generated by one of the most frequent SINE elements, the B2 element, which has been described to be polyadenylated(65,66). As shown in Figure 3B, as in the case of mRNAs, standard direct RNA-seq reads are more enriched in B2 RNAs, however, NERD-seq can still detect full-length B2 RNAs at sufficient levels.

Overall, our findings show that NERD-seq can expand coverage of non-coding genome while still being able to detect poly(A) transcripts detected by the standard protocol.

4.5.3 NERD-seq can efficiently detect various classes of short non-poly(A) ncRNAs in contrast to the standard approach

We then questioned whether NERD-seq is able to detect known classes of non-poly(A) ncRNAs, which has been also one of our major motives for developing this method. To this end, we have first generated the relative read density distribution plots around the TSS of four classes of ncRNAs: snoRNAs, snRNAs, scRNAs and srpRNAs. As mentioned above, for validating our results, in order to exclude any lab-specific technical systematic errors, in addition to the data generated by us through the standard approach for the same RNA pool, we have also employed external direct RNA-seq data generated from the same organ (brain) and the same standard

approach(64) for the comparison with our NERD-seq data (presented in Supplementary Figures S2 and S3).

As shown in Figure 4, for three out of four of these classes, snoRNAs, snRNAs and srpRNAs, the standard approach can hardly detect any of them, while it also significantly underperforms in the case of scRNAs ($KS < 0.05$) compared to NERD-seq. In contrast, NERD-seq produces robust distributions for all four classes (see also Supplementary Figure S2 for a comparison with external data). Then, we asked whether NERD-seq can detect another important class of non-poly(A) ncRNAs, tRNAs, that have been shown to be very frequent target of RNA modifications. As shown in Figure 5A, NERD-seq produces a robust distribution around the TSS of these RNAs (see also Supplementary Figure S3 for a comparison with external data). Interestingly though, as in the case of scRNAs, the standard approach can still detect some of them, though almost three-fold less than NERD-seq. Although tRNAs are known to be non-polyadenylated, it appears that they can still be detected by a poly(A)-selecting approach, such as the standard direct RNA-seq, presumably when they are marked with poly(A)s during degradation(67). As discussed below, this finding denotes the importance of an RNA-seq approach that can detect the non-poly(A) RNAs, as those may represent an entirely different biological context compared to the ones marked for degradation.

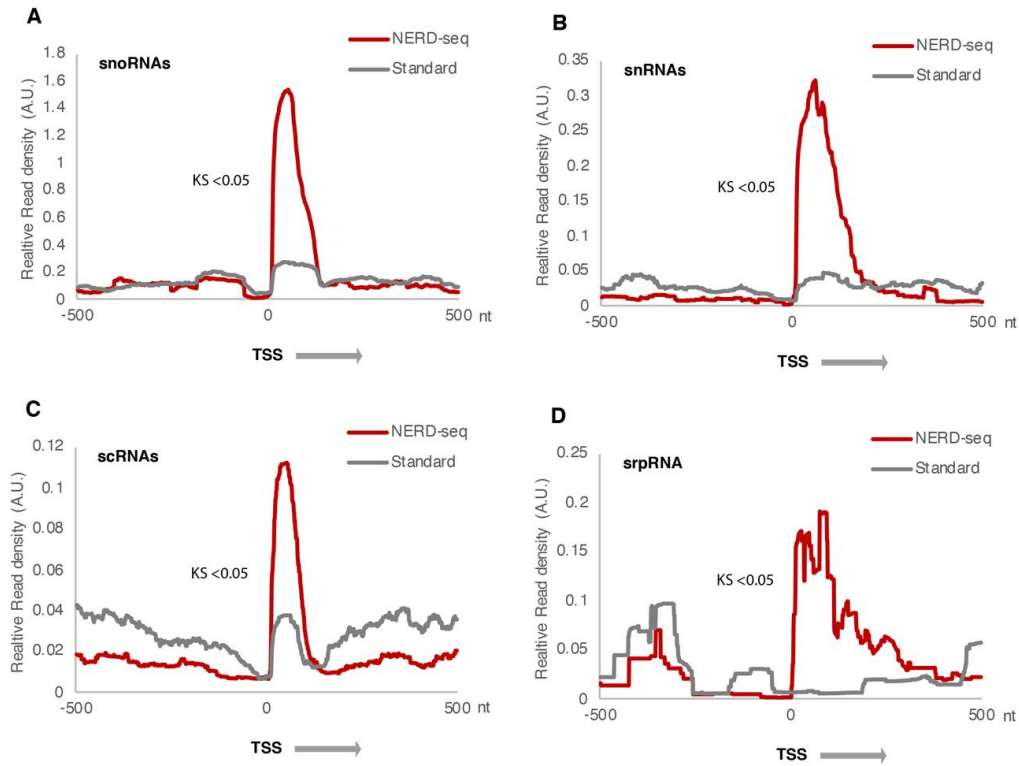


Figure 4. NERD-seq can detect snoRNAs, snRNAs, scRNAs and srpRNAs.

(A) Relative read density distribution around the Transcription Start Site (TSS) of snoRNAs for standard and NERD-seq. X, Y axis and KS-test as in Figure 3.

(B) Relative read density distribution around the Transcription Start Site (TSS) of snRNAs.

(C) Relative read density distribution around the Transcription Start Site (TSS) of scRNAs.

(D) Relative read density distribution around the Transcription Start Site (TSS) of srpRNA.

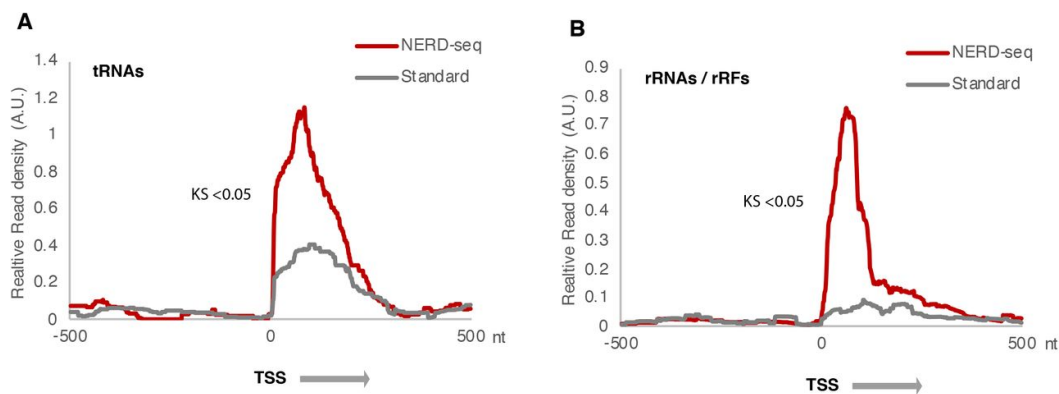


Figure 5. NERD-seq can detect tRNAs and ribosomal RNA fragments (rRFs).

(A) Relative read density distribution around the Transcription Start Site (TSS) of known tRNAs (mm10) for standard and NERD-seq. X, Y axis and KS-test as in Figure 3.

(B) Relative read density distribution around the Transcription Start Site (TSS) of rRNAs (mm10).

Finally, we examined the level of rRNA contamination in our data. As in the case of tRNAs, rRNAs are also marked with poly(A) tails for degradation, so even the standard approach is expected to detect some long rRNAs as shown in Fig.5B, which also applies to NERD-seq. Interestingly, in addition to the background rRNA detected in the standard approach, NERD-seq distribution around the rRNA TSS depicts a strong peak approx. 120-140nt wide, suggesting the existence of short rRNA fragments generated from this position that are missed by the standard approach (see Supplementary Figure S3 for a comparison with external data). Such rRNA fragments called rRFs, have been described as being capable of modulating rRNA transcription and function(68). As well, these fragments are important to cell survival and proliferation(69) in a sex, tissue, and population specific manner(70).

These results show that NERD-seq can efficiently expand direct RNA-seq capabilities to detect multiple classes of short non-poly(A) ncRNAs that may be missed by standard RNA-seq.

4.5.4 NERD-seq allows for the detection of LINE 1 produced ncRNAs

As shown in Figure 2C, reads mapping to LINE elements are overrepresented in NERD-seq. ncRNAs from LINEs have been described in the past to be important for preventing neurodegeneration with its interaction with homeoprotein b in dopaminergic neurons(71). Also, they function to aid SINE RNAs in retrotransposition(72,73). LINEs span genomic regions around 5KB or more. We wanted to identify hotspots across the LINE elements producing the RNAs detected by NERD-seq. To this end, we mapped reads generated by both the NERD-seq and standard approach across the LINE metagene, which was constructed by aligning all known LINE elements in mm10 at their start site (Figures 6A and B, respectively). As shown in these figures,

consistent with our findings in Figure 2C, LINE elements in NERD-seq data present multiple locations mapping ncRNAs compared to the standard approach. We then attempted to pinpoint the exact LINE elements producing these ncRNAs. To this end, we compared the read coverage of NERD-seq reads across all the major LINE elements listed in the UCSC mm10 repeat masker annotation track and presented at Fig.6C the three top-ranking ones based on read coverage numbers. As shown in this figure, NERD-seq read coverage of LINE elements comes predominantly from the L1 family, which far outweighs the following two families. This is confirmed by plotting NERD-seq reads at the respective L1 metagene (Figure 6D) and comparing this plot with that of other LINE families, such as L2 (Figure 6E). These results reveal the ability of NERD-seq to detect ncRNAs derived from LINE elements.

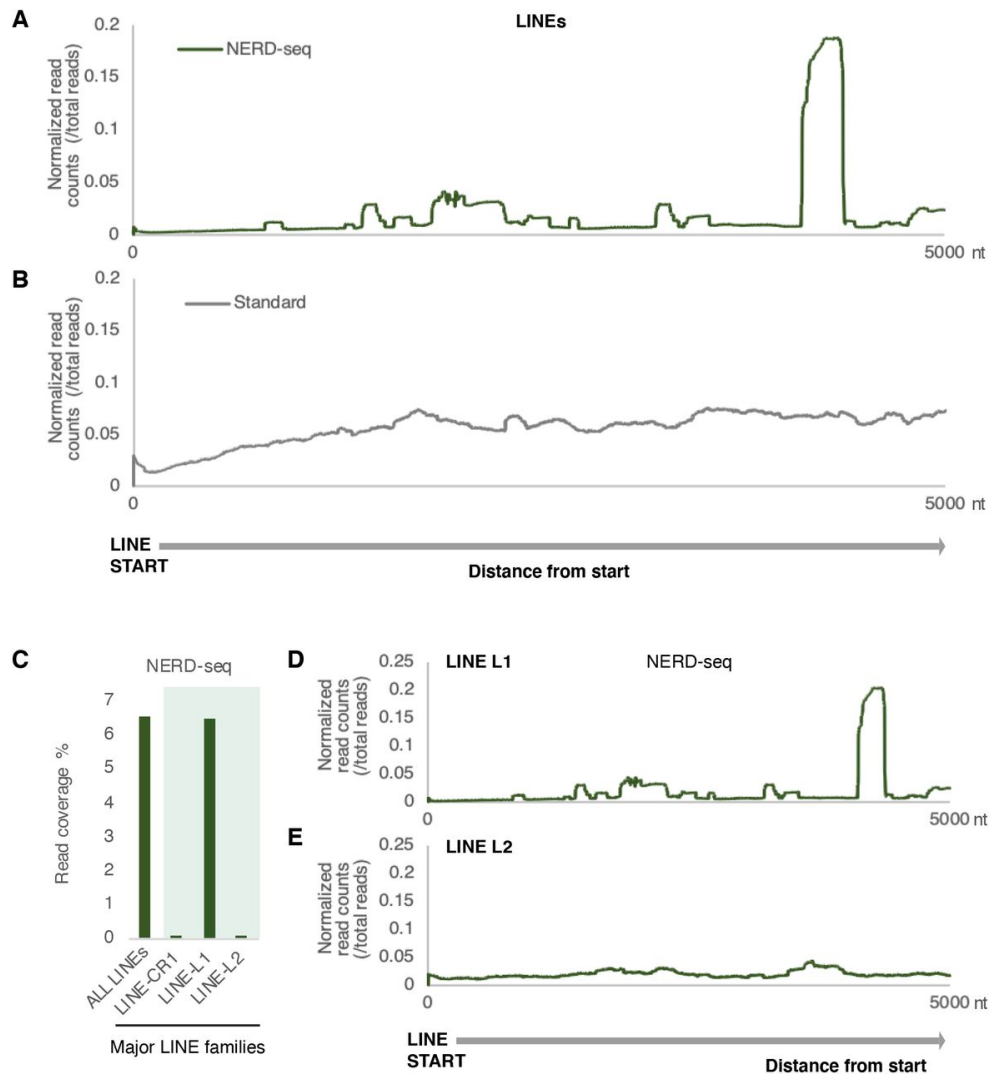


Figure 6. NERD-seq detects LINE L1 associated ncRNAs.

(A) Normalized counts of reads from NERD-seq mapped across the first 5000nt of a metagene constructed by all known LINE elements (repeat masker UCSC mm10). The metagene model has been constructed by aligning all known LINEs at their start site (base #1 in each element's DNA sequence). Distances at X axis correspond to absolute distance (in nucleotides; nt) 5000nt downstream from the start. The arrow below the graph corresponds to the sense direction of the elements. Normalized read counts per position are calculated by dividing the number of reads aligning to each position to the total number of reads.

(B) Same as in (A) but for standard direct RNA-seq reads.

(C) Coverage (percentage) across all LINEs and across the three LINE subfamilies for NERD-seq reads. The three LINE subfamilies with most aligned reads have been selected and are depicted here.

(D) Normalized counts of reads from NERD-seq mapped across the first 5000nt of a metagene constructed as in (A) but by all known LINE L1 family elements.

(E) Normalized counts of reads from NERD-seq mapped across the first 5000nt of a metagene constructed as in (A) but by all known LINE L2 family elements.

4.6 Discussion

Expanding our understanding of the epitranscriptome and its constituting RNA modifications, and why and where RNA edits occur, is expected to have profound influence on our understanding of how cells use RNA in different signalling and functional contexts(74,75), and how cells recognize internally transcribed RNAs compared to dysfunctional RNAs and RNA-related threats. The Nanopore platform and direct RNA-seq have opened new avenues for such studies. However, the standard direct RNA-seq approach prevents us from leveraging the full potential of this technology by excluding the detection of critical RNA classes that constitute some of the main targets of such modifications. Here, we present an approach that addresses these issues. NERD-seq can not only successfully detect mRNAs and poly(A) ncRNAs detected by the standard approach, but expands the detection of ncRNAs to all major classes of non-poly(A) ncRNAs. At the expense of reducing coverage of protein-coding RNAs for only 15% (decrease of 95% of overall reads to 80%), NERD-seq allows the successful detection of snoRNAs, snRNAs, scRNAs, srpRNAs and tRNAs. With the evolution of the Nanopore technology and the significant increase in the sequencing yield through ONT platforms such as PromethION, this small decrease in coverage of mRNAs can easily be tolerated. Moreover, while not amplifying full-length rRNAs, which is often undesirable, it can still detect rRFs that has recently proved to be critical in various biological contexts such as cell proliferation and survival(69) and rRNA transcription modulation(68).

For our comparison between NERD-seq and the standard protocol we chose hippocampal tissue due to its rich epitranscriptome variability throughout neurodevelopment and upon conditional treatment(61–63). Hippocampi exhibit high transcriptional activity and genome-environment interactions, while also being responsive to environmental and physiological changes(76–79).

Additionally, it is a tissue that has received considerable attention for its role in memory formation and learning, and its early dysfunction in disorders such as Alzheimer's and Parkinson's disease and during cancer treatment(45,52,80–85).

Application of NERD-seq revealed similar metrics regarding base yield and a slightly improved N50, while generating almost three times the number of reads generated by the standard approach. This is a critical point, given that a significant increase in the number of RNA fragments to be sequenced could have reduced pore viability, motor protein fuel reserves and, thus, overall yield. An important observation that has arisen from our optimization experiments for this approach is the importance of adding random primers during the first-strand synthesis. Performing the first-strand synthesis at a higher temperature has undisputable advantages in resolving RNA structures that could impede the process. However, at this temperature RNA degradation is highly accelerated, which could result in a huge increase in the proportion of very short fragments (and overall of RNA fragments to be sequenced), which, as mentioned above, could be overwhelming for fuel reserves and detrimental for yield. Addition of random primers and a preliminary first-strand synthesis step at a lower temperature (50°C) before proceeding to higher temperature seems to prevent this and protect RNA from such degradation.

Our approach has enabled for the first time in-depth Nanopore PromethION sequencing in a neural mouse tissue, and the first direct RNA-seq in hippocampus ever with both the standard and our new NERD-seq approach. All classes that are detected by NERD-seq have been shown to bear critical roles in neural cells, and, thus, by expanding capabilities for studying their epitranscriptomes, NERD-seq will help further understand the mechanisms underlying this type of

regulation. For example, snoRNAs have been shown to be key players in the RNA modification machinery but, until now, the standard approach does not allow to test for the impact of such changes on snoRNAs themselves in terms of potential self-regulatory loops. Moreover, marking of ncRNAs with poly(A) for degradation makes them the only ncRNAs in their class to be detected by the standard sequencing methods, adding a significant confounding factor, as any detected RNA modifications detected may be only connected with their degradation process and not with their other functions. NERD-seq can now identify those otherwise undetectable ncRNAs and increase the overall portfolio of RNAs

Nevertheless, this study leaves unanswered some important questions. For example, NERD-seq is able to detect more ncRNAs derived by LINEs compared to the standard approach, a finding that we have narrowed down to L1 elements, however, this finding is complicated by the large length of L1 elements. In particular, the identity of these L1-associated RNAs and whether they are related with previous ncRNAs described in LINEs, such as clusters of piRNAs, is unclear. Our attempt to associate the location of these ncRNAs with known locations of piRNAs was inconclusive, partially due to the large number of piRNA locations and their wide genomic distribution. Unfortunately, Nanopore sequencing does not allow for accurate basecalling of small RNAs such as piRNAs, which means that we cannot use the current datasets to search for a correlation between piRNA locations and that of our identified L1-associated ncRNAs. Application of Illumina sequencing has been neither conclusive due to the repetitive nature of these elements. Elucidating the identity of these ncRNAs lies beyond the scope of the current study, and future studies should elucidate the nature of these RNAs.

Overall, based on our findings, NERD-seq presents a simple but powerful approach for transcriptome and epitranscriptome analysis that expands our ability to exploit in full the potential of the Nanopore sequencing technology.

4.7 References

1. Emrich, S.J., Barbazuk, W.B., Li, L. and Schnable, P.S. (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res*, 17, 69–73.
2. Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133, 523–536.
3. Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10, 57–63.
4. Wagh, K., Ishikawa, M., Garcia, D.A., Stavreva, D.A., Upadhyaya, A. and Hager, G.L. (2021) Mechanical Regulation of Transcription: Recent Advances. *Trends Cell Biol*.
5. Nam, K.N., Mounier, A., Wolfe, C.M., Fitz, N.F., Carter, A.Y., Castranio, E.L., Kamboh, H.I., Reeves, V.L., Wang, J., Han, X. et al. (2017) Effect of high fat diet on phenotype, brain transcriptome and lipidome in Alzheimer's model mice. *Sci Rep*, 7, 4307.
6. Dal Santo, S., Zenoni, S., Sandri, M., De Lorenzis, G., Magris, G., De Paoli, E., Di Gaspero, G., Del Fabbro, C., Morgante, M., Brancadoro, L. et al. (2018) Grapevine field experiments reveal the contribution of genotype, the influence of environment and the effect of their interaction (GxE) on the berry transcriptome. *Plant J*, 93, 1143–1159.
7. Wagner, P.J., Park, H.R., Wang, Z., Kirchner, R., Wei, Y., Su, L., Stanfield, K., Guilarte, T.R., Wright, R.O., Christiani, D.C. et al. (2017) In Vitro Effects of Lead on Gene Expression in Neural Stem Cells and Associations between Up-regulated Genes and Cognitive Scores in Children. *Environ Health Perspect*, 125, 721–729.
8. Cieslik, M. and Chinnaiyan, A.M. (2018) Cancer transcriptome profiling at the juncture of clinical translation. *Nat Rev Genet*, 19, 93–109.
9. Twine, N.A., Janitz, K., Wilkins, M.R. and Janitz, M. (2011) Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One*, 6, e16266.
10. Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I. and Marden, J.H. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol*, 17, 1636–1647.

11. Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D.H., Headland, L.R., Peng, J., Maloof, J.N. and Sinha, N.R. (2012) A High-Throughput Method for Illumina RNA-Seq Library Preparation. *Front Plant Sci*, 3, 202.
12. Syddall, C.M., Reynard, L.N., Young, D.A. and Loughlin, J. (2013) The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet*, 9, e1003557.
13. Tilgner, H., Jahanbani, F., Gupta, I., Collier, P., Wei, E., Rasmussen, M. and Snyder, M. (2018) Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. *Genome Res*, 28, 231–242.
14. Bolisetty, M.T., Rajadinakaran, G. and Graveley, B.R. (2015) Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol*, 16, 204.
15. Kennedy, K., Hall, M.W., Lynch, M.D., Moreno-Hagelsieb, G. and Neufeld, J.D. (2014) Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol*, 80, 5717–5722.
16. Lahens, N.F., Kavakli, I.H., Zhang, R., Hayer, K., Black, M.B., Dueck, H., Pizarro, A., Kim, J., Irizarry, R., Thomas, R.S. et al. (2014) IVT-seq reveals extreme bias in RNA sequencing. *Genome Biol*, 15, R86.
17. Ozsolak, F. and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 12, 87–98.
18. Karst, S.M., Ziets, R.M., Kirkegaard, R.H., Sorensen, E.A., McDonald, D., Zhu, Q., Knight, R. and Albertsen, M. (2021) High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods*, 18, 165–169.
19. Zhao, L., Zhang, H., Kohnen, M.V., Prasad, K., Gu, L. and Reddy, A.S.N. (2019) Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Front Genet*, 10, 253.
20. Chen, S.Y., Deng, F., Jia, X., Li, C. and Lai, S.J. (2017) A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing. *Sci Rep*, 7, 7648.
21. Gradel, C., Terrazos Miani, M.A., Baumann, C., Barbani, M.T., Neuenschwander, S., Leib, S.L., Suter-Riniker, F. and Ramette, A. (2020) Whole-Genome Sequencing of Human Enteroviruses from Clinical Samples by Nanopore Direct RNA Sequencing. *Viruses*, 12.
22. Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.J., Buck, D. and Au, K.F. (2017) Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*, 6, 100.
23. Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J. et al. (2019) Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*, 16, 1297–1305.

24. Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M.D. and Hussain, S. (2019) A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun*, 10, 3359.
25. Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A. et al. (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*, 15, 201–206.
26. Price, A.M., Hayer, K.E., McIntyre, A.B.R., Gokhale, N.S., Abebe, J.S., Della Fera, A.N., Mason, C.E., Horner, S.M., Wilson, A.C., Depledge, D.P. et al. (2020) Direct RNA sequencing reveals m(6)A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun*, 11, 6016.
27. Wang, L., Si, Y., Dedow, L.K., Shao, Y., Liu, P. and Brutnell, T.P. (2011) A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS One*, 6, e26426.
28. Smith, A.M., Jain, M., Mulroney, L., Garalde, D.R. and Akeson, M. (2019) Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One*, 14, e0216709.
29. Ontiveros, R.J., Stoute, J. and Liu, K.F. (2019) The chemical diversity of RNA modifications. *Biochem J*, 476, 1227–1245.
30. Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem*, 79, 321–349.
31. Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*, 71, 817–846.
32. Roth, S.H., Levanon, E.Y. and Eisenberg, E. (2019) Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat Methods*, 16, 1131–1138.
33. Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H. and Bujnicki, J.M. (2006) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res*, 34, D145–149.
34. Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res*, 39, D195–201.
35. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S. et al. (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, 159, 148–162.

36. Crooke, P.S., 3rd., Tossberg, J.T., Porter, K.P. and Aune, T.M. (2021) Cutting Edge: Reduced Adenosine-to-Inosine Editing of Endogenous Alu RNAs in Severe COVID-19 Disease. *J Immunol*, 206, 1691–1696.
37. Tossberg, J.T., Heinrich, R.M., Farley, V.M., Crooke, P.S., 3rd. and Aune, T.M. (2020) Adenosine-to-Inosine RNA Editing of Alu Double-Stranded (ds)RNAs Is Markedly Decreased in Multiple Sclerosis and Unedited Alu dsRNAs Are Potent Activators of Proinflammatory Transcriptional Responses. *J Immunol*, 205, 2606–2617.
38. Di Timoteo, G., Dattilo, D., Centron-Broco, A., Colantoni, A., Guarnacci, M., Rossi, F., Incarnato, D., Oliviero, S., Fatica, A., Morlando, M. et al. (2020) Modulation of circRNA Metabolism by m(6)A Modification. *Cell Rep*, 31, 107641.
39. Ramos, J., Proven, M., Halvardson, J., Hagelskamp, F., Kuchinskaya, E., Phelan, B., Bell, R., Kellner, S.M., Feuk, L., Thuresson, A.C. et al. (2020) Identification and rescue of a tRNA wobble inosine deficiency causing intellectual disability disorder. *RNA*, 26, 1654–1666.
40. Lorenz, D.A., Sathe, S., Einstein, J.M. and Yeo, G.W. (2020) Direct RNA sequencing enables m(6)A detection in endogenous transcript isoforms at base-specific resolution. *RNA*, 26, 19–28.
41. Liu, H., Begik, O., Lucas, M.C., Ramirez, J.M., Mason, C.E., Wiener, D., Schwartz, S., Mattick, J.S., Smith, M.A. and Novoa, E.M. (2019) Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat Commun*, 10, 4079.
42. Gong, J., Liu, C., Liu, W., Xiang, Y., Diao, L., Guo, A.Y. and Han, L. (2017) LNCediting: a database for functional effects of RNA editing in lncRNAs. *Nucleic Acids Res*, 45, D79–D84.
43. Athanasiadis, A., Rich, A. and Maas, S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol*, 2, e391.
44. Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matisse, T.C., Buyske, S. and Gabriel, A. (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res*, 14, 1719–1725.
45. Cheng, Y., Saville, L., Gollen, B., Isaac, C., Belay, A., Mehla, J., Patel, K., Thakor, N., Mohajerani, M.H. and Zovoilis, A. (2020) Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. *Elife*, 9.
46. Agis-Balboa, R.C., Arcos-Diaz, D., Wittnam, J., Govindarajan, N., Blom, K., Burkhardt, S., Haladyniak, U., Agbemenyah, H.Y., Zovoilis, A., Salinas-Riester, G. et al. (2011) A hippocampal insulin-growth factor 2 pathway regulates the extinction of fear memories. *EMBO J*, 30, 4071–4083.
47. Hernandez, A.J., Zovoilis, A., Cifuentes-Rojas, C., Han, L., Bujisic, B. and Lee, J.T. (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc Natl Acad Sci U S A*, 117, 415–425.

48. Zovoilis, A., Cifuentes-Rojas, C., Chu, H.P., Hernandez, A.J. and Lee, J.T. (2016) Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell*, 167, 1788–1802 e1713.
49. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
50. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.
51. Down, T.A. and Hubbard, T.J. (2002) Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res*, 12, 458–461.
52. Cheng, Y., Saville, L., Gollen, B., Veronesi, A.A., Mohajerani, M., Joseph, J.T. and Zovoilis, A. (2021) Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes. *EMBO Rep*, e52255.
53. Mraz, M., Malinova, K., Mayer, J. and Pospisilova, S. (2009) MicroRNA isolation and stability in stored RNA samples. *Biochem Biophys Res Commun*, 390, 1–4.
54. Wilson, B.D., Eisenstein, M. and Soh, H.T. (2019) High-Fidelity Nanopore Sequencing of Ultra-Short DNA Targets. *Anal Chem*, 91, 6783–6789.
55. Kono, N. and Arakawa, K. (2019) Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ*, 61, 316–326.
56. Boivin, V., Deschamps-Francoeur, G., Couture, S., Nottingham, R.M., Bouchard-Bourelle, P., Lambowitz, A.M., Scott, M.S. and Abou-Elela, S. (2018) Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA*, 24, 950–965.
57. Chander, Y., Koelbl, J., Puckett, J., Moser, M.J., Klingele, A.J., Liles, M.R., Carrias, A., Mead, D.A. and Schoenfeld, T.W. (2014) A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front Microbiol*, 5, 395.
58. Aonuma, H., Iizuka-Shiota, I., Hoshina, T., Tajima, S., Kato, F., Hori, S., Saijo, M. and Kanuka, H. (2020) Detection and discrimination of multiple strains of Zika virus by reverse transcription-loop-mediated isothermal amplification. *Trop Med Health*, 48, 87.
59. Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M. and Mead, D. (2008) Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol*, 74, 4164–4174.
60. Moser, M.J., DiFrancesco, R.A., Gowda, K., Klingele, A.J., Sugar, D.R., Stocki, S., Mead, D.A. and Schoenfeld, T.W. (2012) Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One*, 7, e38371.

61. Li, L., Zang, L., Zhang, F., Chen, J., Shen, H., Shu, L., Liang, F., Feng, C., Chen, D., Tao, H. et al. (2017) Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis. *Hum Mol Genet*, 26, 2398–2411.
62. Merkurjev, D., Hong, W.T., Iida, K., Oomoto, I., Goldie, B.J., Yamaguti, H., Ohara, T., Kawaguchi, S.Y., Hirano, T., Martin, K.C. et al. (2018) Synaptic N(6)-methyladenosine (m(6)A) epitranscriptome reveals functional partitioning of localized transcripts. *Nat Neurosci*, 21, 1004–1014.
63. Zhang, Z., Wang, M., Xie, D., Huang, Z., Zhang, L., Yang, Y., Ma, D., Li, W., Zhou, Q., Yang, Y.G. et al. (2018) METTL3-mediated N(6)-methyladenosine mRNA modification enhances long-term memory consolidation. *Cell Res*, 28, 1050–1061.
64. Sessegolo, C., Cruaud, C., Da Silva, C., Cologne, A., Dubarry, M., Derrien, T., Lacroix, V. and Aury, J.M. (2019) Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Sci Rep*, 9, 14908.
65. Borodulina, O.R. and Kramerov, D.A. (2008) Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA*, 14, 1865–1873.
66. Ustyantsev, I.G., Borodulina, O.R. and Kramerov, D.A. (2020) Identification of nucleotide sequences and some proteins involved in polyadenylation of RNA transcribed by Pol III from SINES. *RNA Biol*, 1–14.
67. Toompuu, M., Tuomela, T., Laine, P., Paulin, L., Dufour, E. and Jacobs, H.T. (2018) Polyadenylation and degradation of structurally abnormal mitochondrial tRNAs in human cells. *Nucleic Acids Res*, 46, 5209–5226.
68. Zhu, C., Yan, Q., Weng, C., Hou, X., Mao, H., Liu, D., Feng, X. and Guang, S. (2018) Erroneous ribosomal RNAs promote the generation of antisense ribosomal siRNA. *Proc Natl Acad Sci U S A*, 115, 10082–10087.
69. Chen, Z., Sun, Y., Yang, X., Wu, Z., Guo, K., Niu, X., Wang, Q., Ruan, J., Bu, W. and Gao, S. (2017) Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS One*, 12, e0176458.
70. Cherlin, T., Magee, R., Jing, Y., Pliatsika, V., Loher, P. and Rigoutsos, I. (2020) Ribosomal RNA fragmentation into short RNAs (rRFs) is modulated in a sex- and population of origin-specific manner. *BMC Biol*, 18, 38.
71. Blaudin de The, F.X., Rekaik, H., Peze-Heidsieck, E., Massiani-Beaudoin, O., Joshi, R.L., Fuchs, J. and Prochiantz, A. (2018) Engrailed homeoprotein blocks degeneration in adult dopaminergic neurons through LINE-1 repression. *EMBO J*, 37.
72. Elbarbary, R.A., Lucas, B.A. and Maquat, L.E. (2016) Retrotransposons as regulators of gene expression. *Science*, 351, aac7247.

73. Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L. and Moran, J.V. (2015) The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol Spectr*, 3, MDNA3-0061-2014.
74. Galeano, F., Tomaselli, S., Locatelli, F. and Gallo, A. (2012) A-to-I RNA editing: the “ADAR” side of human cancer. *Semin Cell Dev Biol*, 23, 244–250.
75. Kishi, H. (1986) [Recent trend in pathogenic bacteria of urinary tract infections--simple and complicated urinary tract infections]. *Nihon Rinsho*, 44, 2552–2557.
76. Fernandez-Albert, J., Lipinski, M., Lopez-Cascales, M.T., Rowley, M.J., Martin-Gonzalez, A.M., Del Blanco, B., Corces, V.G. and Barco, A. (2019) Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat Neurosci*, 22, 1718–1730.
77. Schulz, H., Ruppert, A.K., Herms, S., Wolf, C., Mirza-Schreiber, N., Stegle, O., Czamara, D., Forstner, A.J., Sivalingam, S., Schoch, S. et al. (2017) Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun*, 8, 1511.
78. Solvsten, C.A.E., de Paoli, F., Christensen, J.H. and Nielsen, A.L. (2018) Voluntary Physical Exercise Induces Expression and Epigenetic Remodeling of VegfA in the Rat Hippocampus. *Mol Neurobiol*, 55, 567–582.
79. Peleg, S., Sananbenesi, F., Zovoilis, A., Burkhardt, S., Bahari-Javan, S., Agis-Balboa, R.C., Cota, P., Wittnam, J.L., Gogol-Doering, A., Opitz, L. et al. (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science*, 328, 753–756.
80. Lombardi, G., Crescioli, G., Cavedo, E., Lucenteforte, E., Casazza, G., Bellatorre, A.G., Lista, C., Costantino, G., Frisoni, G., Virgili, G. et al. (2020) Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer’s disease in people with mild cognitive impairment. *Cochrane Database Syst Rev*, 3, CD009628.
81. Hoozemans, J.J., van Haastert, E.S., Nijholt, D.A., Rozemuller, A.J., Eikelenboom, P. and Scheper, W. (2009) The unfolded protein response is activated in pretangle neurons in Alzheimer’s disease hippocampus. *Am J Pathol*, 174, 1241–1251.
82. Foo, H., Mak, E., Chander, R.J., Ng, A., Au, W.L., Sitoh, Y.Y., Tan, L.C. and Kandiah, N. (2017) Associations of hippocampal subfields in the progression of cognitive decline related to Parkinson’s disease. *Neuroimage Clin*, 14, 37–42.
83. Yang, M. and Moon, C. (2015) Effects of cancer therapy on hippocampus-related function. *Neural Regen Res*, 10, 1572–1573.
84. Pantazi, A. and Zovoilis, A. (2013) Vector-free methods for manipulating miRNA activity in vitro and in vivo. *Methods Mol Biol*, 936, 231–245.

85. Zovoilis, A., Agbemenyah, H.Y., Agis-Balboa, R.C., Stilling, R.M., Edbauer, D., Rao, P., Farinelli, L., Delalle, I., Schmitt, A., Falkai, P. et al. (2011) microRNA-34c is a novel target to treat dementias. *EMBO J*, 30, 4299–4308.
86. Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.

Chapter 5 - ncRNA modifications: an unmined source of information in amyloid beta neuropathology.

5.1 Foreword

This chapter constitutes yet unpublished data. A small portion (and the related software) of the chapter has been contributed by my colleague, Mr. Liam Mitchell (Figure 1) and it is presented to help contextualize the study. Liam also developed the XPore script I used to analyze the RNA modifications in APP + mice. As well, my colleague Dr. Babita Gollen performed the sequencing of the APP+ NERD-seq mice.

The research in this chapter investigates how RNA modifications are dysregulated in ncRNAs in Alzheimer's disease models and puts forth a use case for the NERD-Seq method I developed in the previous chapter. Largely, we make use of data from the APP NL-G-F model and the hippocampus HT22 cell line. We also develop multiple cell assays to show how B2 RNAs respond to changes in A-to-I expression.

Altogether, these data show that RNA modifications are an important source of dysregulation in Alzheimer's disease pathology.

5.2 Abstract

Alzheimer's disease is a complex pathology with multiple molecular contributors to neural dysfunction. At the forefront of the disease is the amyloid pathology whereby amyloid plaques accumulate in brain tissue and lead to molecular dysregulation. A relatively understudied phenomenon of Alzheimer's disease is the role of differing RNA modifications and how they may also contribute to cell dysfunction. In this chapter we show evidence for dysregulation of A-to-I editing in B2 RNAs and using the NERD-Seq sequencing protocol, we show that many other

ncRNAs may exhibit differing RNA modifications. This chapter presents a body of work that suggests RNA modifications might have important roles in pathological processes, including those underlying amyloid pathology in the hippocampus tissue from the APP NL-G-F mouse model.

5.3 Introduction

Alzheimer's disease (AD) is the most diagnosed form of dementia with an estimated incidence of 1 in 85 people as of 2050 (Brookmeyer *et al*, 2007; Gao *et al*, 2019). Primarily thought to be a proteinopathy produced by misfolded amyloid beta peptides and hyperphosphorylated tau (Bloom, 2014; Sebastian Monasor *et al*, 2020), the molecular mechanisms by which neurodegeneration and cell dysfunction occur remain elusive. As such, therapeutic targeting of amyloid and tau dysregulation has resulted in little success, despite immense funding and effort (Asher & Priefer, 2022; Cornutiu, 2015; Huang *et al*, 2019).

A relatively understudied component of AD is the epitranscriptome: the dynamic process of post-transcriptional RNA modifications in cells (Begik *et al*, 2022; Gassner *et al*, 2021; Mansi *et al*, 2021). RNA modifications vary in their stoichiometry and function, from the highly represented pseudouridine which is thought to promote RNA self-recognition (Anderson *et al*, 2010) to the double-stranded RNA (dsRNA) specific adenosine to inosine modification which allows, amongst other effects, wobble base pairing in codon: anticodon binding (Brachova *et al*, 2019).

Within the interplay of ncRNAs' playing a role in AD etiology, lies the role of the epitranscriptome on ncRNA regulation. For instance, dsRNA structures are immunogenic in the cytoplasm and if too much dsRNA species remain, the cell will initiate interferon pathways, including a

hyperediting state by Adar's p150 isoform (Pestal *et al*, 2015). Hyperediting develops as an effort to inhibit MDA5-MAVS recognition of dsRNA which stimulates an apoptotic and interferon stimulated genes response (Bajad *et al*, 2020). A common contributor to high dsRNA concentrations in the cytoplasm are SINE RNAs. SINE RNA accumulation has been shown to contribute to pathologies such as systemic lupus erythmatosus and age related macular degeneration (Kim *et al*, 2014; Wang *et al*, 2021). Interestingly, mounting evidence suggests AD patients may also feature reduced A-to-I editing than no-AD patients in the hippocampus vasculature (Crooke *et al*, 2022). Coincidentally, neuroinflammation is a common hallmark of AD and has been postulated as a potential contributor of neurodegeneration (Burg *et al*, 2018; Polesskaya *et al*, 2018; Saito *et al*, 2014; Sebastian Monasor *et al.*, 2020).

Beyond SINE RNAs, AD patients feature lowered detectable rRNA expression (Ding *et al*, 2005), though the reasoning is not well explained. In the case of 5S rRNA, it may be due to increased oxidation, which in turn could affect 5S rRNA's biogenesis (Ding *et al*, 2012). Reactive oxygen species are elevated in AD patients and the exact impact of ROS on ncRNAs in AD patients is not well understood (Izumi *et al*, 2020; Llanos-Gonzalez *et al*, 2019). If 5S rRNA is indeed more highly oxidized, other similarly important RNAs may be as well.

Moreover, RNA-protein interactions are extremely common in cells, inhibiting or enhancing activities. As discussed in Chapters 2 and 3, the Hsf1-Alu/B2 interactions are substantially inhibitory of Alu and B2's inhibition of RNA polymerase II in their respective organism during stress (Cheng *et al*, 2020; Cheng *et al*, 2021). An important component of RNA-protein interactions is the RNA modifications present. An RNA protein interaction that has seen some

interest in AD is the FMRP and BC200/ BC1 interactions. The FMRP protein binds different mRNA targets to inhibit their translation (Zalfa *et al*, 2003). BC-1 RNA in mice, evolutionarily derived from Ala-tRNAs retains some RNA modification sites and guides some of FMRP's mRNA interactions (Lacoux *et al*, 2012; Zalfa *et al.*, 2003; Zhang *et al*, 2018). Interestingly, the 2'-O-methylation of G32 and C33 are inhibitory of FMRP-BC-1 synaptic localization (Lacoux *et al.*, 2012), permitting translation of downstream targets, such as APP (Zhang *et al.*, 2018). Some of FMRP's other targets include: Arc, A-CAMKII, FXR1, FXR2, MAP1B, MAP2, GlyRa1, GluRq (Zalfa *et al.*, 2003). BC-1 is orthologous to the human BC200, an Alu-like RNA which is also thought to have AD involvement being that it is highly upregulated and similarly to BC-1, it is thought to regulate stress related gene translation (Luo & Chen, 2016).

Here we present data exploring the A-to-I editing stoichiometry of B2 RNAs in APP⁺ mice and how its dysregulation may lead to dysregulation in other cellular functions. We also report that changes in A-to-I editing in Ht22 mouse hippocampal cells changes the amount of full length B2 RNAs. As well, using our NERD-seq protocol, we report RNAs that are differentially modified between APP⁺ and wild type mice.

5.4 Material and methods

5.4.1 NERD-seq of hippocampus tissue

The NERD-seq experiments were prepared as described in Chapter 4.3 using the same APP NL-G-F mouse hippocampus RNA from Chapter 2.3.

5.4.2 Nanopore sequencing of Human cerebral cortex tissue

Cerebral cortex total RNA from human were sourced through Takara Biosciences (636561). For the standard approach we used the ONT SQK_RNA002 kit and the direct RNA-seq protocol, listed at the Nanopore Community portal (<https://nanoporetech.com/community>) with 1.5µg total RNA as starting material and 10ng of RNA sequins mix B spiked in (PMID: 27502218). In both the standard and the NERD-seq protocol, ligations were performed for 15min. For NERD-seq, the standard protocol was further modified as follows: 1.5ug total RNA was PNK digested using 50U PNK (NEB, M0201) and 10µL 10X PNK buffer for 1 hour at 37°C, then directly separated into two fractions (short and long RNA) using the Invitrogen MirVana kit (AM1561) with a modified protocol as described before in our short RNA-seq approach for illumina (45–48). 90uL of the short RNA elution fraction was polyadenylated using 12µL 10x polyA polymerase buffer (NEB, B0276), 12µL 10mM ATP (NEB, B0756A), 6µL polyA polymerase (NEB, M0276) and incubated at 37°C for 30min. The polyadenylated short fraction and the long fraction were combined, purified and concentrated using the RNeasy MinElute kit (Qiagen, 74204). 12µL of the RNA was adaptor ligated using 4uL NEBNext Quick ligation buffer (B6058), 1.3uL RTA adaptor (Nanopore), 4000U T4 DNA ligase (NEB, M0202) and topped up to 20µL with nuclease free water. The reaction was incubated at room temperature for 15 minutes. The sample was subsequently reverse transcribed using 150U GspSSD2.0 DNA polymerase (Optigene, GSPSSD2-002HC), 10µL 10X buffer, 8µL dNTP (NEB, N0447), 12µL 50mM MgSO₄, 10µL Betaine, 5µL random primer mix (NEB, S1330), and 33.5µL nuclease free H₂O. The mixture was incubated for 50°C – 10min, 70°C – 20min. Samples were cleaned by 2.88x Omega Mag-Bind® TotalPure NGS beads and eluted in 23µL. The 23µL library was ligated to the Nanopore RMX adaptor using 6µL RMX adaptor (Nanopore), 8µL NEBNext quick ligation buffer, and 6000U T4 DNA ligase and incubated for 10

minutes at room temperature. The adapted library was 1x bead cleaned using the Omega Mag-Bind® TotalPure NGS beads, eluted in Nanopore elution buffer. Library was loaded onto Nanopore PromethION according to manufacturer's instructions (flow cells were version 9.4.1).

5.4.3 7SK RNA sequencing

The following sequence was ordered as a g-block from IDT: 5' - taatagactcactataGGATGTGAGGGCGATCTGGCTGCGACATCTGTCACCCCATTGATCGC CAGGGTTGATTTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTTCCTCCCTCACCG CTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCTGAAGAGGACGACCATCCCCG ATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTGCTA GAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAGC TTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTT TTACATATAATAAATAAATAAATCTTTAAAAAAAAA – 3'. The g-block was prepared for in vitro transcription as described in (Cheng *et al.*, 2020). Synthesized RNA was assessed for purity using the Agilent Bioanalyzer RNA pico assay and 1000ng was subsequently polyadenylated as described above. Afterwards, the RNA was sequenced using the standard direct RNA sequencing approach described above.

5.4.4 Cell culture

Ht22 cells were cultured as described in (Cheng *et al.*, 2020). siRNA transfections of anti-Adar and anti-Adarb1 (IDT) were performed in an equal pool of 10nM final concentration as described in (Cheng *et al.*, 2020). 8-Aza incubations occurred at a concentration of 1µM in 100% DMSO.

Amyloid incubations occurred at a final 10 μ M in 100% DMSO and amyloid peptides (custom synthesized by Genscript) were prepared as described in (Cheng *et al.*, 2020)

5.4.5 Endonuclease V assay

250ng total RNA was incubated with 5 μ L 10X NEB 4 buffer, 1 μ L endonuclease V enzyme (NEB, M0305S) in a 50 μ L reaction for 1 hour at room temperature. Untreated samples were incubated with 1 μ L of nuclease free water.

5.4.6 qPCR

quantitative PCR was performed as described in (Cheng *et al.*, 2020) with the same primers for B2 RNA. Adar primers used were Forward: 5' – GCCAAAGACAGTGGTCAACCAG – 3'; Reverse: 5' – GAACAAGGATGTTGCTGAGGAGC – 3'. Adarb1 primers used were Forward: 5' – GTATGACGCCAGACTCTCACCA – 3'

5.4.7 Bioinformatics analysis and SINE A-to-I editing method rational

The A-to-I editing methodology was performed using a custom REDIttools pipeline whereby a custom reference was constructed from unique B2 elements within the B2 families mm1a, mm1t and mm2 (125nt or longer). Successfully mapped sequences (bwa-0.7.17 aln with maximum 10 allowed mismatches (Li & Durbin, 2010)) were converted to the bam format using samtools-1.6 (view) (Li *et al.*, 2009). REDIttools v1 (denovo script) (Picardi & Pesole, 2013) was utilized to produce tables of base alignments for each file using -c 1, -q 1, -m 1, -O 0, -V 1, -n 0 parameters. REDIttools tables were combined and within each sample and for each position (1-180) of the B2 RNAs, we defined two proportion statistics: (i) expected proportion of bases, which is four proportions A, T, C, G (sum to 1) that describe how many of the unique reference B2 RNAs have

each base in that position and, (ii) observed. which is four more base proportions that describe how many of the reads have each base in that position.

5.4.8 Differential expression analysis

The differential expression analysis was performed as described in (Cheng *et al.*, 2020). Gene ontology analysis employed the use of DAVID (<https://david.ncifcrf.gov/summary.jsp>) (Huang *et al.*, 2009; Sherman *et al.*, 2022). Interferon gene analysis employed the use of the Interferome database (<http://www.interferome.org/>) (Rusinova *et al.*, 2013).

5.4.9 Gene expression analysis of 5S rRNA

5S rRNA levels were assessed using the 5S features in a rRNA annotation produced from Biomart (<http://useast.ensembl.org/biomart/martview/1dd9354f76a77a42dcd072afb5c70e65>). Mapped reads were uploaded to Seqmonk and were annotated to the 5S rRNA feature. The annotated probe report function was used to estimate per million reads normalized expression values for each 5s rRNA gene and the average of all reads was taken within each sample.

5.4.10 Differential modification analysis

All of the reads within each sample group (APP+ and wild type) were concatenated into a pooled fastq and analyzed using the Xpore package (Pratanwanich *et al.*, 2021). The read annotation was taken from Ensembl's mouse ncRNA reference (https://ftp.ensembl.org/pub/release-109/fasta/mus_musculus/ncrna/) with the corresponding GTF annotation. The final diffmod table was filtered for coverage of 50 and a pvalue < 0.05.

5.5 Results

5.5.1 B2 RNAs' A-to-I editing stoichiometry is reduced in APP NL-G-F mice

SINE RNAs are one of the most common targets of A-to-I editing by Adar and Adarb1 (Bakhtiarizadeh *et al*, 2018; Funkhouser *et al*, 2017; Schaffer & Levanon, 2021). This led us to explore if A-to-I editing of B2 RNAs is dysregulated in APP mice as they age (Fig 2A-B). By modifying the REDITools pipeline (Mansi *et al.*, 2021; Picardi & Pesole, 2013) to detect reference to read A to G mismatches within each B2 to a consensus match (Fig 1), we show that as WT and APP mice age, their A-to-I editing ratios lessen (Fig 2C-D) but not always at the same life stages as when B2 RNA's processing increases as discussed in chapter 2 (Fig 2A (right)). And, compared to WT, APP mice's A-to-I ratios trend much higher in the 3 month old mice (Fig 2C-D).

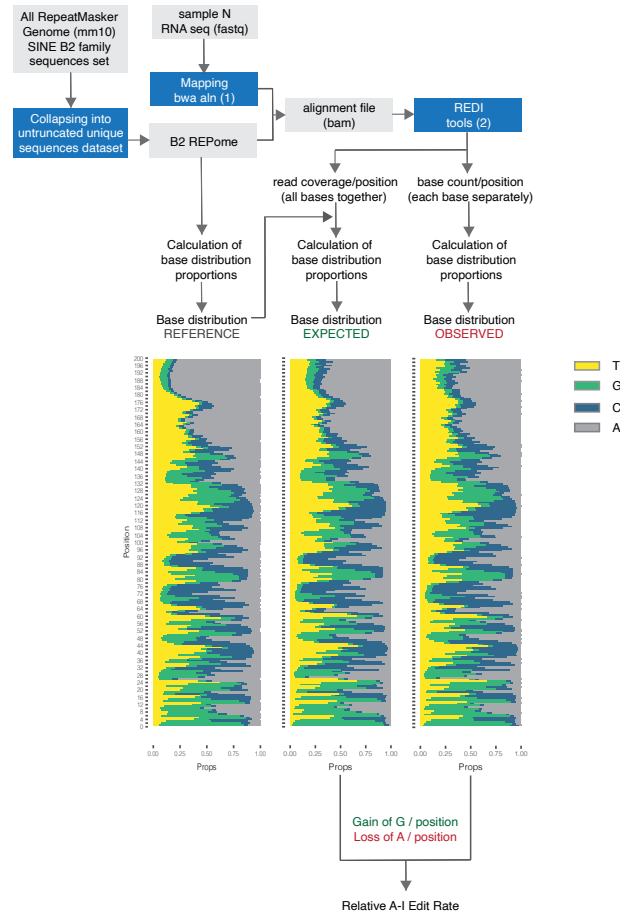


Figure 1. Workflow describing the method for estimating relative A-to-I editing rate across SINE B2 RNAs.

Arrows depict the flow of analysis. Plots depict the distribution of base proportions (for A,T,C,G) across B2 RNAs based on i) sequences in the reference (Repeat Masker, mm10) (REFERENCE - left panel), ii) read coverage in each sample with regard to the reference (EXPECTED – middle panel) and iii) actual base coverage observed in each sample (OBSERVED – right panel). Y axis of each plot represents the position in a B2 metagene combining all unique B2 RNA sequences (REPome) aligned at the start site of their consensus sequence with numbers representing the distance from the sequence start (+1). Props: cumulative proportions of four bases (cumulative proportion of all bases added together = 1). (1) as described in (Li & Durbin, 2009), (2) as described in (Picardi & Pesole, 2013).

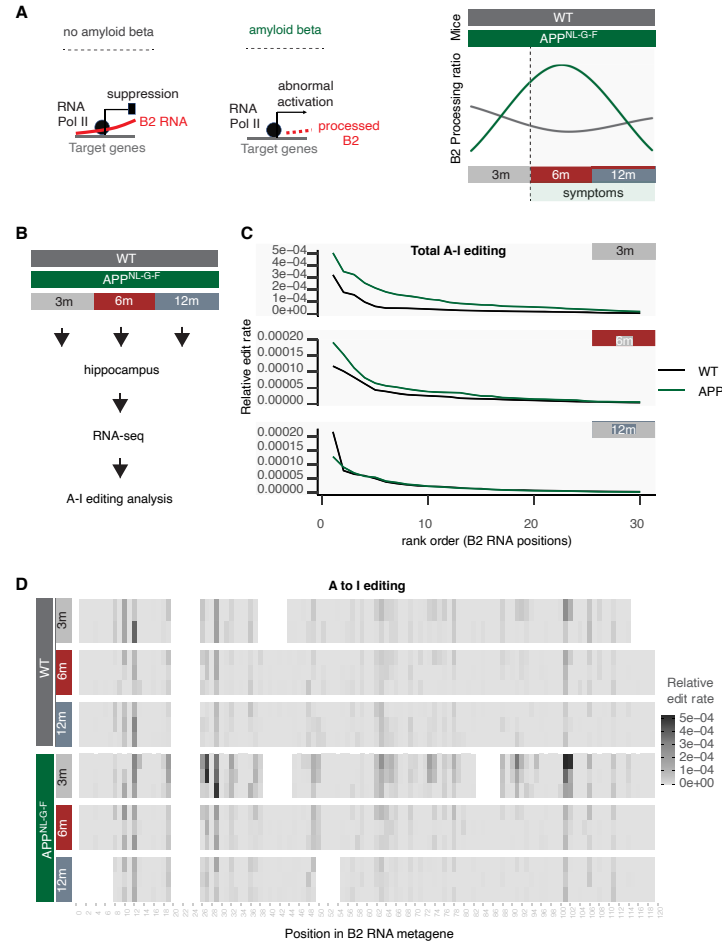


Figure 2. A-to-I editing is increased at specific B2 RNA positions and time points in a mouse model of amyloid beta pathology.

(A) Graphical overview of the mode of regulation of transcription by B2 RNAs and the connection of increased B2 RNA processing with amyloid beta pathology. Left panel depicts the suppression of RNA Polymerase II by full length B2 RNAs (based on (Espinoza, Goodrich, & Kugel, 2007; Ponicsan, Kugel, & Goodrich, 2010, 2015; Yakovchuk, Goodrich, & Kugel, 2009)) and the increased processing of B2 RNA and its destabilization in amyloid beta pathology (Cheng et al., 2020). Right panel is a graphical representation of the B2 RNA processing dynamics through time observed in the hippocampus of an amyloid pathology mouse model (APP) and the respective wild type (WT) control (based on (Cheng et al., 2020)).

(B) Analysis of RNA-seq datasets used for the study of A-to-I editing in the hippocampus during amyloid beta pathology. RNA-seq data sets from the hippocampus of an amyloid pathology mouse model (APP) and the respective wild type (WT) control are derived from our previous study (Cheng et al., 2020). 3m, 6m and 12 m stand for the age of mice in months.

(C) Comparison of total A-to-I editing rates between APP and WT mice per age group. A-to-I editing rate for each position and sample has been calculated as per Fig.1 and editing rates for samples of the same age group (3m, 6m or 12m old) and genotype (WT or APP) have been averaged and ranked in descending order. Only the 30 highest ranking positions are depicted.

(D) Relative A-to-I editing rate per sample (rows) and B2 RNA position (columns). Heatmap represents the editing rate calculated as in Fig.1 for all samples and the first 120 positions in B2 RNA that include the identified hotspots to increase image resolution.

Because figure 3F in Chapter 2 shows 6 month mice are the only age group with significantly higher full length levels (APP vs WT), while 3 month old WT mice have significantly lessened B2 RNA processing, we then queried what the consequences of lowered A-to-I editing might have on B2 RNAs' integrity. So, we designed a qPCR assay to measure relative B2 RNA levels. In the Ht22 mouse hippocampal cell line, we used the Adar and Adarb1 inhibitor 8-aza adenosine (8-aza) (Zhang & Slack, 2016). 8-aza is thought to be a competitive inhibitor of Adar and Adarb1, preventing their association with adenosine sites (Zhang & Slack, 2016). Additionally, we compared the incubation of 8-aza to an orthogonal approach, employing siRNAs to inhibit the expression of Adar and Adarb1 simultaneously. In both, we observed the overall expression of Adar and Adarb1 was significantly lowered (Fig 3A-B). Yet, when we assessed full length B2 RNA levels, they were significantly lowered when Adar and Adarb1 are inhibited (Figure 3C). As well, when we co-incubated amyloid peptides (as described in Chapter 2, figure 6A) with 8-Aza, we see that full length B2 RNA levels significantly recover (Figure 3D). To attribute these changes to difference in A-to-I editing, we used the enzyme Endonuclease V to process the RNAs at inosine sites. Endonuclease V is an inosine specific endonuclease when in the presence of magnesium (Knutson & Heemstra, 2020). Where the sample expresses less full length B2 RNA compared to untreated RNA, indicates the presence of relatively high inosine. If there is no change, the assay cannot detect high levels of inosine. What we see is for samples which showed lessened expression of Adar and Adarb1 in Fig 3A-B (8-Aza and Anti-adar), no significant change in full length B2 expression was detected (Fig 3E). However, the samples with normal Adar and Adarb1 expression in Fig 3A-B (DMSO and Scramble) showed significant reduction in full length B2 expression.

Interestingly, when we co-incubated amyloid with 8-Aza (Abeta(42)-Aza), amyloid apparently rescues the A-to-I edits compared to the reverse peptide incubation with 8-Aza (Abeta(Rev)-Aza) for which the assay did not produce a significant difference in full length B2 RNA levels (Fig 3E).

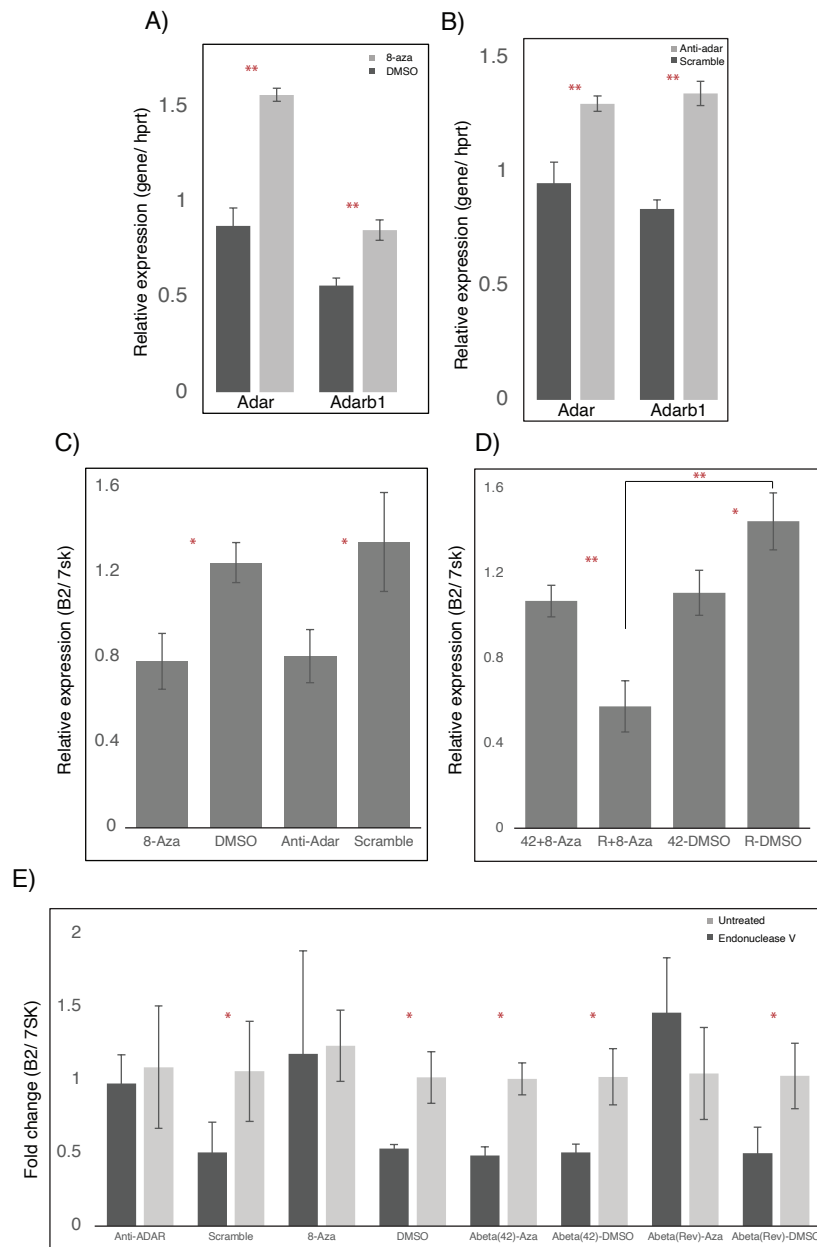


Figure 3 When A-to-I editing is reduced, full length B2 RNA levels are reduced.

(A-B) Relative Adar and Adarb1 expression normalized to Hprt when Ht-22 hippocampal cells are incubated with 8-aza and DMSO (A) and transfected with anti-Adar and anti-Adarb1 siRNA, and scramble siRNA (B) for 24 hours.

(C-D) Relative full length B2 RNA expression normalized to 7SK when Ht-22 hippocampal cells are treated as in (A) and (B), and with (1-42) amyloid beta peptides (42) or reverse peptides (R) in the presence of 8-Aza or DMSO for 6 hours (D).

(E) Fold change expression of full length B2 RNA normalized to 7SK when total RNA from the treated Ht-22 samples in (A-D) are incubated with endonuclease V or incubated in the assay buffer without enzyme (untreated).

Error bars represent standard deviation and * denotes a pvalue <0.05; ** denotes a pvalue <0.01. n=4 biological replicates.

The recovery of A-to-I editing despite incubation with 8-Aza suggests a substantial strength of effect by amyloid to stimulate A-to-I editing on B2 RNA. Indeed, Adarb1 is amongst the B2 SRGs list in supplementary table 1 and is significantly upregulated in the amyloid incubated Ht22 cells in supplementary table 1. Furthermore, Adar is upregulated in 6mo old APP mice (supplementary table 1).

Altogether, these results suggest there is a direct role by amyloid beta in stimulating an A-to-I editing response upon B2 RNAs. Conversely, A-to-I editing lessens as mice age, whereas in APP mice, 3mo and 6mo APP mice have significantly more A-to-I editing, but by the time the mice are 12mo, their A-to-I editing ratios in B2 RNAs normalized to wild type levels. Furthermore, there appears to be a direct relationship between A-to-I editing in B2 RNAs and B2 RNAs' full-length levels which is partially recovered by amyloid inoculation.

5.5.2 Interferon stimulated gene expression is elevated in 12 month APP NL-G-F mice

Because dsRNAs, including B2 RNAs are known to elicit an interferon response if they are too often cytoplasmic localized, and A-to-I editing is lessened, we assessed our model within the lens of interferon signaling with particular interest in the interferon stimulated gene (ISG) response pathway. Because the stoichiometry of A-to-I editing in B2 RNAs is lessened to a greater extent

between 3mo and 12mo mice as summarized in Fig 4A, we thought to assess the gene expression patterns by age. We performed a differential gene expression analysis between the differing age groups of our mice, within their respective APP⁺ vs wild type groups (3mo vs 6mo and 6mo vs 12mo) (Supplementary table 2). Of the 343 transcripts that had higher expression in 12mo APP⁺ mice (upregulated), 179 belong to known interferon inducible genes (Rusinova *et al.*, 2013) (Fig 4B). This was compared to only 80 out of the 155 transcripts that had higher expression in 12mo Wild type mice (Supplementary table 2). Moreover, while there were no significantly upregulated pathways and biological processes in wild type 12mo mice, 12mo APP⁺ mice are enriched in many immune regulation processes (Fig 4D). As well, there is significant enrichment in transcripts belonging to genes that localize to transport organelles like the endosome and vesicles, and the lysosome (Fig 4C). The lysosome pathway is also enriched (Fig 4E).

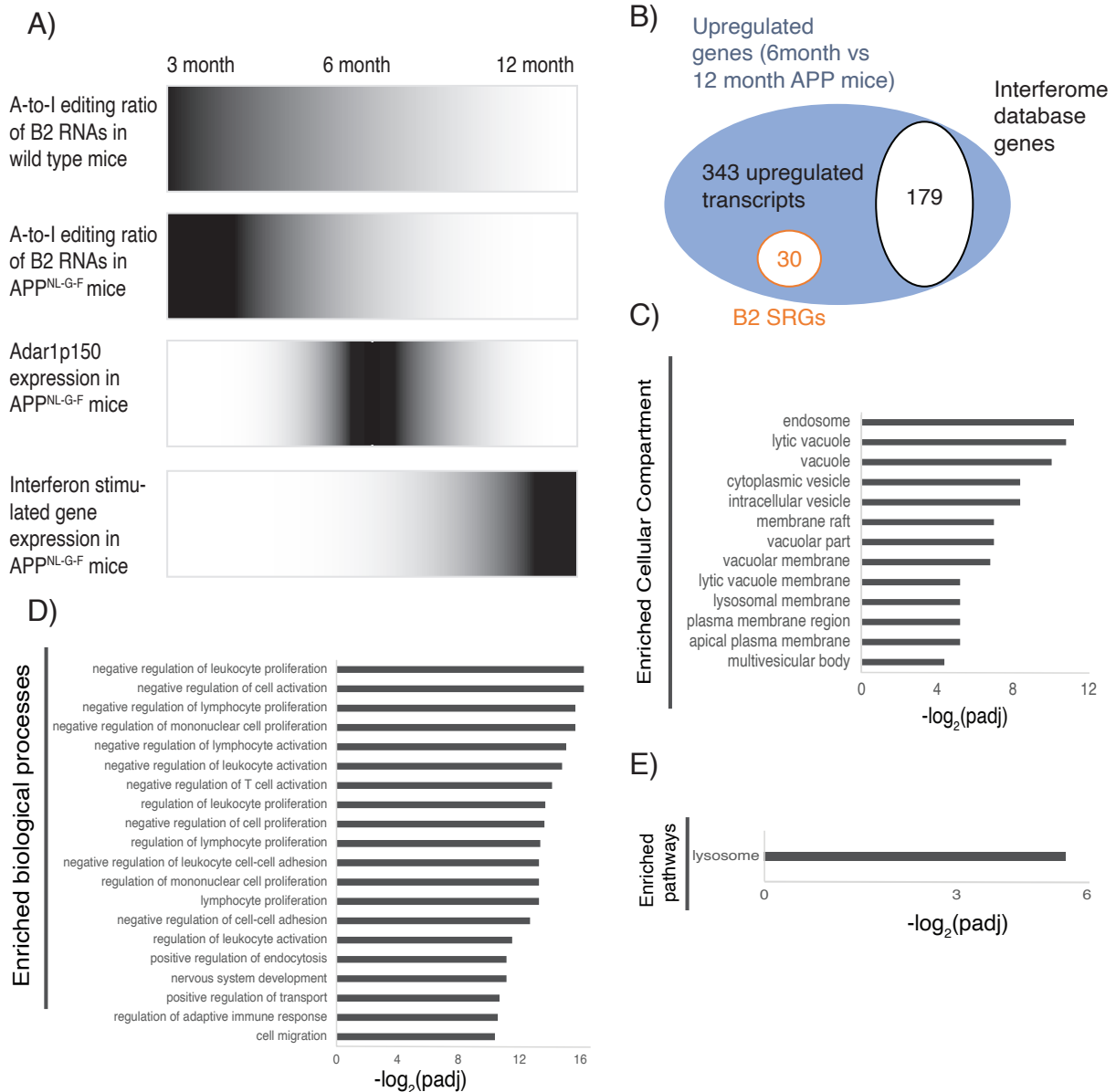


Figure 4. Interferon stimulated genes emerge in 12mo APP+ mice

(A) Timelapse schematic summarizing the overall changes in A-to-I editing of B2 RNA in wild type mice as they age (top row) and APP mice (Second row). Relative to the 3mo position, the lessening of A-to-I editing of B2 RNA is more gradual in Wild type mice whereas, in APP+ mice, the expression is relatively high in 3 and 6mo mice and the same between wild type and APP+ wild type mice. Moreover, Fig 5C shows increased Adar p150 expression in 6mo APP mice (third row) and a depletion by 12mo. 12mo APP mice are enriched in ISG expression.

(B) Summary of Interferome database overlap with the differential transcript expression. B2 SRGs are used as a comparison.

(C-E) Gene ontology analysis from DAVID of transcripts that are higher in 12mo APP+ mice. Padj values are inverse log (base2) for visualization.

- (C) Cellular compartment enrichment analysis.
- (D) Biological process enrichment analysis.
- (E) Kegg pathway enrichment analysis.

Taking the interferon stimulated genes that are more highly expressed in 12mo APP⁺ mice, we saw that indeed the normalized expression of these genes were significantly higher in the 12mo APP⁺ mice than the 6mo APP⁺ mice and the 12mo Wild type mice (Fig 5D-E). This suggested that in fact, an interferon response was indeed elicited. To attribute this response to changes in A-to-I stoichiometry in B2 RNAs, we report the transcript level analysis for the interferon stimulated Adar p150 (Vogel *et al*, 2020) which blocks the MDA5-MAVS response, limiting inflammation and ISG response (Pestal *et al.*, 2015). We find that 6mo APP⁺ mice have elevated normalized expression of Adar p150 compared to 6mo Wild type mice and 12mo APP⁺ mice (Fig 5C). In fact, 6mo APP⁺ mice also have elevated expression of the entire Adar and Adarb1 genes, despite in Wild type mice, 3mo mice having elevated expression of Adar and Adarb1 compared to 6mo Wild type mice (Fig 5A-B). These data suggest Adar and particularly, Adar p150 expression is elevated before ISGs are elevated.

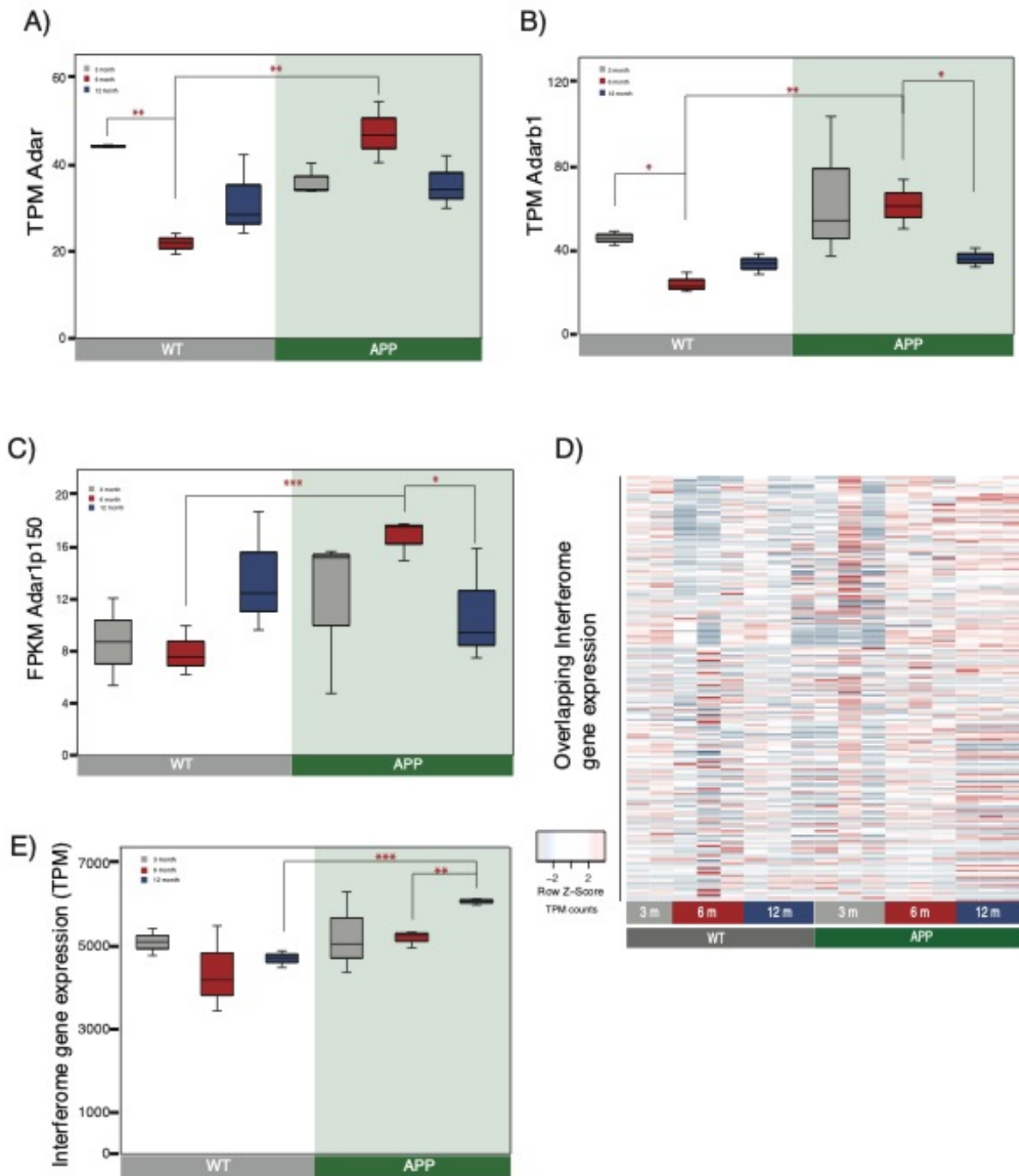


Figure 5. Adar and Adarb1 expression is elevated before interferon stimulated genes.

(A-C, E) Box and whisker plots depicting the transcripts per million (A,B), fragments per kilobase of exon per million mapped fragments (C) and sum of the transcripts per million (E) of the Adar gene (A), the Adarb1 gene (B), the Adar p150 transcript (C) and the 179 SRGs with higher expression in 12mo APP+ mice (E).

(D) Heatmap depicting the genes with transcripts upregulated in 12mo APP⁺ mice that overlap with the interferome database genes. Normalized within rows.

* depicts pvalue < 0.05, ** depicts pvalue < 0.01, *** depicts a pvalue < 0.001.

Altogether, we show that A-to-I editing in B2 RNAs is elevated in 3mo and 6mo APP⁺ mice, normalizing to wild type levels in 12mo mice. When we inoculate HT22 cells with amyloid beta peptides, we similarly see higher A-to-I editing and higher Adar and Adarb1 expression. Inhibiting Adar and Adarb1 with 8-Aza and siRNAs lowers full length B2 RNA levels, but full length B2 RNA levels are rescued when we inoculated with amyloid beta peptides. We are able to attribute the recovery of full length B2 RNA levels to a recovery in A-to-I editing. Finally, we show that ISGs are more highly expressed in 12mo APP⁺ with multiple elevated biological processes with roles in immune function. Further, the lysosome pathway is significantly elevated, while the enriched cellular compartments belong to many transport and lytic structures in the cell. Furthermore, Adar p150 expression is elevated before the ISG response.

5.5.3 NERD-Seq can enable RNA modification detection

After looking at A-to-I editing in B2 RNAs, we decided to expand our RNA modification analysis beyond B2 RNAs and A-to-I editing. To confirm that the NERD-seq protocol (developed in chapter 4) is capable of detecting modifications, we *in vitro* transcribed (IVT) the human 7SK construct and performed the protocol on externally sourced human cerebral cortex total RNA. We then used Nanopolish tools (Simpson *et al*, 2017) to parse kmers by transcript and using a student's t-test, determined which regions were producing significantly different voltage patterns (Fig 6A). Doing so, we were able to detect a significant shift in two kmers with known pseudouridine modifications (pseudo 242 and 244) and that matched many m6A modification sites reported in (Leger *et al*, 2021). In fact, the standard protocol was unable to capture enough 7SK reads to reproduce modification sites, compared to the no-edit IVT product (Fig 9B-C).

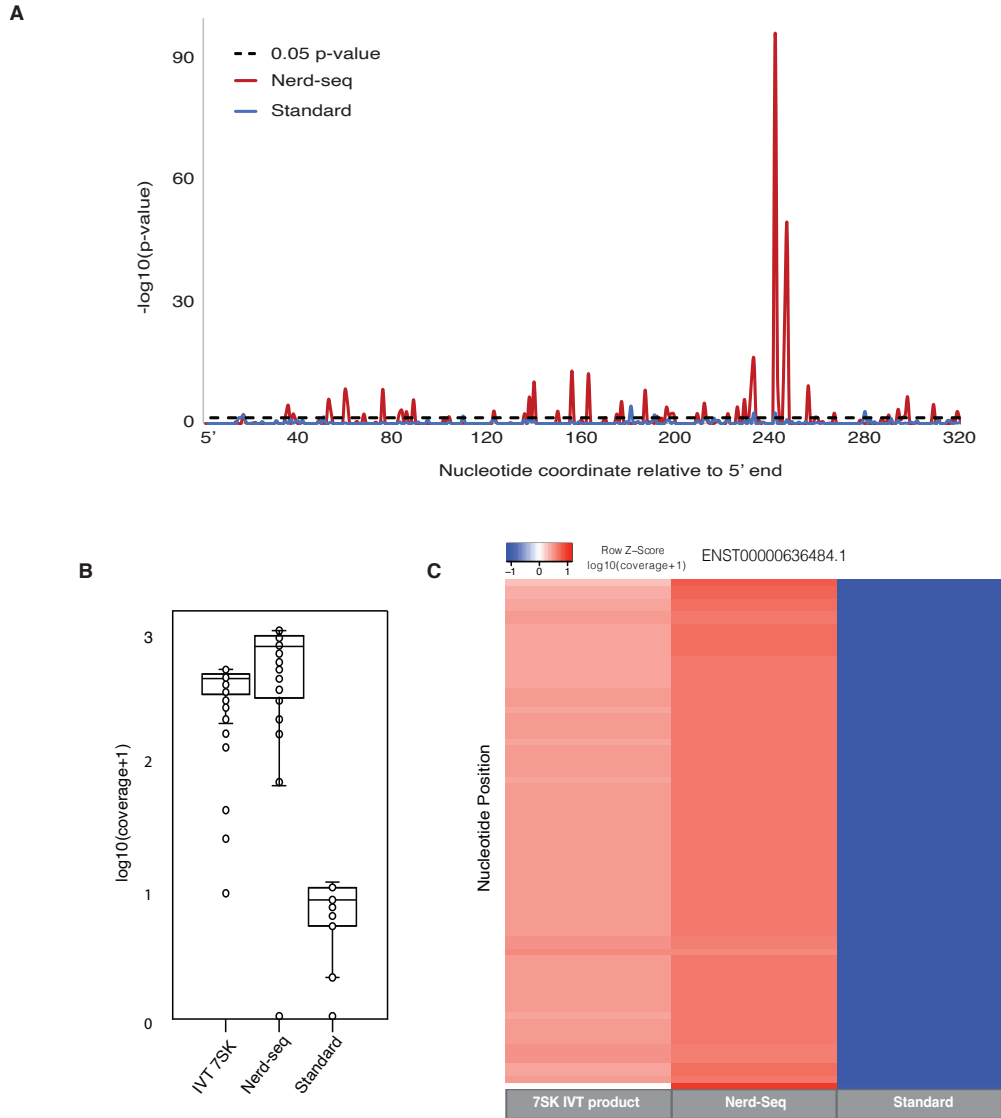


Figure 6. NERD-seq detects RNA modifications in Rn7SK

(A) Line plot depicting the inverse log of the student's t-test pvalue of all kmers between Nerd-seq and IVT (red), and Standard and IVT (blue). The black dotted line denotes the 0.05 pvalue cutoff.

(B) Box and whisker plot depicting the $\log(\text{coverage}+1)$ of each kmer from 7SK.

(C) Heatmap depicting the coverage of each kmer by nucleotide position in 7SK. Values are normalized by row and taken as a function of $\log(\text{coverage}+1)$.

5.5.4 APP+ mice have differentially modified RNAs

Using the NERD-seq methodology, we resequenced the APP+ and wild type mice and applied an RNA modification analysis called XPore (Pratanwanich *et al.*, 2021) which uses Nanopolish to assess the RNAs which were differentially modified between the two samples. According to the XPore output, we uncovered 2611 significantly different gene and kmer combinations with more than 50 kmer read coverage from 141 unique genes (Supplementary table 3). Of the unique genes with significant modification predictions, a significant number are involved in the translation related pathways (Fig 7A) and biological processes (Fig 7B), with many belonging to mitochondrial rRNAs and tRNAs, 5S rRNA, snoRNAs and BC1 RNA.

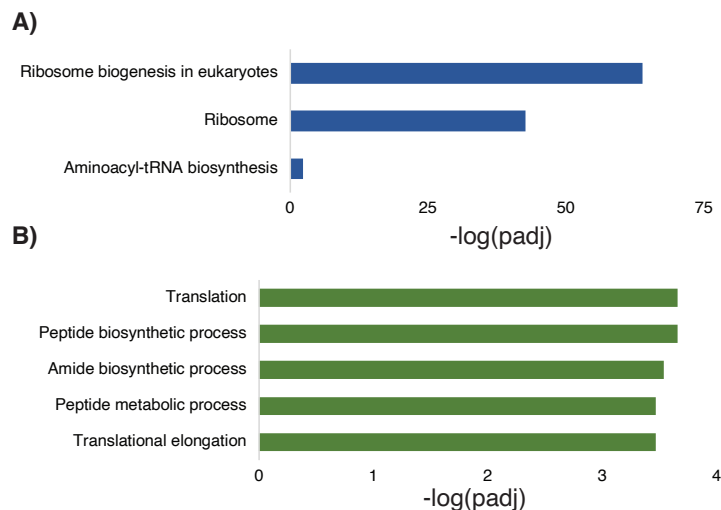


Figure 7. APP+ modified RNAs are primarily involved in translation.

(A-B) Gene ontology analysis from DAVID of genes from the XPore output. Padj values are inverse log (base2) for visualization.

(A) Kegg pathway enrichment analysis

(B) Biological process (BP5) enrichment analysis.

When looking at BC1 RNA specifically, of the two predicted modification sites, both fall within predicted 2'-O-methylation sites including G32, C33, and G61 (Lacoux *et al.*, 2012) (Fig 8B). As well, BC1's expression is significantly lower between 6mo APP+ and wild type, and 12mo APP+ and wild type mice. It is unclear if the changes to the modification stoichiometry and the expression levels are related but both data together suggest further that BC1 is dysregulated in AD models.

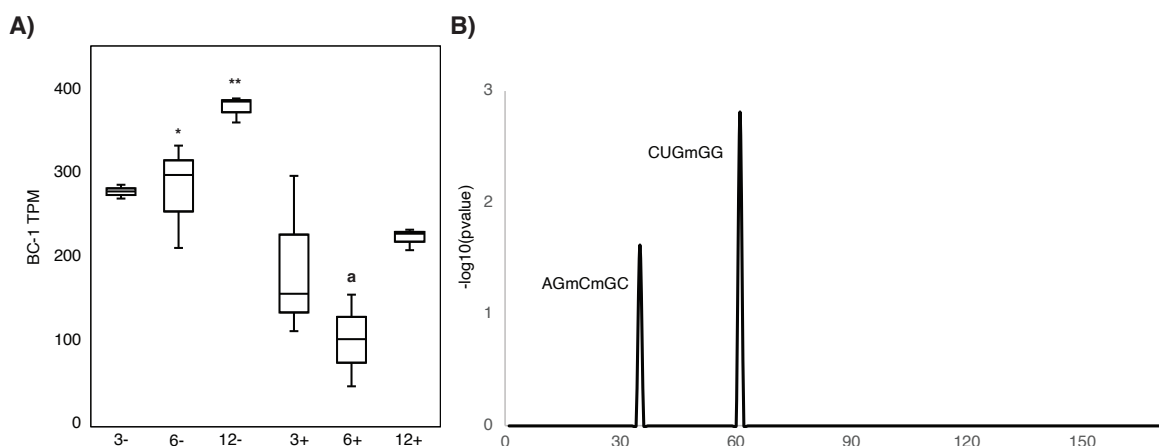


Figure 8. BC1 is dysregulated in APP+ mice.

(A) Box and whisker plot showing BC1 TPM normalization expression values from Illumina sequencing data. *denotes pvalue <0.05 between 6mo wild type and 6mo APP+, ** denotes pvalue <0.01 between 12mo wild type and 12mo APP+ and “a” denotes a pvalue <0.05 between 6mo APP+ and 12mo APP+.

(B) Line plot depicting the inverse log of the pvalue for significant prediction kmers across the BC1 sequence. Methylation sites from the literature are shown within the kmer sequences by an “m” adjacent to its corresponding nucleotide.

Additionally, when looking more closely at 5S rRNA, we can see that as a class their average expression is significantly lower in APP+ mice, than wild type (Fig 9). Whether the modifications reported in supplementary table 3 are related to the lowered expression is also unclear; however, because of the elevated ROS in APP+ mice and the fact that 5S rRNA are sensitive to ROS, it is

possible many of the differential modification predictions in supplementary table are in fact oxidation products.

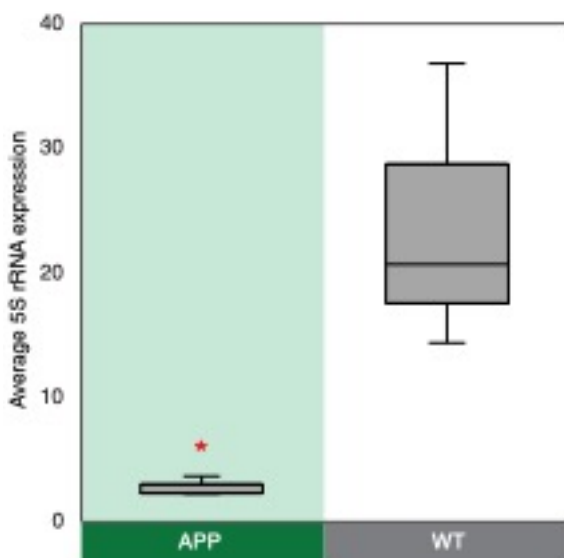


Figure 9. 5S rRNA levels are reduced in APP+ mice.

Box and whisker plot depicting the average expression of all 5S rRNA genes produced from the Seqmonk annotated probe report. * denotes a pvalue <0.05.

Taken together, these data contribute further evidence that ncRNAs are dysregulated in AD models and in particular, RNA modification chemistry may play an important role in AD. As well, the predicted modification sites produced in supplementary table will be an important resource for researchers interested in assessing RNA modifications in the APP NL-G-F model. Although the predicted modifications are agnostic, meaning the exact modification it refers to is unknown, it can still serve as a useful guide to suggest what RNAs to look at and where within the RNAs to look.

5.6 Discussion

In human health, RNA modifications are known contributors to cell function. Dysregulation of RNA modification processes can be predictive of poor health. In leukaemia stem cells, high m6A

levels maintain high levels of translation for pro-survival genes (Barbieri & Kouzarides, 2020). As well, dementias have differing levels of enrichment in various RNA modifications (Cm, Um, m7G, m1G, m2,2,7G and Psi) (Zhang *et al*, 2020b). It is still not entirely clear what effect RNA modifications have on RNAs themselves. It is known that A-to-I editing disturbs double stranded regions because inosine base pairing to uracil is less stable than uracil to adenosine (Bass & Weintraub, 1987; Rebagliati & Melton, 1987). However, it is not well understood what consequence that has on the RNAs' stability, as well as the consequences of other RNA modifications on RNAs.

Here we unveil that amyloid beta stimulates A-to-I editing in 3mo and 6mo mice, representing a before unreported mechanism of transcriptome dysregulation. Many studies have reported that A-to-I editing is decreased in AD patients which is similar to our finding that 12mo APP+ mice have much less editing than 3mo or 6mo mice (Khermesh *et al*, 2016; Wu *et al*, 2021). In addition, it has also been reported that Alu RNA features reduced editing in hippocampus vasculature in AD patients (Crooke *et al.*, 2022). It is unclear how well the 3mo and 6mo mice's response to amyloidosis would translate to human patients but assessing individuals early in their amyloidosis stages may help translate these findings. Furthermore, our results show that when A-to-I editing is reduced, less full length B2 RNAs can be detected, meaning A-to-I editing is important in some way for stability or prevents an interaction in the cell that processes B2 RNA. Dicer for example can process dsRNAs and is known to target B2 RNAs, whereby deficits in Dicer can be toxic for cells because of the accumulation of RNAs like B2 RNA (Kaneko *et al*, 2011; White *et al*, 2014; Yusuke Shiromoto*, 2020). If Adar unwinds the structure of B2 RNA by A-to-I editing, it could

prevent Dicer's action upon it, but more evidence is needed to causally link enzymatic processing of B2 to the phenomenon of lowered full-length levels.

Additionally, we report that at 12mo, APP⁺ mice show features of interferon response, whereby Adarp150 is downregulated between 6mo and 12mo APP⁺ mice, possibly allowing the activation of MDA5-MAVS signalling (Pestal *et al.*, 2015) and thus, the activation of ISGs. Furthermore, multiple immune system biological pathways are upregulated which is a common feature of interferon signalling in brain tissue (Kramer *et al.*, 2019). Cellular compartment analysis also showed enrichment of transport molecules such as the endosome and vesicles. The endosome specifically has been shown to house IFN and enable a low-level, long-term response (Altman *et al.*, 2020). Lastly, the lysosome pathway is upregulated in 12mo APP⁺ mice. The lysosome pathway is known to be remodeled by the ISG response, changing its pH and structure in preparation for the breakdown of potentially pathogenic molecules (Zhang *et al.*, 2020a). The lysosomal pathway is also known to be dysregulated in AD (Whyte *et al.*, 2017) and the APP NL-G-F model (Whyte *et al.*, 2020). Altogether, these results suggest the interferon pathway is activated in 12mo APP⁺ mice, which has been shown to lead to synaptic loss and neurodegeneration in AD (Roy *et al.*, 2020). If it is true that the depletion of A-to-I editing stoichiometry in B2 RNAs leads to an over accumulation of dsRNAs in APP⁺ mice, this could be an important feature of AD pathology. Of course, they may be other factors which contribute to increased interferon activity, including ROS accumulation (Izumi *et al.*, 2020) and microglia activation (Sebastian Monasor *et al.*, 2020).

Beyond A-to-I editing in B2 RNAs we also applied a transcriptome-wide approach to RNA modification differences between 6mo APP⁺ and wild type mice. The results show enrichment in modified RNAs primarily within protein synthesis pathways. Because the ribosome is dysregulated early on in AD (Ding *et al.*, 2005; Ding *et al.*, 2012), it is possible these results suggest RNA modification changes are responsible and that further attention on RNA modifications in the ncRNA space is due. Additionally, we are able to replicate RNA modifications in known modification sites for BC1, particularly the 2'-O-methylation sites G32, C33, and G61 (Lacoux *et al.*, 2012). These methylation sites are thought to regulate the interaction of BC1 with FMRP, preventing synaptic localization. While the agnostic nature of the XPore comparison does not allow us to infer whether BC1 is more or less modified, we can suggest FMRP and BC1's interactions are dysregulated in APP⁺ mice. Because FMRP regulates APP expression (Zhang *et al.*, 2018) it is possible BC1's modification dysregulation also plays a role in the changes to APP translation.

Altogether our results show that RNA modifications in APP⁺ mice are dysregulated, leading to dysfunction in important cellular processes. RNA modification and ncRNAs, particularly with regard to their impact on cell function is a growing field of study and employing that lens of study in AD may lead to better understanding of the molecular dysregulation that takes place.

5.7 References

Altman JB, Taft J, Wedeking T, Gruber CN, Holtmannspotter M, Piehler J, Bogunovic D (2020) Type I IFN is siloed in endosomes. *Proc Natl Acad Sci U S A* 117: 17510-17512

- Anderson BR, Muramatsu H, Nallagatla SR, Bevilacqua PC, Sansing LH, Weissman D, Kariko K (2010) Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res* 38: 5884-5892
- Asher S, Priefer R (2022) Alzheimer's disease failed clinical trials. *Life Sci* 306: 120861
- Bajad P, Ebner F, Amman F, Szabo B, Kapoor U, Manjali G, Hildebrandt A, Janisiw MP, Jantsch MF (2020) An internal deletion of ADAR rescued by MAVS deficiency leads to a minute phenotype. *Nucleic Acids Res* 48: 3286-3303
- Bakhtiarizadeh MR, Salehi A, Rivera RM (2018) Genome-wide identification and analysis of A-to-I RNA editing events in bovine by transcriptome sequencing. *PLoS One* 13: e0193316
- Barbieri I, Kouzarides T (2020) Role of RNA modifications in cancer. *Nat Rev Cancer* 20: 303-322
- Bass BL, Weintraub H (1987) A developmentally regulated activity that unwinds RNA duplexes. *Cell* 48: 607-613
- Begik O, Mattick JS, Novoa EM (2022) Exploring the epitranscriptome by native RNA sequencing. *RNA*
- Bloom GS (2014) Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* 71: 505-508
- Brachova P, Alvarez NS, Hong X, Gunewardena S, Vincent KA, Latham KE, Christenson LK (2019) Inosine RNA modifications are enriched at the codon wobble position in mouse oocytes and eggshells. *Biol Reprod* 101: 938-949
- Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* 3: 186-191
- Burg N, Bittner S, Ellwardt E (2018) Role of the epigenetic factor Sirt7 in neuroinflammation and neurogenesis. *Neurosci Res* 131: 1-9
- Cheng Y, Saville L, Gollen B, Isaac C, Belay A, Mehla J, Patel K, Thakor N, Mohajerani MH, Zovoilis A (2020) Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. *Elife* 9
- Cheng Y, Saville L, Gollen B, Veronesi AA, Mohajerani M, Joseph JT, Zovoilis A (2021) Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes. *EMBO Rep* 22: e52255
- Cornutiu G (2015) The Epidemiological Scale of Alzheimer's Disease. *J Clin Med Res* 7: 657-666

- Crooke PS, 3rd, Tossberg JT, Heinrich RM, Porter KP, Aune TM (2022) Reduced RNA adenosine-to-inosine editing in hippocampus vasculature associated with Alzheimer's disease. *Brain Commun* 4: fcac238
- Ding Q, Markesbery WR, Chen Q, Li F, Keller JN (2005) Ribosome dysfunction is an early event in Alzheimer's disease. *J Neurosci* 25: 9171-9175
- Ding Q, Zhu H, Zhang B, Soriano A, Burns R, Markesbery WR (2012) Increased 5S rRNA oxidation in Alzheimer's disease. *J Alzheimers Dis* 29: 201-209
- Funkhouser SA, Steibel JP, Bates RO, Raney NE, Schenk D, Ernst CW (2017) Evidence for transcriptome-wide RNA editing among *Sus scrofa* PRE-1 SINE elements. *BMC Genomics* 18: 360
- Gao S, Burney HN, Callahan CM, Purnell CE, Hendrie HC (2019) Incidence of Dementia and Alzheimer Disease Over Time: A Meta-Analysis. *J Am Geriatr Soc* 67: 1361-1369
- Gassner FJ, Zaborsky N, Buchumenski I, Levanon EY, Gatterbauer M, Schubert M, Rauscher S, Hebenstreit D, Nadeu F, Campo E *et al* (2021) RNA editing contributes to epitranscriptome diversity in chronic lymphocytic leukemia. *Leukemia* 35: 1053-1063
- Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57
- Huang YM, Shen J, Zhao HL (2019) Major Clinical Trials Failed the Amyloid Hypothesis of Alzheimer's Disease. *J Am Geriatr Soc* 67: 841-844
- Izumi H, Sato K, Kojima K, Saito T, Saido TC, Fukunaga K (2020) Oral glutathione administration inhibits the oxidative stress and the inflammatory responses in App(NL-G-F/NL-G-F) knock-in mice. *Neuropharmacology* 168: 108026
- Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA *et al* (2011) DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* 471: 325-330
- Khmermesh K, D'Erchia AM, Barak M, Annese A, Wachtel C, Levanon EY, Picardi E, Eisenberg E (2016) Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* 22: 290-302
- Kim Y, Tarallo V, Kerur N, Yasuma T, Gelfand BD, Bastos-Carvalho A, Hirano Y, Yasuma R, Mizutani T, Fowler BJ *et al* (2014) DICER1/Alu RNA dysmetabolism induces Caspase-8-mediated cell death in age-related macular degeneration. *Proc Natl Acad Sci U S A* 111: 16082-16087
- Knutson SD, Heemstra JM (2020) EndoVIPER-seq for Improved Detection of A-to-I Editing Sites in Cellular RNA. *Curr Protoc Chem Biol* 12: e82

- Kramer TJ, Hack N, Bruhl TJ, Menzel L, Hummel R, Griemert EV, Klein M, Thal SC, Bopp T, Schafer MKE (2019) Depletion of regulatory T cells increases T cell brain infiltration, reactive astrogliosis, and interferon-gamma gene expression in acute experimental traumatic brain injury. *J Neuroinflammation* 16: 163
- Lacoux C, Di Marino D, Boyl PP, Zalfa F, Yan B, Ciotti MT, Falconi M, Urlaub H, Achsel T, Mougin A *et al* (2012) BC1-FMRP interaction is modulated by 2'-O-methylation: RNA-binding activity of the tudor domain and translational regulation at synapses. *Nucleic Acids Res* 40: 4086-4096
- Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Miano V, Migliori V, Toolan-Kerr P, Sideri T, Enright AJ *et al* (2021) RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun* 12: 7198
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589-595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079
- Llanos-Gonzalez E, Henares-Chavarino AA, Pedrero-Prieto CM, Garcia-Carpintero S, Frontinan-Rubio J, Sancho-Bielsa FJ, Alcain FJ, Peinado JR, Rabanal-Ruiz Y, Duran-Prado M (2019) Interplay Between Mitochondrial Oxidative Disorders and Proteostasis in Alzheimer's Disease. *Front Neurosci* 13: 1444
- Luo Q, Chen Y (2016) Long noncoding RNAs and Alzheimer's disease. *Clin Interv Aging* 11: 867-872
- Mansi L, Tangaro MA, Lo Giudice C, Flati T, Kopel E, Schaffer AA, Castrignano T, Chillemi G, Pesole G, Picardi E (2021) REDiportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Res* 49: D1012-D1019
- Pestal K, Funk CC, Snyder JM, Price ND, Treuting PM, Stetson DB (2015) Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development. *Immunity* 43: 933-944
- Picardi E, Pesole G (2013) REDiTools: high-throughput RNA editing detection made easy. *Bioinformatics* 29: 1813-1814
- Polesskaya O, Kananykhina E, Roy-Engel AM, Nazarenko O, Kulemzina I, Baranova A, Vassetsky Y, Myakishev-Rempel M (2018) The role of Alu-derived RNAs in Alzheimer's and other neurodegenerative conditions. *Med Hypotheses* 115: 29-34
- Pratanwanich PN, Yao F, Chen Y, Koh CWQ, Wan YK, Hendra C, Poon P, Goh YT, Yap PML, Chooi JY *et al* (2021) Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol*

- Rebagliati MR, Melton DA (1987) Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell* 48: 599-605
- Roy ER, Wang B, Wan YW, Chiu G, Cole A, Yin Z, Propson NE, Xu Y, Jankowsky JL, Liu Z *et al* (2020) Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *J Clin Invest* 130: 1912-1930
- Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, Chapman R, Hertzog PJ (2013) Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res* 41: D1040-1046
- Saito T, Matsuba Y, Mihira N, Takano J, Nilsson P, Itohara S, Iwata N, Saido TC (2014) Single App knock-in mouse models of Alzheimer's disease. *Nat Neurosci* 17: 661-663
- Schaffer AA, Levanon EY (2021) ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions. *Methods Mol Biol* 2181: 149-162
- Sebastian Monasor L, Muller SA, Colombo AV, Tanrioever G, Konig J, Roth S, Liesz A, Berghofer A, Piechotta A, Prestel M *et al* (2020) Fibrillar Abeta triggers microglial proteome alterations and dysfunction in Alzheimer mouse models. *Elife* 9
- Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 50: W216-221
- Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14: 407-410
- Vogel OA, Han J, Liang CY, Manicassamy S, Perez JT, Manicassamy B (2020) The p150 Isoform of ADAR1 Blocks Sustained RLR signaling and Apoptosis during Influenza Virus Infection. *PLoS Pathog* 16: e1008842
- Wang SB, Narendran S, Hirahara S, Varshney A, Pereira F, Apicella I, Ambati M, Ambati VL, Yerramothu P, Ambati K *et al* (2021) DDX17 is an essential mediator of sterile NLRC4 inflammasome activation by retrotransposon RNAs. *Sci Immunol* 6: eabi4493
- White E, Schlackow M, Kamieniarz-Gdula K, Proudfoot NJ, Gullerova M (2014) Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* 21: 552-559
- Whyte LS, Hassiotis S, Hattersley KJ, Hemsley KM, Hopwood JJ, Lau AA, Sargeant TJ (2020) Lysosomal Dysregulation in the Murine App(NL-G-F/NL-G-F) Model of Alzheimer's Disease. *Neuroscience* 429: 143-155
- Whyte LS, Lau AA, Hemsley KM, Hopwood JJ, Sargeant TJ (2017) Endo-lysosomal and autophagic dysfunction: a driving factor in Alzheimer's disease? *J Neurochem* 140: 703-717

- Wu S, Yang M, Kim P, Zhou X (2021) ADeditome provides the genomic landscape of A-to-I RNA editing in Alzheimer's disease. *Brief Bioinform* 22
- Yusuke Shiromoto* MS, Helen Qu, Andrew V Kossenkova, and Kazuko Nishikura* (2020) Processing of Alu small RNAs by DICER/ADAR1 complexes and their RNAi targets. *RNA* 26
- Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, Oostra B, Bagni C (2003) The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112: 317-327
- Zhang H, Zoued A, Liu X, Sit B, Waldor MK (2020a) Type I interferon remodels lysosome function and modifies intestinal epithelial defense. *Proc Natl Acad Sci U S A* 117: 29862-29871
- Zhang T, Pang P, Fang Z, Guo Y, Li H, Li X, Tian T, Yang X, Chen W, Shu S *et al* (2018) Expression of BC1 Impairs Spatial Learning and Memory in Alzheimer's Disease Via APP Translation. *Mol Neurobiol* 55: 6007-6020
- Zhang WC, Slack FJ (2016) ADARs Edit MicroRNAs to Promote Leukemic Stem Cell Activity. *Cell Stem Cell* 19: 141-142
- Zhang X, Trebak F, Souza LAC, Shi J, Zhou T, Kehoe PG, Chen Q, Feng Earley Y (2020b) Small RNA modifications in Alzheimer's disease. *Neurobiol Dis* 145: 105058

Chapter 6 - Conclusion

6.1 Conclusions and future directions

In the above chapters, I examined AD models and AD patient data within the lens of ncRNAs, uncovering novel findings about the function of SINE B2 and Alu RNAs in AD and particularly, amyloid pathology. I found that B2 and Alu RNAs are potent regulators of stress response in normal cells (Chapter 2-3), but in our APP⁺ model, B2 RNAs are aberrantly processed in 6mo mice, leading to a cascade of (SRG) activation (Chapter 2). The study in Chapter 2 unveils for the first time how B2 RNAs' inhibition of RNA Pol II at SRG sites may function within a pathology. Similarly, the findings in Chapter 3 demonstrate that AD patients similarly feature aberrant processing of Alu RNA all the way into late age AD, as well as correlating with many molecular markers and pathological tests used to diagnose AD. The findings suggest strong correlation with Alu processing and the genes upregulated in AD, and when we knockdown Alu in a cell model, many of the genes upregulated in AD feature higher expression (Chapter 3). In both cases I show that B2 and Alu RNAs are processed specifically and efficiently by Hsf1 and gather evidence to show that Hsf1 may be recruited to process B2 and Alu RNAs *in vivo* to elicit a stress response. These findings are further evidence that processing of B2 and Alu RNAs permits an RNA Pol-II mediated stress response (Hernandez *et al*, 2020; Zovoilis *et al*, 2016) and during normal conditions, B2 and Alu RNA are important inhibitors of the stress response (Espinoza *et al*, 2004; Ponicsan *et al*, 2015; Yakovchuk *et al*, 2009).

Beyond the phenomenon of B2 and Alu RNA processing, I also showed in Chapter 5 that B2 RNA is aberrantly A-to-I modified in APP⁺ mice while also providing evidence that the modification dysregulation that occurs in particularly 12mo APP⁺ mice may lead to an interferon response and thus contribute to synaptic loss and neuroinflammation. This is an important finding because it

provides further confirmation to emerging evidence that dysregulation in A-to-I editing in B2 RNAs can elicit this type of response in disease systems (Gal-Mark *et al*, 2017). Similarly, the phenomenon of A-to-I dysregulation has been reported in Alu RNAs in other disease systems (Crooke *et al*, 2021; Vlachogiannis *et al*, 2020) and in AD (Crooke *et al*, 2022). Further evidence is required to assess whether reduced A-to-I editing of Alu RNAs reduces their full-length levels like in B2 RNA, as reported in chapter 5.

Moreover, I demonstrate that other ncRNAs feature dysregulated RNA modification profiles in 6mo APP⁺ mice using a custom RNA sequencing approach called NERD-seq (Chapter 4) and an XPore analysis pipeline (Pratanwanich *et al*, 2021) which shows general dysregulation of RNAs involved in protein synthesis pathways (Chapter 5). Overall, Chapter 5 is a body of work that suggests assessing RNA modifications in AD and AD models may help elucidate some of the more elusive pathologies of AD and amyloid pathology. If targeting amyloid does not slow the progression of AD (Asher & Priefer, 2022), it is time to focus on the underlying causes of amyloid beta dysregulation and clarifying nuances in the molecular etiologies. The supplementary table 3 included in this thesis will be an important guide to researchers looking at epitranscriptomic markers in AD and while the modifications presented are agnostic (not assigned to a modification), they should narrow down which RNAs are differentially modified. Indeed, it is already known that RNA modifications are dysregulated in AD (Crooke *et al*, 2022; Zhang *et al*, 2020) but annotating where the modifications occur, their impact on the host RNA, and the underlying effect on the cells will enable the advancement of our understanding of the molecular mechanisms governing AD.

How all of the different factors contribute synergistically to AD pathology is also unclear and will be important for future study. For instance, elevated processing of B2 and Alu RNAs occurs at the same time as elevated A-to-I editing. In addition to potentially inhibiting ISG expression, does A-to-I editing – which occurs in both cytoplasmic and nuclear space – have an effect on how B2 and Alu RNA inhibit Pol-II? It has long been known A-to-I editing inhibits the formation of dsRNA structures (Bass & Weintraub, 1987;1988), and because of the importance of dsRNA for many RNA:protein interactions (Saunders & Barber, 2003; Ramos *et al*, 2022), it is possible dysregulation in A-to-I editing changes the RNA protein interaction landscape, including B2 and Alu RNAs' interactions with Pol-II and other proteins like Hsf1.

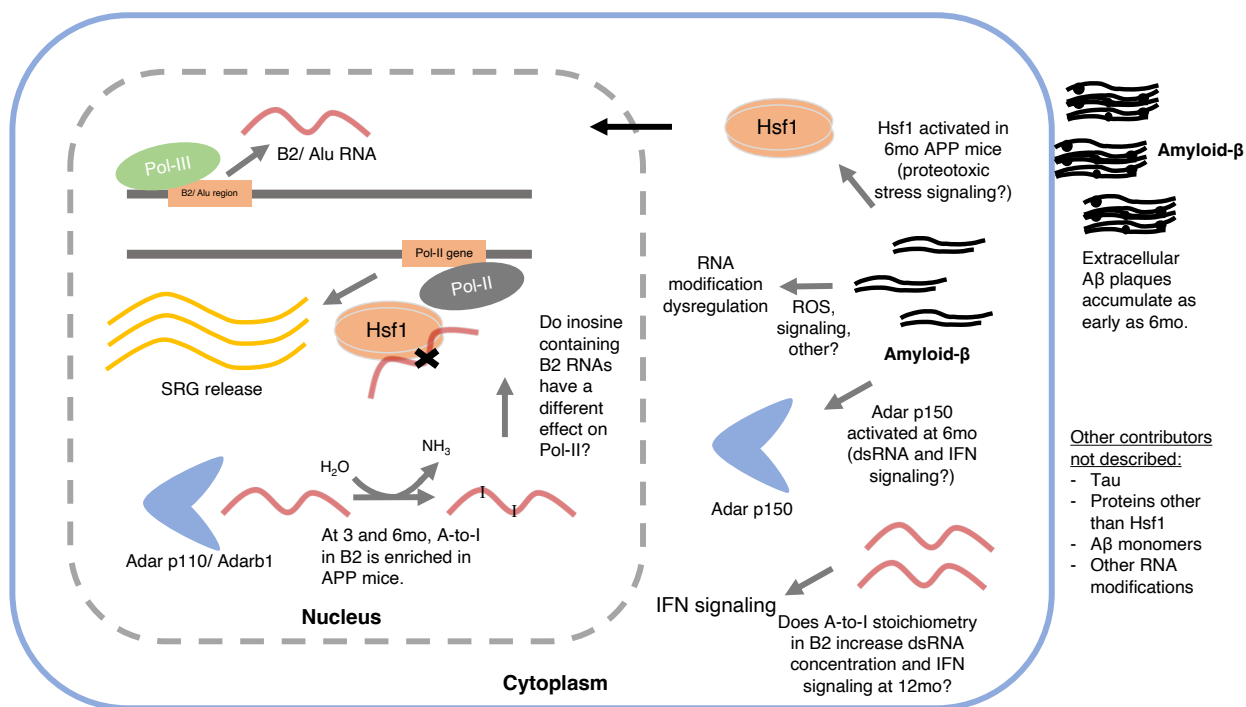


Figure 1. Model of ncRNA derived dysregulation in amyloid beta neuropathology. Summary of how ncRNAs, particularly the differing stages of B2 and Alu RNAs and their corresponding A-to-I editing stoichiometry and processing are involved in AD and how amyloid beta may be situated.

Of course, amyloid beta is not the only contributor to AD pathology. Tau neurofibrillary tangles – which occur with high interplay to amyloid beta (Castanho *et al*, 2020) – are important intracellular aggregates which contribute to neurodegeneration in many neurodegenerative disorders including AD (Braak & Braak, 1991; Lee *et al*, 2001; Dregni *et al*, 2022). It is unclear how Tau dysregulation may interact with ncRNAs. The role of miRNAs (such as miR-27a-3p’s inhibition of an agent of Tau hyperphosphorylation: GSK-3 β) in regulating Tau expression is well reviewed in Millan (2018), but little is known regarding the interplay between Tau and other ncRNAs, especially direct interactions, if any (Millan, 2018). Interestingly, changes in Braak stage and Tau burden do correlate with changes in transposable element expression, including changes in SINE RNA expression (Guo *et al*, 2018; Ramirez *et al*, 2022). Of course, in Chapter 3, we also show Alu RNA processing correlates with Braak stage and Przybyla *et al* (2020) show the P301S mouse tau model exhibit an immediate early gene signature which corresponds to elevated stress response (Przybyla *et al*, 2020). Further assessment of the contributions of Tau and amyloid-beta on transcriptome changes independently and together may help elucidate further how Alu and B2 SINE RNA processing is accelerated and the mechanism by which protein factors may be recruited to enzymatically accelerate SINE RNA’s processing.

The contributions of Hsf1 to AD are also not well explored. Calderwood and Murshid (2017) review Hsf1’s dysregulation in cancer and AD, focusing primarily on the regulatory phosphorylation and acetylation of Hsf1 by GSK3 and Sirt1 (Calderwood and Murshid, 2017). GSK3 phosphorylation which is thought to inhibit Hsf1’s transcription factor activity (Calderwood and Murshid, 2017; Xavier *et al*, 2000), is highly implicated in AD pathology, particularly for its role in hyperphosphorylating Tau, and it has been speculated GSK3 β may also inhibit Hsf1,

leading to altered heat shock protein response in AD (Calderwood and Murshid, 2017). Moreover, Evans, Wisen and Gestwicki (2006) showed HSP70 and HSP90 may be instrumental in inhibiting amyloid beta plaque formation in early self-assembly stages *in vitro* (Evans, Wisen and Gestwicki, 2006) but contrarily, inhibition of HSP90 prevents memory loss and synaptic loss in Tg2576 mice which overexpress Swedish mutant APP (Wang *et al*, 2017). While heat shock proteins like HSP70 and HSP90 are essential chaperones for proper protein misfolding and are therefore unsurprisingly, important regulators of misfolding proteins like amyloid beta and Tau, they are also important controllers of Hsf1 activity whereby, high cellular stoichiometry of HSP90 may cause an arrest-like state in Hsf1 because of the negative feedback loop governing heat shock proteins (Masser, Ciccarelli and Andreasson, 2020). Altogether, with our evidence that Hsf1 accelerates the processing of B2 and Alu RNAs and the literature evidence of dysregulation of the heat shock pathways and Hsf1 post translational modifications in AD pathology, Hsf1 may have a relatively understudied yet outsized role in AD pathology that deserves further attention. Furthermore, how the interplay of GSK3 β and heat shock proteins with the phenomenon of Alu and B2 processing interrelate is unstudied and deserves follow up. Indeed, in the APP 6mo mice, HSP70 family proteins Hsph1 and Hspa1b are upregulated but don't appear to be governed as B2 RNA SRGs (Cheng *et al*, 2020). Could heat shock protein accumulation interact with the Hsf1-Alu and B2 regulatory axis?

Hsf1's acceleration of B2 and Alu RNAs' processing is an important contributor to the literature of B2 and Alu RNAs' inhibition of Pol-II. Many works have been conducted to characterize how B2 and Alu RNAs inhibit Pol-II and even how their inhibition is reversed to elicit a stress response gene cascade (Zovoilis *et al*, 2016; Hernandez *et al*, 2020; Ponicsan, Kugel and Goodrich, 2015).

However, characterizing this in the context of pathology such as AD is new and presented here for the first time. Our evidence points to a varied cellular response to stress and AD pathology and suggests gene control by B2 and Alu RNAs are highly context dependent. In Zovoilis *et al* (2016), the authors employed B2 RNA CHART-Seq and a B2 RNA knockdown to uncover the genes which had close contact with B2 RNAs and which were upregulated post B2 RNA fragmentation. Of course, much of our analysis relies on assessing the genes in common with our model and these datasets, which show high similarity in response, but as the gene dynamics change through an aging model exposed to constant stimuli compared to a cell culture model, we see the model exhibiting dynamic gene expression signatures with specific and age dependent hallmarks which deserve much follow up.

Overall, the works in this thesis contribute evidence that ncRNAs are dysregulated in AD and as such, contribute to the underlying molecular dysfunction of the disease pathology. Future directions of this work include i) Exploring if other RNA: protein interactions regulate B2 and Alu RNA's control over RNA Pol II ii) Assess if B2 RNAs' processing and SRG activation is similarly regulated in other AD mouse models iii) Understand how B2 and Alu RNAs revert to a pre-stress state iv) Examine how A-to-I editing translated to lowered full length B2 RNA levels and v) annotate the differential ncRNA modifications in supplementary table 3 using biochemical and bioinformatic techniques.

6.2 References

Asher S, Priefer R (2022) Alzheimer's disease failed clinical trials. *Life Sci* 306: 120861

Bass BL, Weintraub H (1987) A developmentally regulated activity that unwinds RNA duplexes. *Cell* 48: 607-613

Bass BL, Weintraub H (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55: 1089-1098

Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82: 239-259

Castanho I, Murray TK, Hannon E, Jeffries A, Walker E, Laing E, Baulf H, Harvey J, Bradshaw L, Randall A et al (2020) Transcriptional Signatures of Tau and Amyloid Neuropathology. *Cell Rep* 30: 2040-2054 e2045

Cheng Y, Saville L, Gollen B, Isaac C, Belay A, Mehla J, Patel K, Thakor N, Mohajerani MH, Zovoilis A (2020) Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. *Elife* 9

Crooke PS, 3rd, Tossberg JT, Heinrich RM, Porter KP, Aune TM (2022) Reduced RNA adenosine-to-inosine editing in hippocampus vasculature associated with Alzheimer's disease. *Brain Commun* 4: fcac238

Crooke PS, 3rd, Tossberg JT, Porter KP, Aune TM (2021) Cutting Edge: Reduced Adenosine-to-Inosine Editing of Endogenous Alu RNAs in Severe COVID-19 Disease. *J Immunol* 206: 1691-1696

Dregni AJ, Duan P, Xu H, Changolkar L, El Mammeri N, Lee VM, Hong M (2022) Fluent molecular mixing of Tau isoforms in Alzheimer's disease neurofibrillary tangles. *Nat Commun* 13: 2967

Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* 11: 822-829

Evans CG, Wisen S, Gestwicki JE (2006) Heat shock proteins 70 and 90 inhibit early stages of amyloid beta-(1-42) aggregation in vitro. *J Biol Chem* 281: 33182-33191

Gal-Mark N, Shallev L, Sweetat S, Barak M, Billy Li J, Levanon EY, Eisenberg E, Behar O (2017) Abnormalities in A-to-I RNA editing patterns in CNS injuries correlate with dynamic changes in cell type composition. *Sci Rep* 7: 43421

Guo C, Jeong HH, Hsieh YC, Klein HU, Bennett DA, De Jager PL, Liu Z, Shulman JM (2018) Tau Activates Transposable Elements in Alzheimer's Disease. *Cell Rep* 23: 2874-2880

Hernandez AJ, Zovoilis A, Cifuentes-Rojas C, Han L, Bujisic B, Lee JT (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc Natl Acad Sci U S A* 117: 415-425

Lee VM, Goedert M, Trojanowski JQ (2001) Neurodegenerative tauopathies. *Annu Rev Neurosci* 24: 1121-1159

- Millan MJ (2017) Linking deregulation of non-coding RNA to the core pathophysiology of Alzheimer's disease: An integrative review. *Prog Neurobiol* 156: 1-68
- Ponicsan SL, Kugel JF, Goodrich JA (2015) Repression of RNA Polymerase II Transcription by B2 RNA Depends on a Specific Pattern of Structural Regions in the RNA. *Noncoding RNA* 1: 4-16
- Pratanwanich PN, Yao F, Chen Y, Koh CWQ, Wan YK, Hendra C, Poon P, Goh YT, Yap PML, Chooi JY *et al* (2021) Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol*
- Przybyla M, van Eersel J, van Hummel A, van der Hoven J, Sabale M, Harasta A, Muller J, Gajwani M, Prikas E, Mueller T *et al* (2020) Onset of hippocampal network aberration and memory deficits in P301S tau mice are associated with an early gene signature. *Brain* 143: 1889-1904
- Ramirez P, Zuniga G, Sun W, Beckmann A, Ochoa E, DeVos SL, Hyman B, Chiu G, Roy ER, Cao W *et al* (2022) Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system. *Prog Neurobiol* 208: 102181
- Ramos H, Monette A, Niu M, Barrera A, Lopez-Ulloa B, Fuentes Y, Guizar P, Pino K, DesGroseillers L, Mouland AJ *et al* (2022) The double-stranded RNA-binding protein, Staufen1, is an IRES-transacting factor regulating HIV-1 cap-independent translation initiation. *Nucleic Acids Res* 50: 411-429
- Vlachogiannis NI, Gatsiou A, Silvestris DA, Stamatelopoulos K, Tektonidou MG, Gallo A, Sfrikakis PP, Stellos K (2020) Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. *J Autoimmun* 106: 102329
- Wang B, Liu Y, Huang L, Chen J, Li JJ, Wang R, Kim E, Chen Y, Justicia C, Sakata K *et al* (2017) A CNS-permeable Hsp90 inhibitor rescues synaptic dysfunction and memory loss in APP-overexpressing Alzheimer's mouse model via an HSF1-mediated mechanism. *Mol Psychiatry* 22: 990-1001
- Xavier IJ, Mercier PA, McLoughlin CM, Ali A, Woodgett JR, Ovsenek N (2000) Glycogen synthase kinase 3beta negatively regulates both DNA-binding and transcriptional activities of heat shock factor 1. *J Biol Chem* 275: 29147-29152
- Yakovchuk P, Goodrich JA, Kugel JF (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci U S A* 106: 5569-5574
- Zhang X, Trebak F, Souza LAC, Shi J, Zhou T, Kehoe PG, Chen Q, Feng Earley Y (2020) Small RNA modifications in Alzheimer's disease. *Neurobiol Dis* 145: 105058
- Zovoilis A, Cifuentes-Rojas C, Chu HP, Hernandez AJ, Lee JT (2016) Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* 167: 1788-1802 e1713

Appendix 1 – Supplementary for chapter 2: Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology

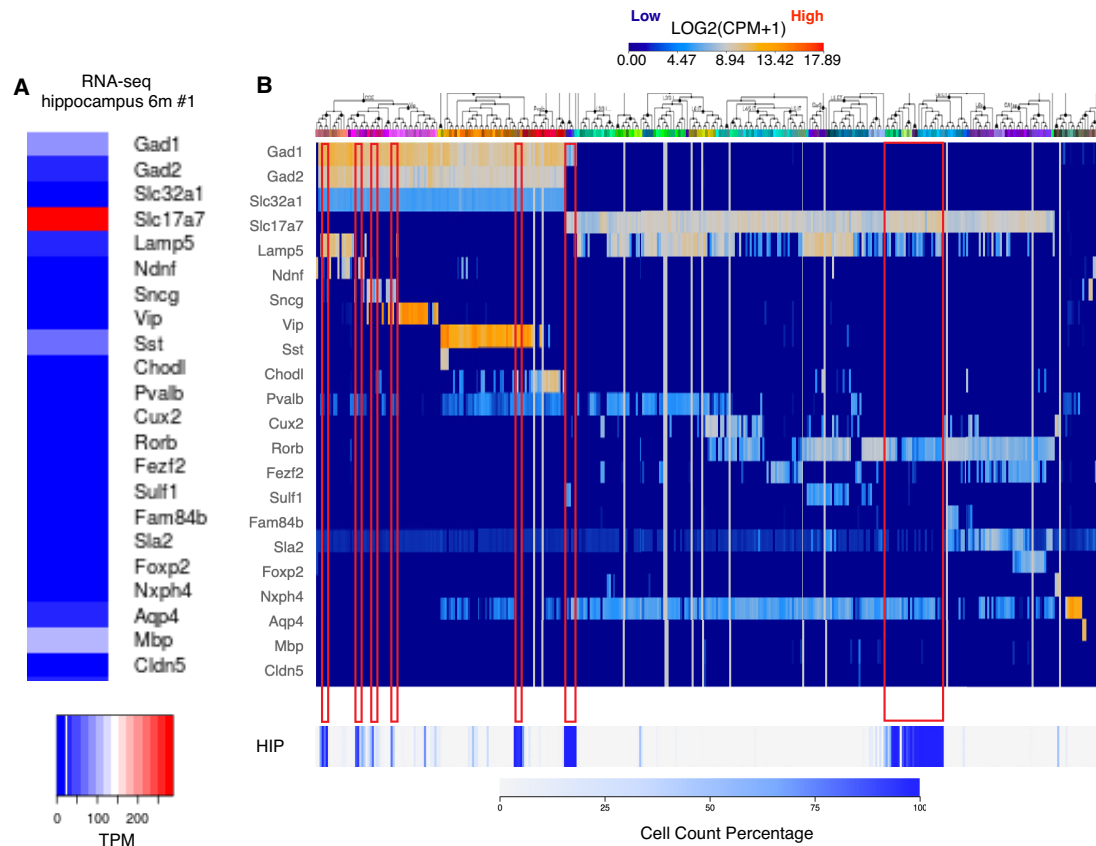


Figure 2—figure supplement 1. Expression of known hippocampal markers in our RNA-seq data of mice hippocampi.

(A–B) Expression levels comparison between RNA-seq data from WT 6-month-old mouse hippocampus used in our study (panel A) and RNA-seq data from the Allen Brain Atlas (Panel B). Comparison is done for the standard panel genes of reference genes used by Allen Brain Atlas. Every column in panel B represents a different cell type in brain with cell types belonging to hippocampus corresponding to increased density (cell count percentage) at the lower part of panel B labeled as HIP. Red rectangles mark the most prominent cell types that are enriched in Hippocampus. © 2015, Allen Institute for Brain Science. Image credit: Panel B was taken from the Allen Mouse Brain Atlas, available at: https://portal.brain-map.org/atlas-and-data/rnaseq#Mouse_Cortex_and_Hip. It is not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

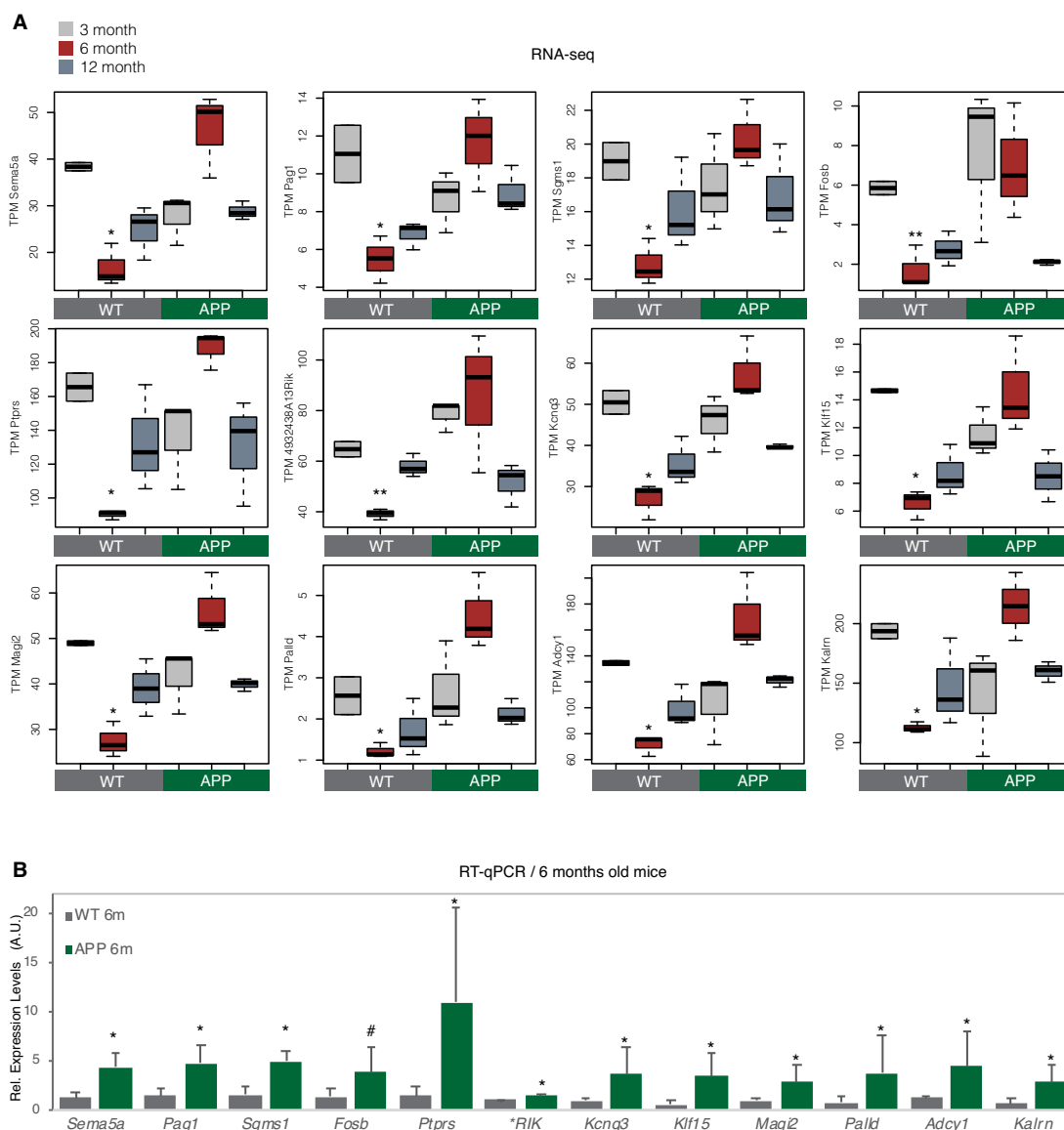


Figure 2—figure supplement 2. Validation of RNA-seq data in WT and APP mice by RT-qPCR.

(A) Expression levels calculated through RNA-seq for the selected genes to be tested through RT-qPCR in (B). These are 12 from the genes that were found to be upregulated in APP 6-month-old mice compared to WT mice of the same age (Supplementary file 1). The selected subset includes both B2-SRGs (first 10 boxplots) that are part of the 72 B2-SRGs upregulated in APP mice (Supplementary file 1) but also two non-B2-SRGs (last two plots; *Adcy1* and *Kalrn*) that are used later as negative controls (see Figure 7C). Values are based on TPM counts of long-RNA-seq data. One asterisk marks statistical significance ($p < 0.05$) for the comparison between 6-month-old APP and 6-month-old WT ($n = 3/\text{group}$, unpaired non-directional t-test), while two asterisks marks statistical significance only for 6-month-old APP greater than 6-month-old WT ($n = 3/\text{group}$,

unpaired directional t-test). All the other comparisons between groups and different ages are non-significant. (B) Confirmation through RT-qPCR of the expression differences between 6-month-old WT and APP mice. Statistical significance (p value threshold 0.05), is depicted as asterisk, for expression in 6-month-old APP mice greater than the expression in 6-month-old WT mice (unpaired directional t-test, n numbers as in (A), error bars represent standard deviation from the mean). # represents a p=0.08 for the same comparison and test. * RIK stands for 4932438A13Rik.

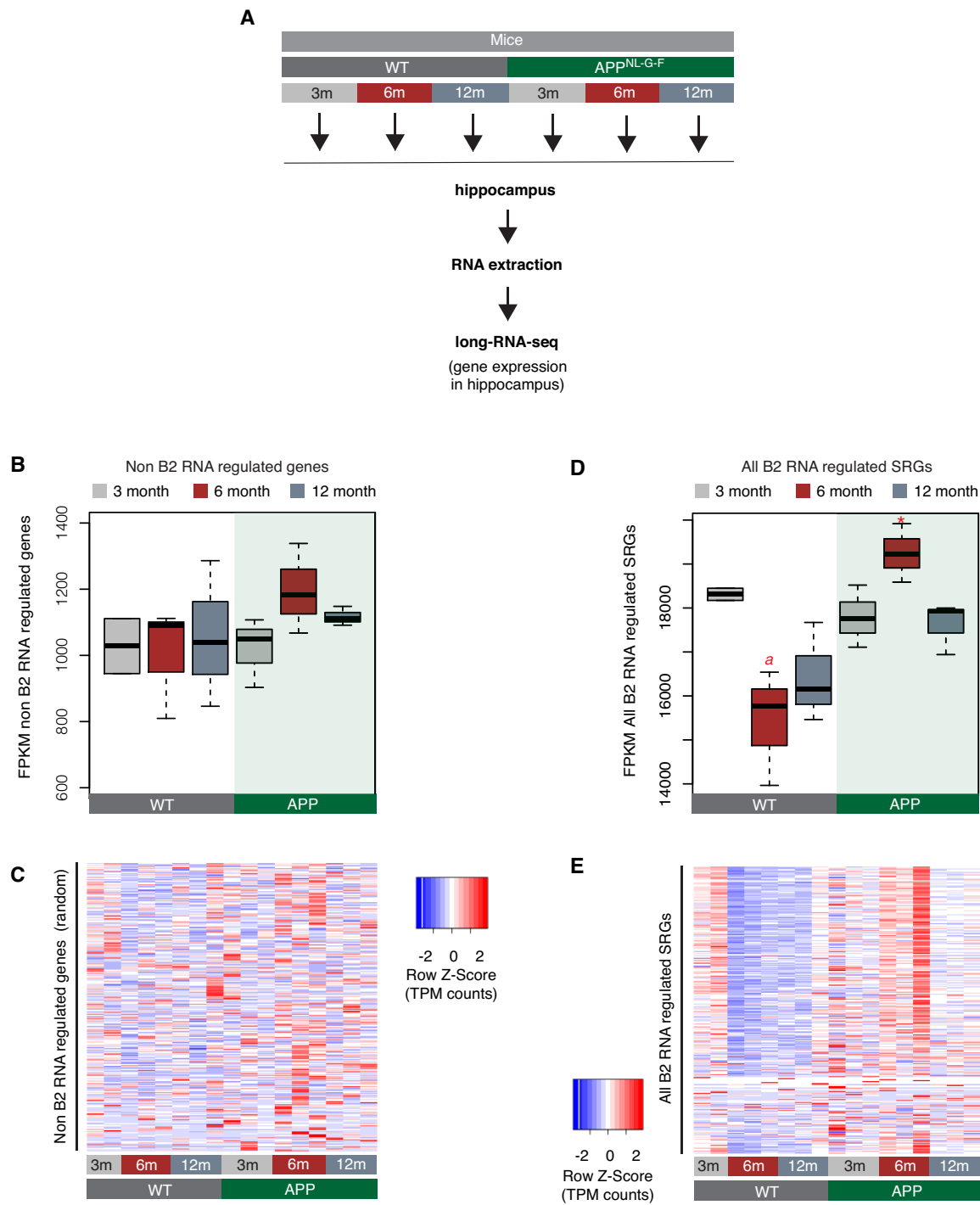


Figure 2—figure supplement 3. Expression levels of all B2-SRGs in amyloid pathology.

(A) Experimental design for study of B2-SRGs in the hippocampus of the amyloid pathology mouse model (APP) and the respective wild type (WT) control. (B) Expression levels of non-B2 RNA regulated genes as defined by long-RNA-seq. Boxplot depicts distribution of expression levels (in FPKM/Fragments per Kilobase per Million) among different age groups of wild type and APP mice. Not significant difference between 6-month-old APP and 6-month-old WT. (C) Gene expression levels (defined by long-RNA-seq) of a random set of genes not associated with B2 RNAs show weak or no association with amyloid pathology status in the hippocampus of APP and WT mice of different ages. Heatmap depicts gene expression with rows representing 1385 non-B2 RNA regulated genes (Supplementary file 1) and columns representing the different mouse samples. Expression values are normalized per row and correspond to TPM values. (D) Expression levels of B2-SRGs (1684 genes, Supplementary file 1) as defined by long-RNA-seq. Boxplot depicts distribution of expression levels (in FPKM/Fragments per Kilobase per Million) of B2 RNA regulated genes among different age groups of wild-type and APP mice. Statistical significance (p value threshold 0.05) for the comparison between 6-month-old APP and 6-month-old WT ($p=0.02$)(depicted as asterisk, unpaired, non- directional t-test, $n = 3/\text{group}$) and the comparison between 3-month-old WT and 6-month-old WT ($p=0.05$)(depicted as a). No significance for the rest of the comparisons between APP and WT in the other two age groups, or between other different ages within the same group (APP or WT). (E) Gene expression levels of B2-SRGs (1684 genes, Supplementary file 1) show strong association with amyloid pathology status in the hippocampus of APP and WT mice of different ages. Heatmap depicts gene expression with rows representing B2 RNA regulated genes and columns representing mouse samples. Expression values are normalized per row and correspond to TPM values. Red color represents higher expression. (Note) When expression dynamics for all B2-SRGs are tested, these are similar to the ones observed for genes in Figure 2, while this pattern in B2 RNA regulated SRG expression was not observed when a random set of genes is tested. Thus, it cannot be excluded that the expression pattern observed in genes of Figure 2 may extend to additional B2-SRGs described in our previous work that have been filtered out by DEseq. This could be attributed to our current sequencing depth and the large dynamic range of some of them as shown in panel E, but also due to genome wide mRNA expression level changes in our system that are not taken into account by DEseq's conservative hypothesis that expression levels of most genes remain unchanged. This is relevant in our case given the ability of full-length B2 RNAs (which are elevated in our system) to suppress global gene expression and thus, it may have resulted in a number of genes being filtered out during DESeq correction for multiple testing.

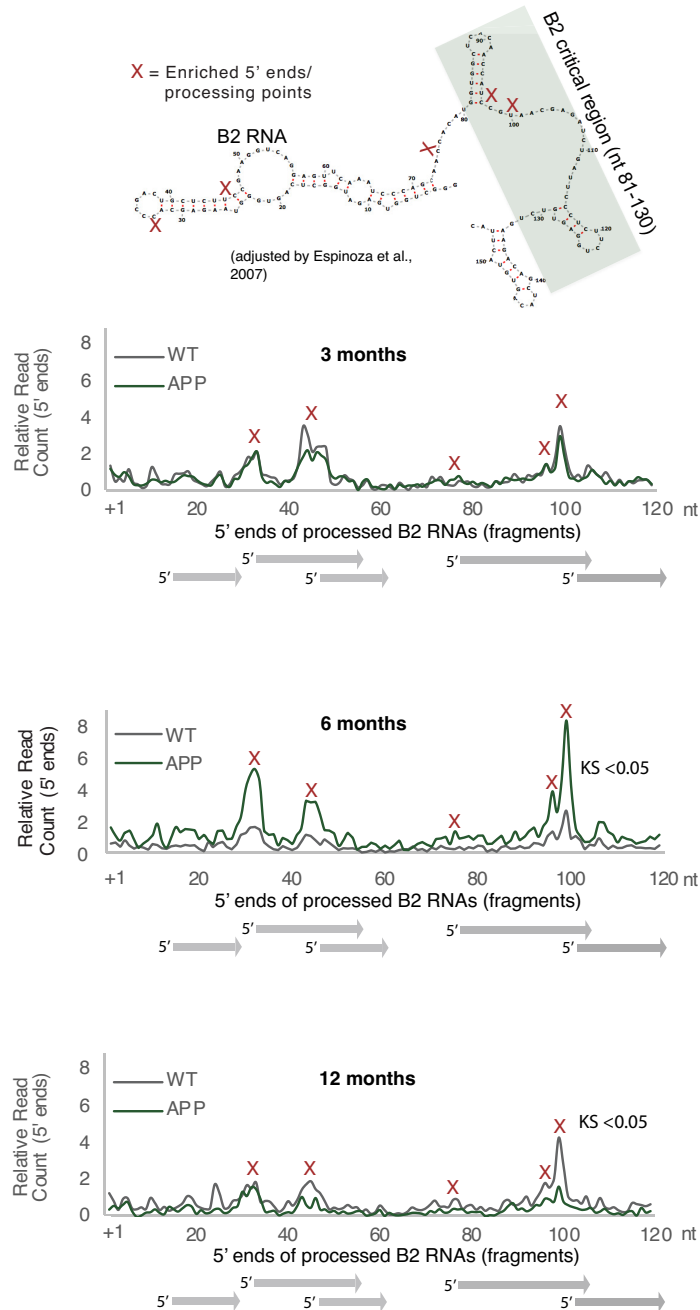


Figure 3—figure supplement 1.

Plotting of the position of the first base (5' end) of B2 RNA fragments across the B2 loci to compare levels of B2 RNA fragments between APP and WT mice in the three different age groups. This figure relates to Figure 3A. Upper panel: As in Figure 3A. Secondary structure and processing points of B2 RNA. Secondary structure of B2 RNA adapted from Espinoza and colleagues (Espinoza et al., 2007). Lower panels: Distribution of the 5' ends of B2 RNA fragments across the B2 RNA loci based on mapped short RNA-seq from the hippocampi of 3-month, 6-month, and 12-month-old APP and WT mice. The x axis represents a metagene combining all B2 RNA loci

aligned at the start site of their consensus sequence (position +1) and the y axis shows the relative 5' end count for B2 RNA fragments aligning to any position downstream of position +1. Figure depicts the difference between B2 RNA fragment distribution of 6-month-old APP mice (higher peaks) vs 6-month-old WT mice (lower peaks) (Kolmogorov Smirnov test, $KS < 0.05$) and the reverse difference of 6-month-old APP mice (higher peaks) vs 6-month-old WT mice (lower peaks) (Kolmogorov Smirnov test, $KS < 0.05$).

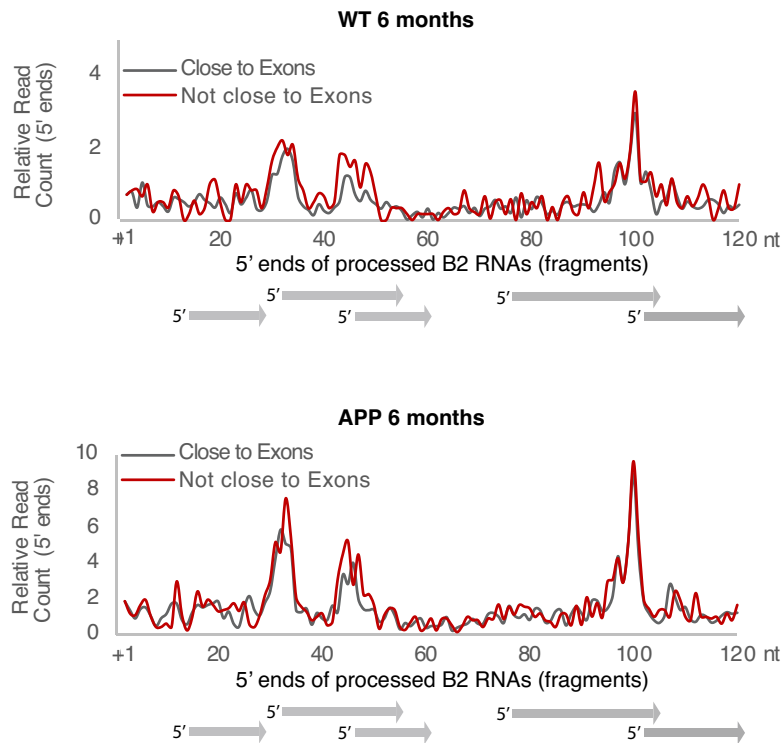


Figure 3—figure supplement 2.

Plotting of the position of the first base (5' end) of B2 RNA fragments across the B2 loci to compare levels of B2 RNA fragments in 6-month-old mice between B2 elements that overlap exonic/ genic regions and those that do not.

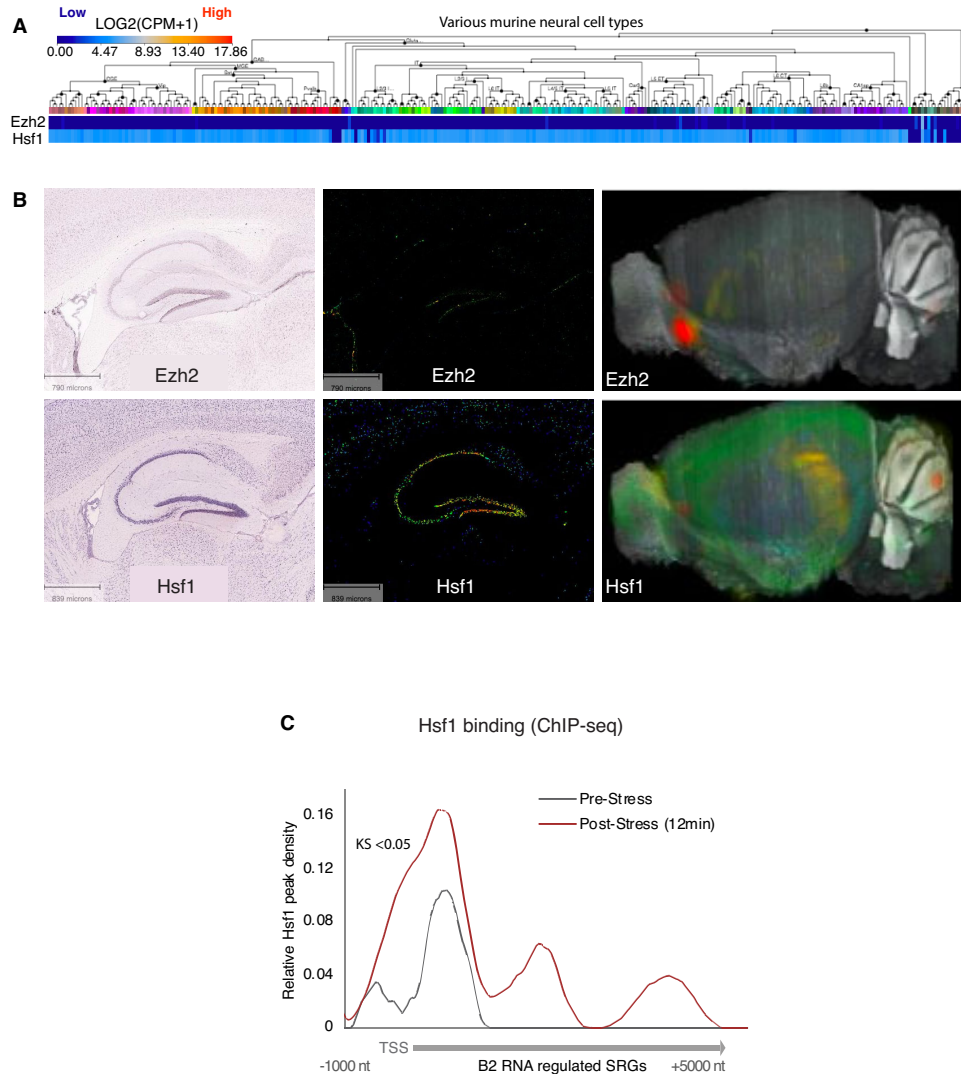


Figure 4—figure supplement 1. Expression of Hsf1 in neural tissues.

(A) Mouse cortex and hippocampus gene expression levels for Ezh2 and Hsf1 depicted in the Allen Brain Atlas Transcriptomics explorer showing limited expression of Ezh2 across multiple neural tissues compared to Hsf1. Levels are per cell type based on RNA-seq data. (B) In situ hybridization brain image data with cellular-level resolution for Ezh2 (upper panel) and Hsf1 (lower panel) depicting low expression levels for Ezh2. Left panel, ISH images, middle panel, gene expression images, right panel, expression in 3D reconstructed whole mouse brain. (C) Metagene analysis of increase in Hsf1 binding (ChIP-seq data) across the start of B2 RNA regulated genes between pre-stress and 12 min post-stress (KS <0.05). ChIP-seq data are from Mahat et al., 2016. TSS represents the Transcription Start Site of these genes. This figure provides the rational why we started investigating Hsf1 as a candidate with a potential relation to B2 Biology and a potential effect on B2 RNA processing. © 2015, Allen Institute for Brain Science. Image credit: The image

in panel A was taken from the Allen Mouse Brain Atlas, available at: https://portal.brain-map.org/atlas-and-data/rnaseq#Mouse_Cortex_and_Hip. The images for panel B were taken from the Allen Mouse Brain Atlas, available at <https://mouse.brain-map.org/experiment/show/100142521> for Eh2 and <https://mouse.brain-map.org/experiment/show/68196972> for Hsf1. These panels are not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

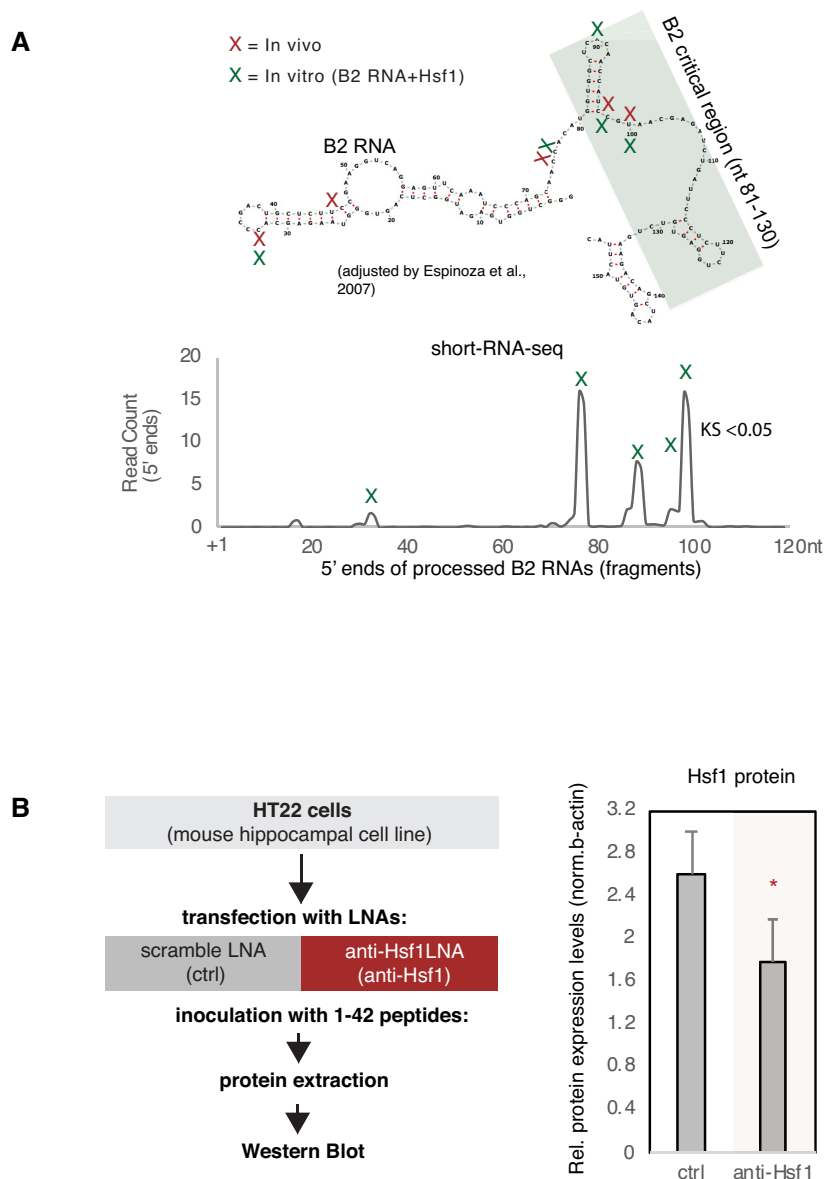


Figure 5—figure supplement 1. Position of B2 RNA fragments generated in vitro and downregulation of Hsf1 protein levels. (A) Plotting of the position of the first base (50 end) of B2 RNA fragments across the B2 loci produced by B2 RNA that has been processed in vitro in the presence of Hsf1. Mapping was done as in case of the in vivo samples (Figure 3) except the

normalization step. (B) Estimation by western blotting of Hsf1 protein levels after treatment of HT22 cells with the anti-Hsf1 LNA. Left panel represents the experimental design. Asterisk depicts statistical significance $p < 0.05$ for expression in anti-Hsf1 LNA-treated samples being less than that of samples treated with the scramble LNA (ctrl) (unpaired, directional t-test, $n = 3/\text{group}$).

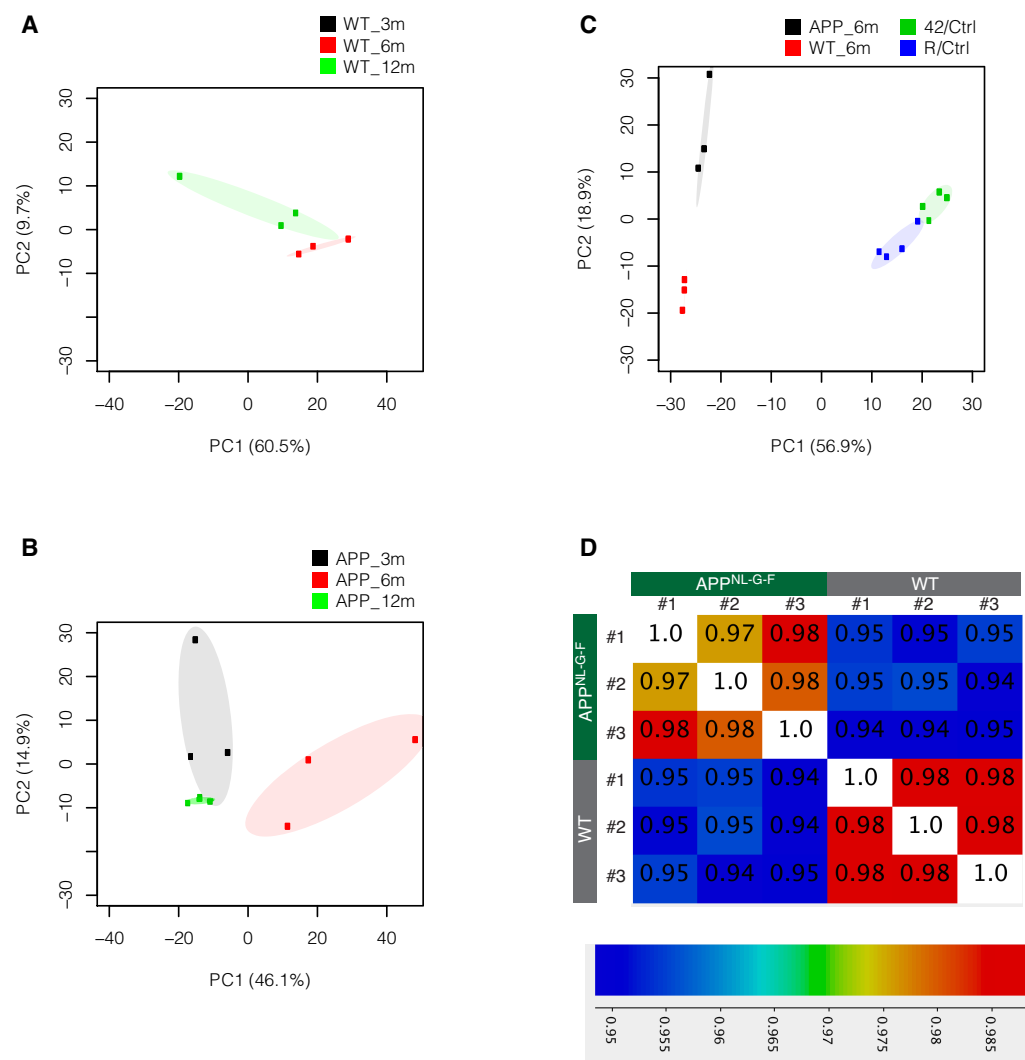


Figure 6—figure supplement 1. PCA plots and correlation matrix for sequenced samples in amyloid pathology and amyloid toxicity models. Plots are based on amyloid pathology genes (Supplementary file 1). Correlation matrix was constructed using Semonk for the reads per million per gene length counts.

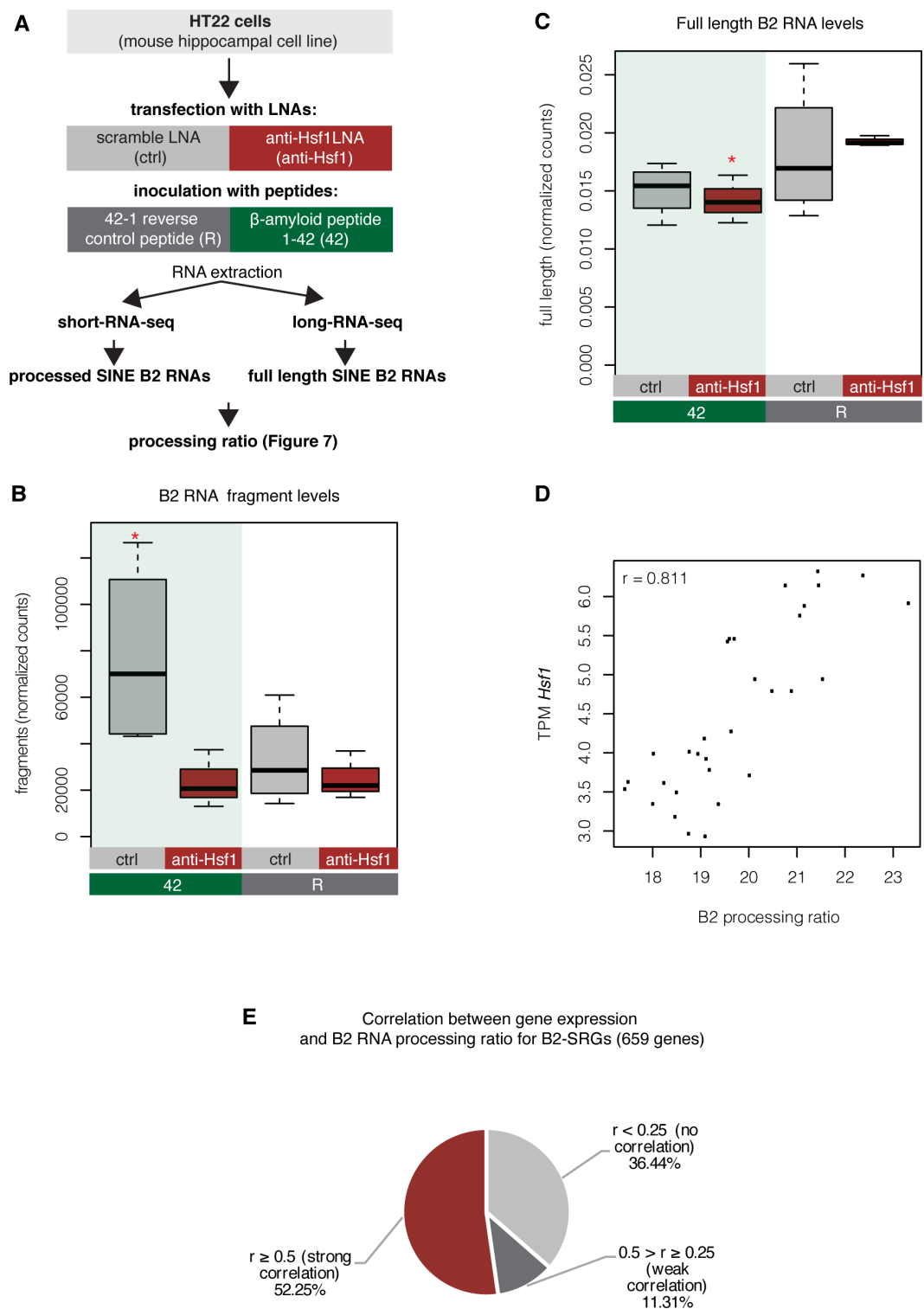


Figure 8—figure supplement 1. B2 RNA levels in HT22 cells and relationship with Hsf1 levels.

(A) Experimental design for estimation of B2 RNA processing ratio based on short and long-RNA seq data in our HT22 amyloid toxicity model. (B) Boxplot depicts distribution of levels of

processed SINE B2 RNA fragments among different conditions. Statistical significance (p value threshold 0.05) for 42/ctrl greater than R/ctrl ($p=0.03$)($n = 4/\text{group}$, unpaired directional t-test, depicted as asterisk). No significant difference was observed between 42/anti-Hsf1 and R/anti-Hsf1 (NS, $n = 3/\text{group}$) and between R/anti-Hsf1 and R/ctrl. (C) Boxplot depicts distribution of levels of full-length SINE B2 RNAs among different conditions. Statistical significance (p value threshold 0.05) for the comparison between 42-anti-Hsf1 and R-anti-Hsf1 ($p=0.04$)(depicted as asterisk, $n = 3/\text{group}$, unpaired non-directional t- test). No significant difference was observed for the other comparisons. (D) Correlation plot for the relationship between Hsf1 levels and processing ratio in all samples sequenced in the current study (Pearson correlation, r : correlation co-efficient, $p<0.001$). (E) Correlation between B2-SRGs expression and B2 RNA processing ratio calculated based on RNA-seq data and short-RNA-seq data from hippocampal samples in the current study. Only B2-SRGs, independently from amyloid beta status, that had read coverage across all samples were considered. For every gene a correlation coefficient was calculated (Pearson correlation) together with the respective p-value. Genes with a correlation p value less than 0.05 were considered in the subsequent analysis (659 genes) that classified them into three categories based on r : (i) no correlation ($r < 0.25$), (ii) weak correlation ($r = 0.25$ or $r < 0.05$), and (iii) strong correlation ($r = 0.5$ or $r > 0.05$). The pie diagram represents the percentage of each category. The exact r and p values of these genes are listed in Supplementary file 1.

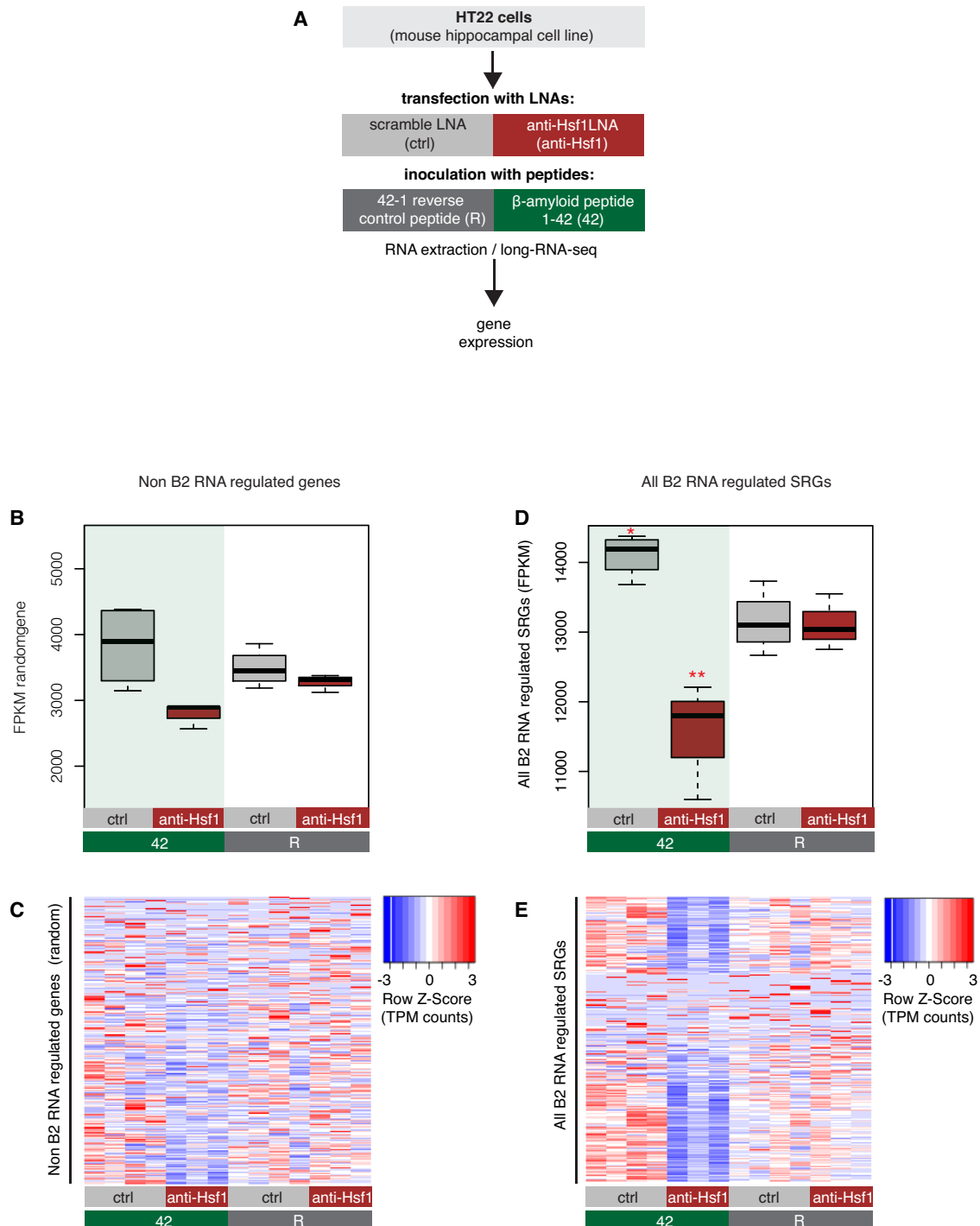
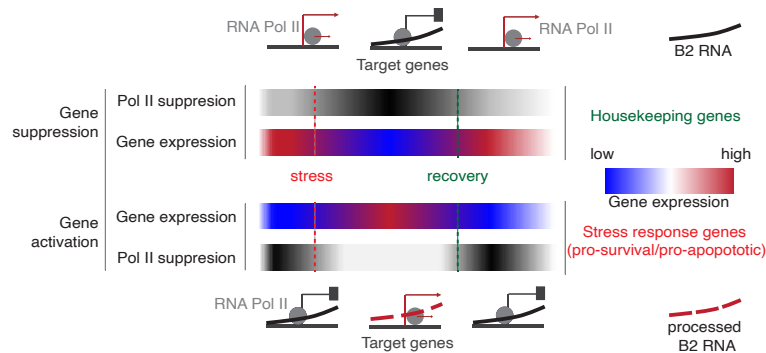


Figure 8—figure supplement 2. Expression levels of all B2-SRGs in amyloid beta toxicity.
 (A) Experimental design of the combined Hsf1 Knock Down – amyloid toxicity assay in HT22 cells followed by short- and long-RNA-seq. (B) Expression levels of non-B2 RNA regulated genes as defined by long-RNA-seq. Boxplot depicts distribution of expression levels (in

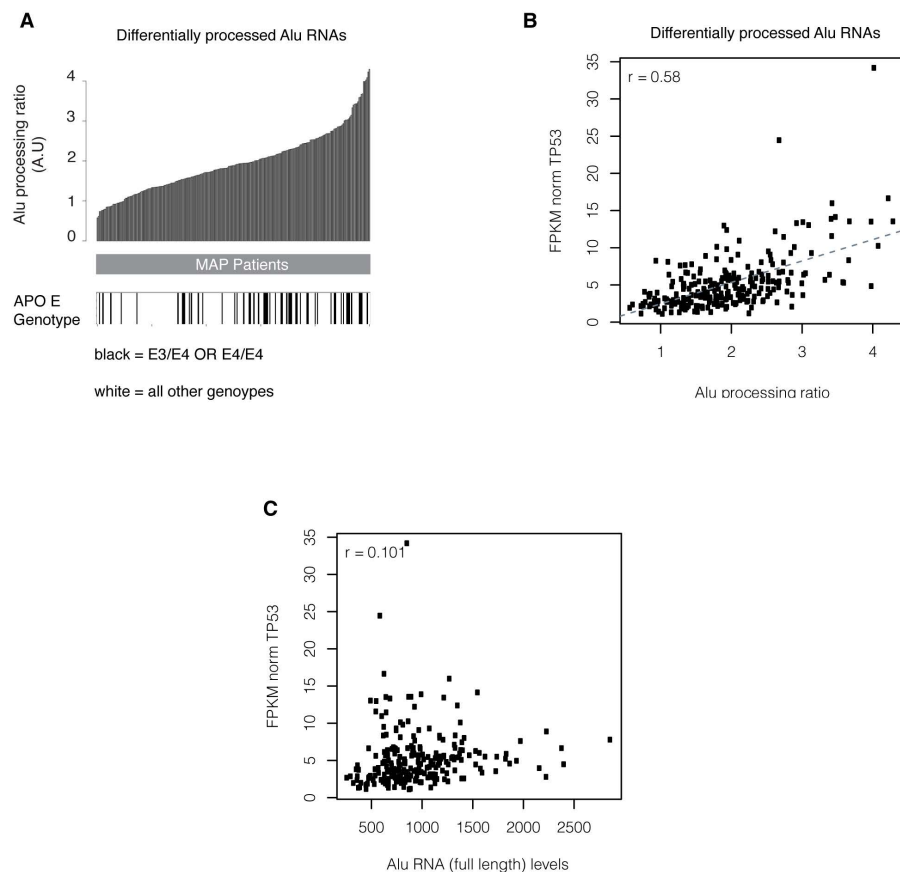
FPKM[Fragments per Kilobase per Million]) among different groups of HT22 cells. (C) Gene expression levels (defined by long-RNA-seq) of a random set of genes not associated with B2 RNAs (Supplementary file 1) show weak or no association with Hsf1 treatment during response to amyloid toxicity in HT22 cells. Heatmap depicts gene expression with rows as random genes (non- B2 RNA regulated) and columns corresponding to the different HT22 cell treatments. Expression values are normalized per row and correspond to TPM values, with red color represents higher expression. (D) B2 RNA regulated SRG expression levels based on long-RNA-seq data (FPKM values). Boxplot depicts distribution of expression levels of B2 RNA regulated genes among different groups of HT22 cells between 42 and R (1684 genes, Supplementary file 1). Statistical significance (p value threshold 0.05) for the comparison between 42/ctrl and R/ctrl ($p=0.01$)($n = 4$ /group, unpaired directional t-test, depicted as one asterisk) and for 42/anti-Hsf1 less than R/anti-Hsf1 ($p=0.03$)($n = 3$ /group, depicted as two asterisks, unpaired directional t-test). (E) Gene expression levels (long-RNA-seq) of B2-SRGs (Supplementary file 1) show strong association with Hsf1 treatment during response to amyloid toxicity in HT22 cells. Heatmap depicts gene expression with rows as B2 RNA regulated genes (1684 genes, Supplementary file 1) and columns as different HT22 cell treatments. TPM values are normalized per row. Red color represents higher expression.

Appendix 2 – Supplementary for chapter 3: Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes



Appendix Figure S1.

Graphical representation of the mode of action of SINE RNAs and SINE RNA processing in mouse (B2 RNAs) with regard to transcriptome changes based on previous studies from the Kugel/Goodrich labs (Espinoza et al, 2007; Ponicsan et al, 2010, 2015; Yakovchuk et al, 2009) and our studies (Cheng et al, 2020; Hernandez et al, 2020; Zovoilis et al, 2016).



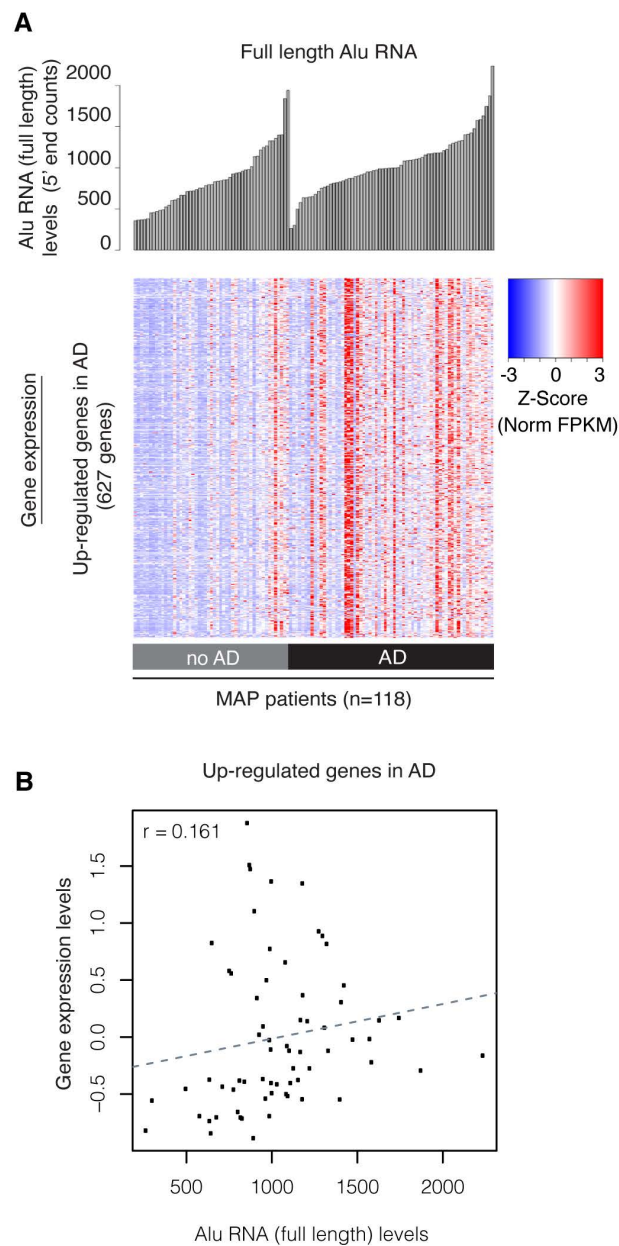
Appendix Figure S2.

(A) Association between SINE Alu RNA ratio (upper panel) and APO E genotype (lower panel). Every column in both panels corresponds to the same MAP patient (n=241). Patients are sorted from left to right in an ascending order with regard to Alu RNA processing ratio. Patients with at least one E4 allele that confers higher genetic risk to developing late onset AD are marked with black in the lower panel. The figure shows that as processing ratio climbs higher, it is more likely to identify an individual with this phenotype.

Here it should be noted though that the connection between APO E and AD, despite APO E being one of the few reliable genetic markers shown to confer some degree of higher susceptibility to AD, is not a one to one absolute match (i.e. the vast majority of people with these genotypes will not develop AD). For this reason, we also do not expect the correlation between Alu RNA processing and this genotype to be a one to one absolute match, especially since Alu RNA processing is a continuous variable while Apo E genotype is a discrete one. The reason why within the higher processing ratio, distribution of genotypes seems to be biphasic (with one density area in the middle of the range, and one towards the right end of the range) remains unknown and likely should be attributed to the large heterogeneity among AD patients and the multi-factorial nature of this disease.

(B) Scatterplot depicting the positive correlation between TP53 expression values and Alu RNA processing ratio in all MAP patients ($r=0.58$, $p<0.001$).

(C) Scatterplot depicting lack of correlation between TP53 mRNA expression values and full length Alu RNA levels in all MAP patients (n=241).

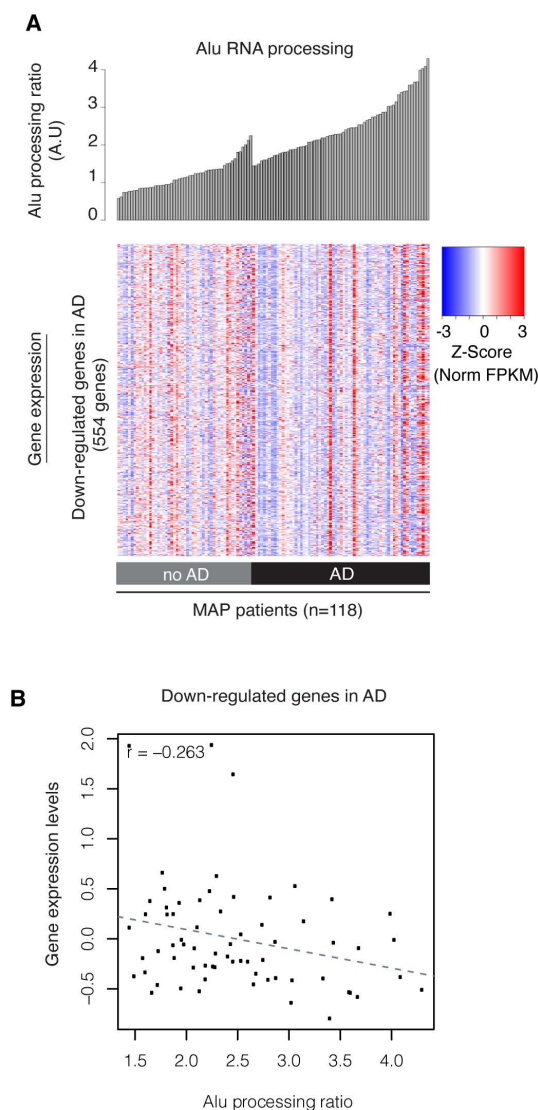


Appendix Figure S3.

(A) No association between full length Alu RNA levels (upper panel) and gene expression levels (lower panel) of genes up-regulated in AD (Dataset EV3, $|\log_2\text{FoldChange}| > 0.5$) is found in contrast to our observations with regard to Alu RNA processing ratio. Columns and rows as described in Figure 6 with every column in both panels corresponding to the same MAP patient

sorted from left to right in an ascending order with regard to Alu RNA levels for either the no AD or AD group.

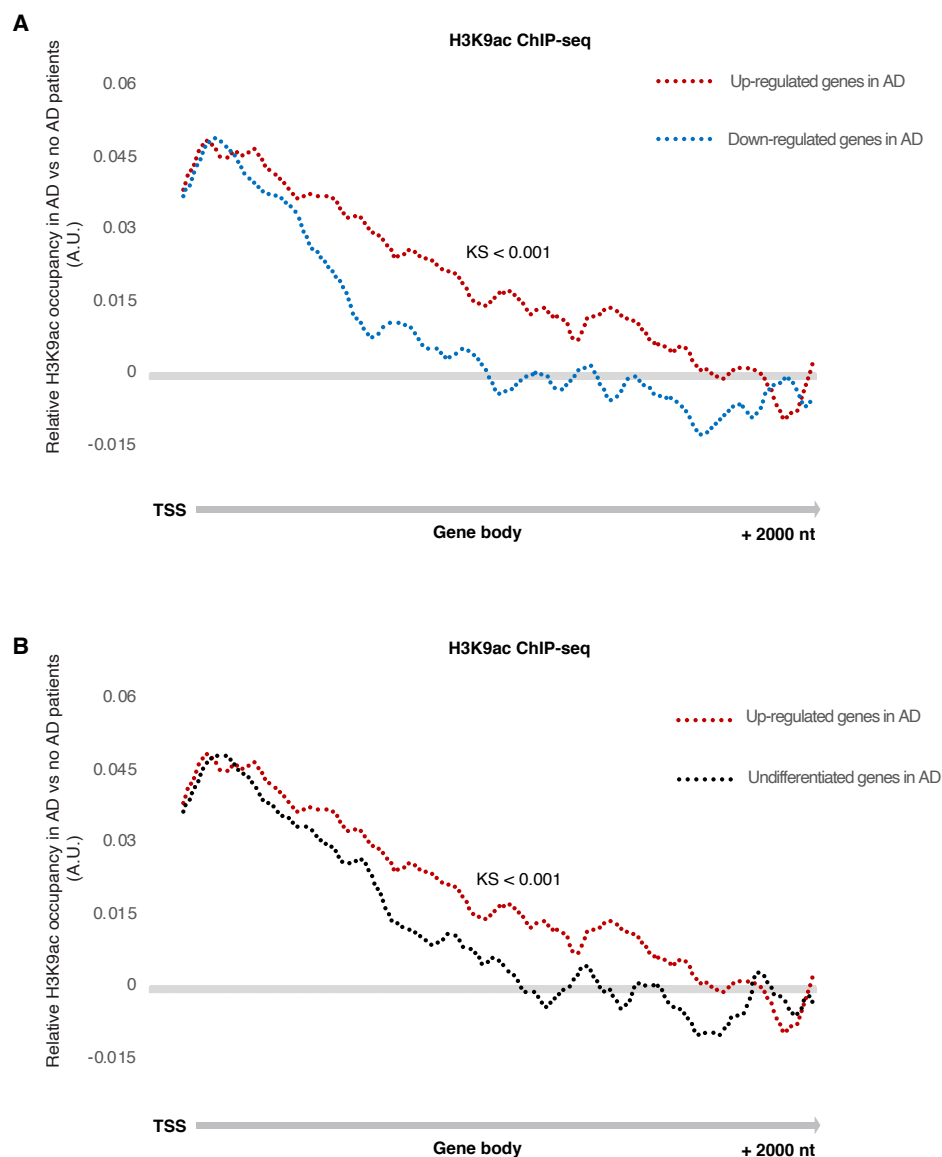
(B) Scatterplot depicting lack of correlation between average gene expression values of upregulated genes in panel A and full length Alu RNA levels in AD patients ($r=0.16$).



Appendix Figure S4.

(A) Only a weak association between Alu RNA processing ratio (upper panel) and gene expression levels (lower panel) of genes down-regulated in AD (Dataset EV4, $|\log_2\text{FoldChange}| > 0.5$) is found in contrast to our observations with regard to full length Alu RNA levels. Columns and rows as described in Figure 6 with every column in both panels corresponding to the same MAP patient sorted from left to right in an ascending order with regard to Alu RNA processing ratio for either the no AD or AD group.

(B) Scatterplot depicting the weak negative correlation between average gene expression values of down-regulated genes in panel A and Alu RNA processing ratio in AD patients ($r=-0.26$).

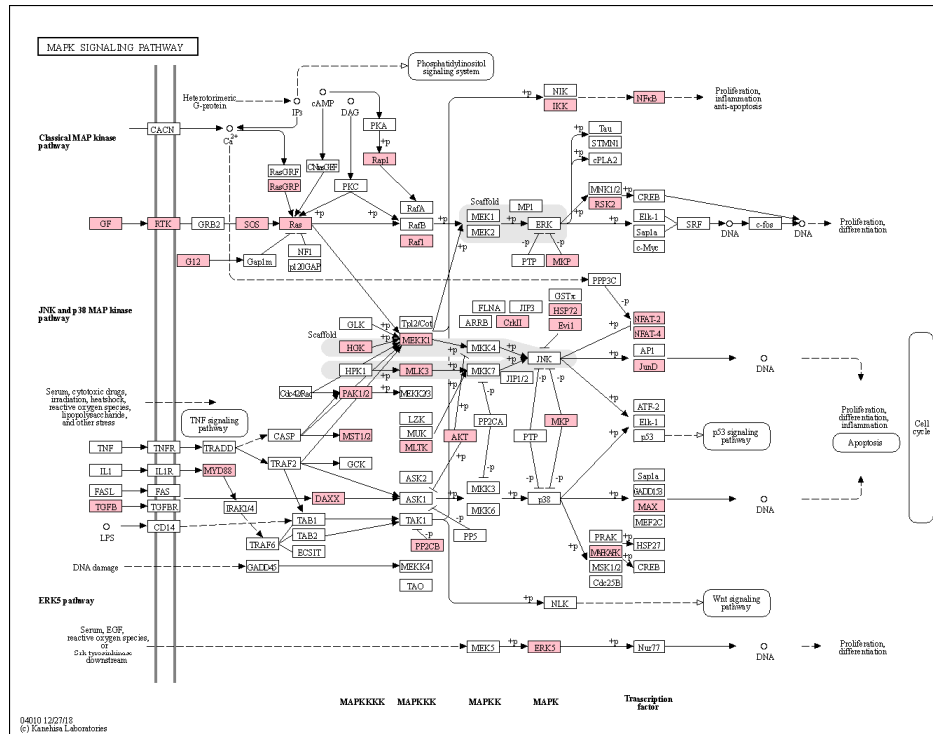


Appendix Figure S5. H3K9ac ChIP-seq for MAP patients.

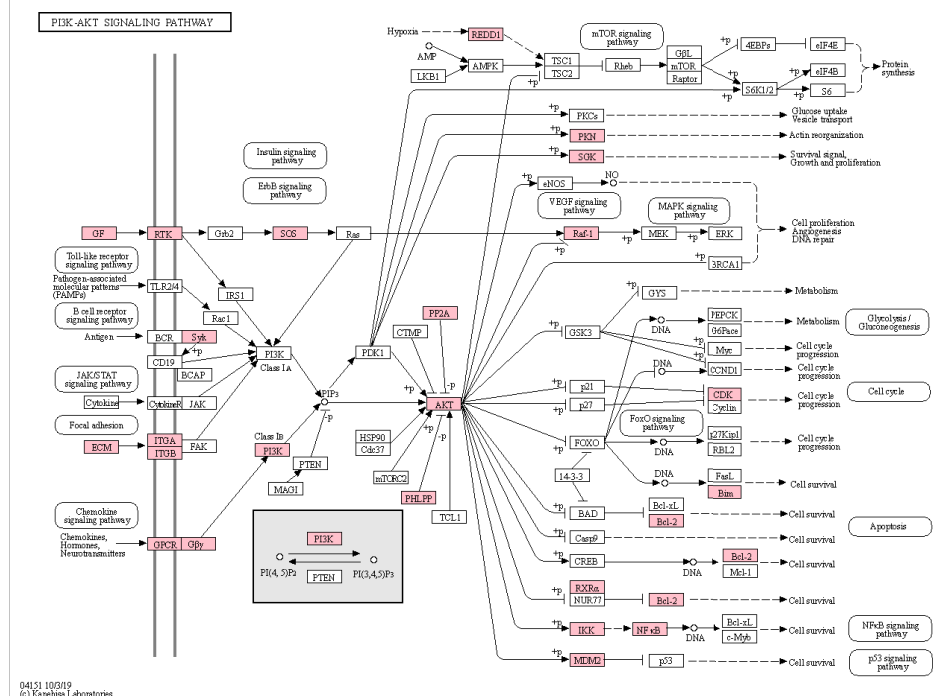
(A) Comparison of H3K9ac occupancy downstream of Transcription Start Site (TSS) between AD and no AD patients for genes up-regulated and down-regulated in AD (Dataset EV3 and Dataset EV4).

(B) Comparison of H3K9ac occupancy downstream of Transcription Start Site (TSS) between AD and no AD patients for genes up-regulated and a random set of non-differentiating genes (Dataset EV3 and Dataset EV5).

A



B



Appendix Figure S6. Pathways impaired in AD up-regulated genes that are strongly correlated with Alu RNA processing ratio.

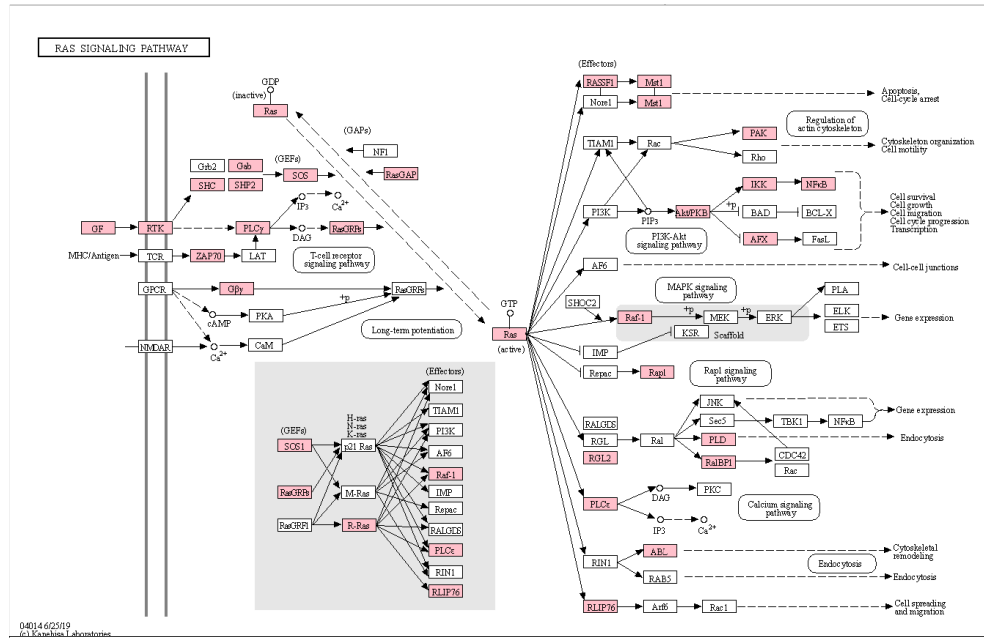
Pathways identified by KEGG pathway term enrichment analysis of AD upregulated genes the expression of which is strongly correlated with Alu RNA processing ratio ($r > 0.5$, p value < 0.05 Pearson correlation, Dataset EV8). Selected pathways are among those with a $p_{adj} < 0.05$ in the KEGG enrichment term analysis by the DAVID platform (see methods). Pathways have been

selected based on their association with functional terms such as cellular stress, cell survival, cell death and proliferation.

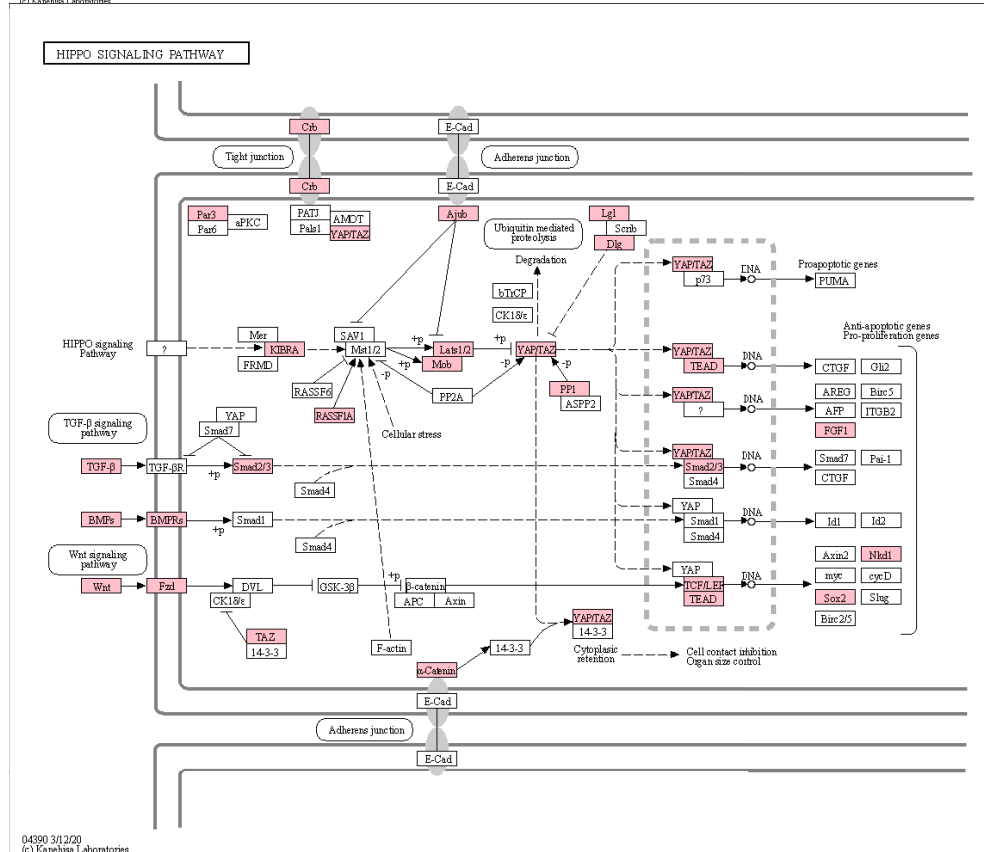
(A) MAPK signalling pathway. Multiple members (colored with pink) belong to the set of AD upregulated genes correlated with Alu RNA processing ratio. This pathway is upstream of the TP53 signalling pathway and cell death and proliferation.

(B) PI3K-AKT pathway. This pathway intersects with various signalling pathways regulating TP53 signalling and cell cycle. Pink color as in A.

A



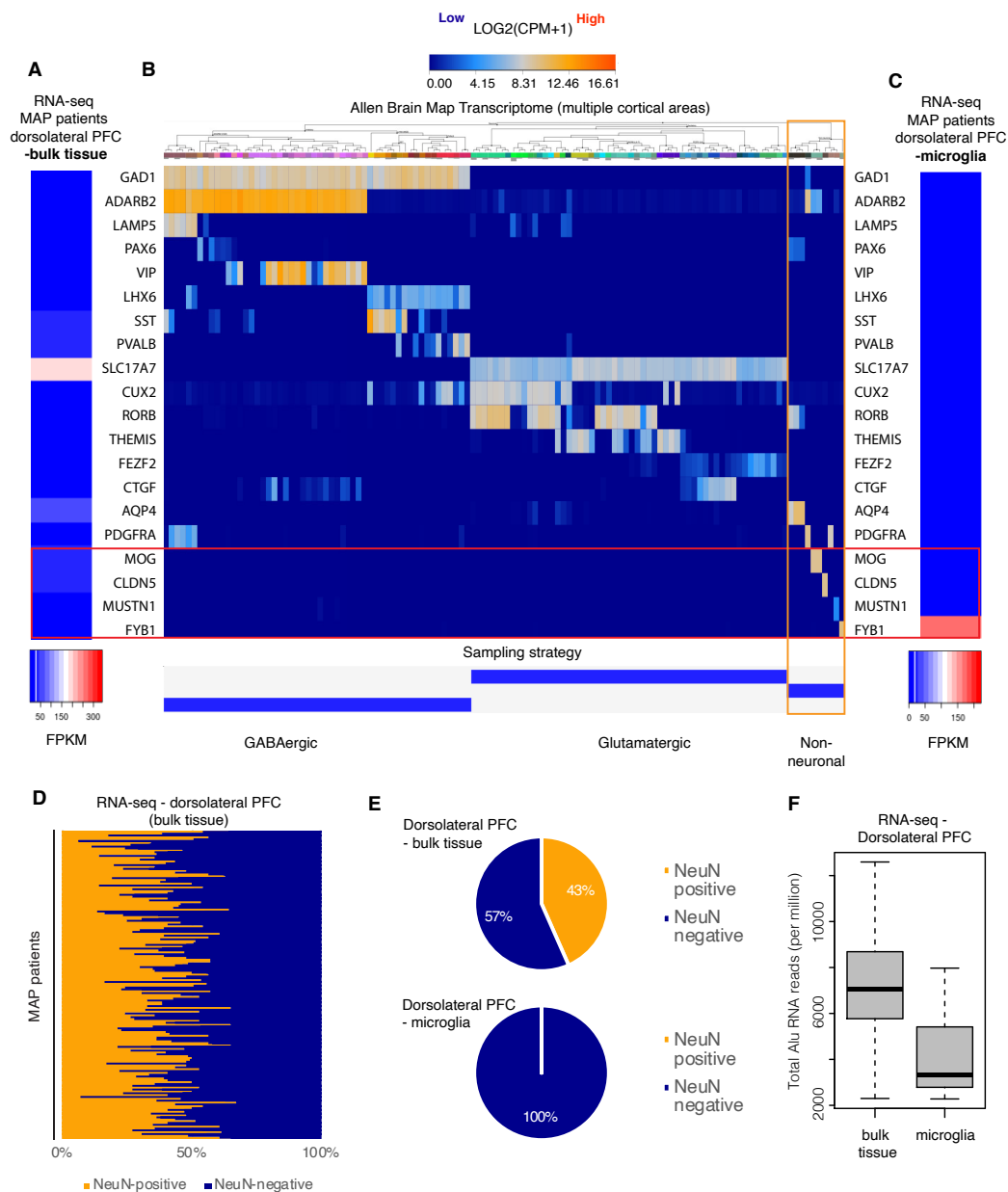
B



Appendix Figure S7. Additional pathways impaired in AD up-regulated genes that are strongly correlated with Alu RNA processing ratio.

(A) RAS signalling pathway. Multiple members (colored with pink) belong to the set of AD upregulated genes correlated with Alu RNA processing ratio. This pathway is important for cell survival and also upstream TP53.

(B) HIPPO signalling pathway. This pathway intersects with various signalling pathways regulating cellular response to stress. Pink color as in A.

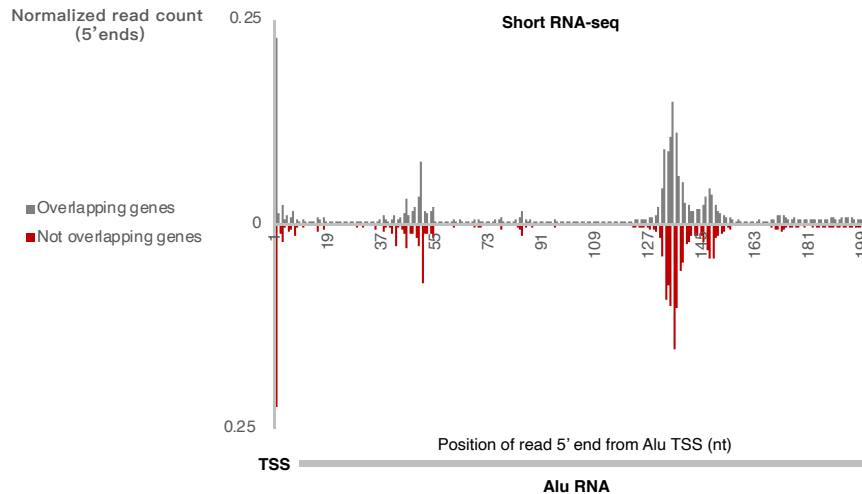


Appendix Figure S8. Expression of known hippocampal markers and cellular deconvolution in the MAP RNA-seq data

(A-B-C) Expression levels comparison between RNA-seq data from MAP RNA-seq data used in our study (panel A), RNA-seq data from the Allen Brain Atlas (Panel B) and RNA-seq data generated from microglia cells within the ROSMAP study (Panel C). Comparison is done for the standard panel genes of reference genes used by Allen Brain Atlas. Every column in panel B represents a different cell type in brain with cell types belonging to neuronal and non neuronal cells depicted at the lower part of panel B labelled as GABAergic, Glutamatergic and Non-neuronal. Markers that are enriched in neuronal vs non-neuronal cells were extracted from: [https://celltypes.brain-map.org/rnaseq/human_ctx_smart-seq] based on (Hodge et al, 2019; Tasic et al, 2018).

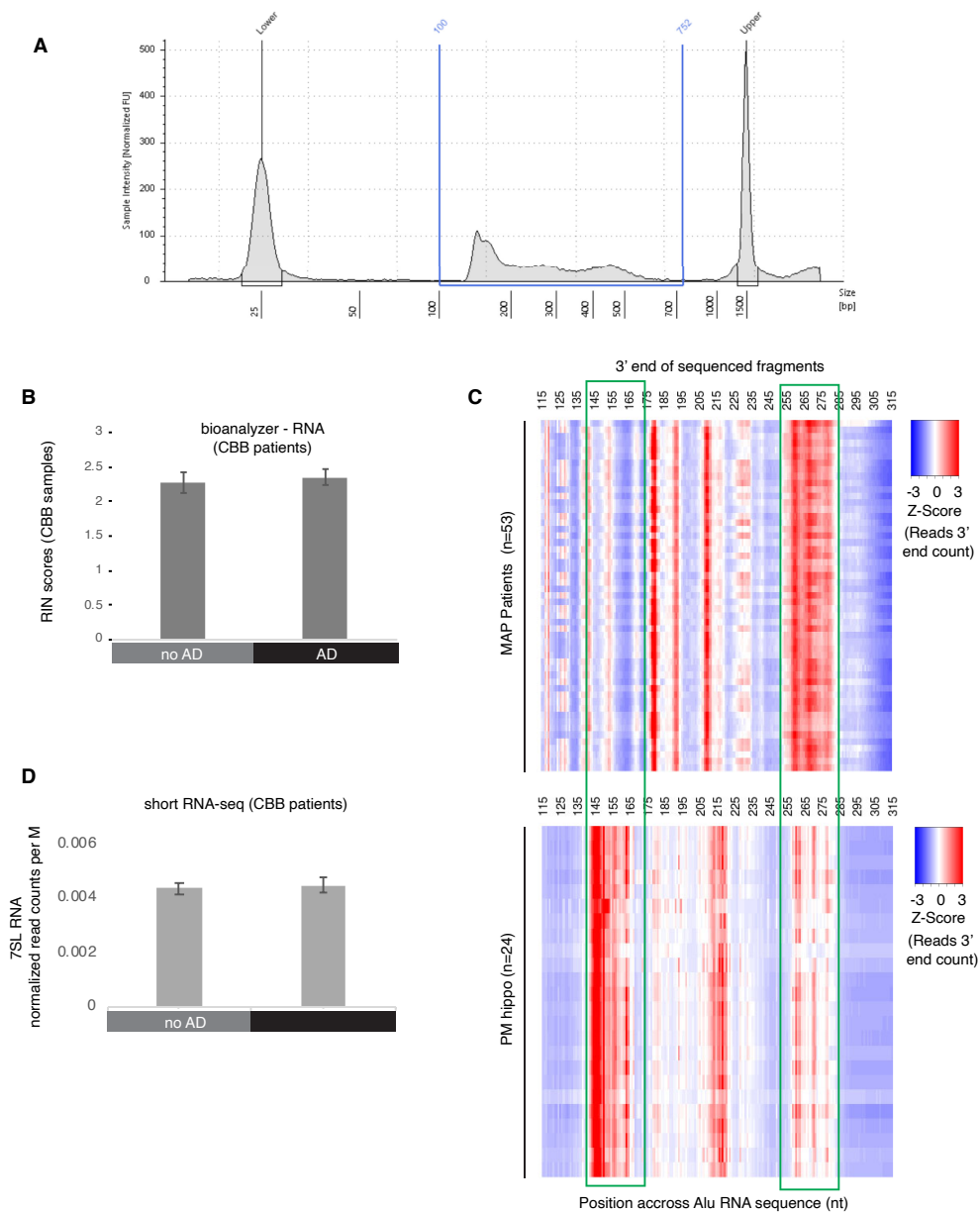
(D-E) Cellular deconvolution of RNA-seq data from MAP patients (bulk tissue-dorsolateral prefrontal cortex) and microglia isolated cells within the ROSMAP study. D represents the percentage of neuronal (NeuN positive) vs non-neuronal (NeuN negative cells) per sequenced MAP sample used in our study. E represents the overall percentage of each of these categories in each RNA-seq data set.

(F) Alu RNA levels comparison between the MAP patients (bulk tissue-dorsolateral prefrontal cortex) and microglia isolated cells within the ROSMAP study.



Appendix Figure S9

Plotting of the position of the first base (5' end) of Alu RNA fragments in a CBB patient against Alus that fall within known Ensemble genes (primary within introns) (upper panel) and Alus outside genic areas (intergenic space) (lower panel). The x axis represents an Alu RNA metagene aligned at the TSS of the Alu elements and the y axis shows the 5' end count for Alu RNA fragments aligning to any position downstream of position +1.



Appendix Figure S10

(A) Agilent Bioanalyzer electropherogram of one of the short-RNA-seq libraries we produced from RNA from the CBB patients' hippocampi. It shows a fragment range of 150-650nt, that corresponds to initial RNA inserts of 30nt to approx. 530 nt (adapter is approx. 120nt).

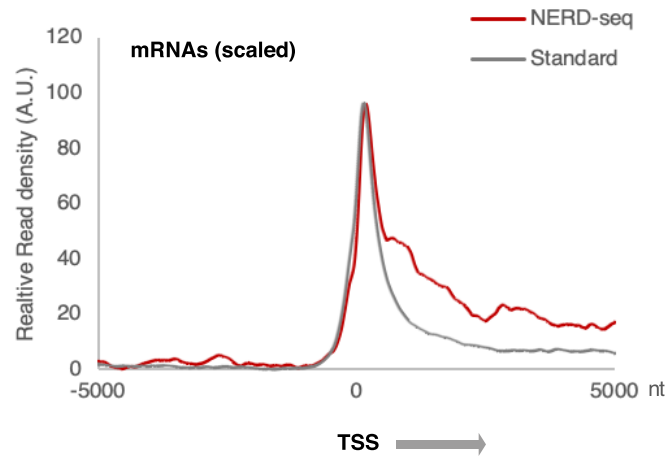
(B) RIN scores of RNAs sequenced from CBB patients.

(C) Plotting of the position of the 3' end of Alu RNA fragments across the end of the consensus Alu metagene to depict potential 3' end positions of Alu RNAs in all post-mortem hippocampal tissues (PM hippo) from patients from the CBB (lower panel) and in MAP patients (upper panel). Each row in the heatmap depicts the distribution of counts of the 3' ends of reads mapped across the Alu metagene for each patient. The x-axis represents a metagene combining all unique Alu

RNA sequences (ALUome) as in Figure 1. Heatmap density corresponds to normalized counts of the 3'end of the reads with red corresponding to higher density of these 3'ends at a specific position. Orange rectangles mark expected end positions for full length fragments (right) and fragments generated by the XR1 processing point (left).

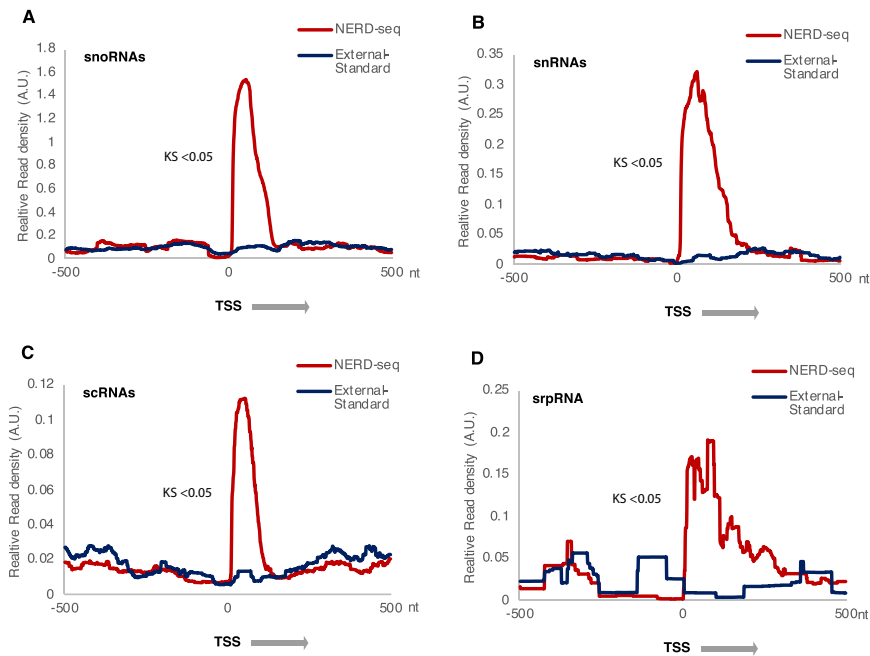
(D) 7SL RNA levels based on short-RNA-seq in CBB patients used for the normalization of total Alu RNA base coverage in Figure 2.

Appendix 3 – Supplementary for chapter 4: NERD-seq: A novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs



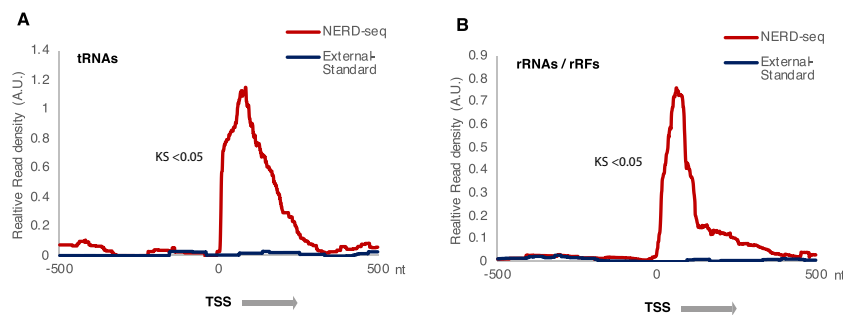
Supplementary Figure 1.

Relative read density distribution around the Transcription Start Site (TSS) of known genes for standard and NERD-seq. A metagene model has been as in Figure 3. Distances at X axis correspond to absolute distance (in nucleotides; nt) 500nt upstream (left) and downstream (right) from TSS. The arrow next to TSS depicts the direction of transcription. Read density is as in figure 3, but has subsequently been scaled within each dataset to allow qualitative comparison of the two distributions.



Supplementary Figure 2. Validation of results on snoRNAs, snRNAs, scRNAs and srpRNAs using an external dataset that has been generated by the standard direct RNA-seq approach(64).

- (A) Relative read density distribution around the Transcription Start Site (TSS) of snoRNAs for external standard and NERD-seq. X, Y axis and KS-test as in Figure 3.
- (B) Relative read density distribution around the Transcription Start Site (TSS) of snRNAs.
- (C) Relative read density distribution around the Transcription Start Site (TSS) of scRNAs.
- (D) Relative read density distribution around the Transcription Start Site (TSS) of srpRNA.



Supplementary Figure 3. Validation of results on tRNAs and rRFs using the external dataset that has been generated by the standard direct RNA-seq approach(64).

- (A) Relative read density distribution around the Transcription Start Site (TSS) of known tRNAs (mm10) for external standard and NERD-seq. X, Y axis and KS-test as in Figure 3.
- (B) Relative read density distribution around the Transcription Start Site (TSS) of rRNAs (mm10).